THE DESIGN AND EVALUATION OF A TEMPORAL-AURAL-VISUAL REPRESENTATION TO SUPPORT MIDDLE SCHOOL STUDENTS' CONCEPTUALIZATION OF THE RANGE OF IMPERCEPTIBLE SIZES

by

Min Young Song

**A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Educational Studies)
in the University of Michigan
2013**

Doctoral Committee:

Associate Professor Chris L. Quintana, Chair
Associate Professor Barry J. Fishman
Professor Nadine B. Sarter
Associate Professor Priti R. Shah

To my parents.

ACKNOWLEDGEMENTS

I cherish learning. Learning happens everywhere and at every moment. Every little action and bit of thinking we conduct results in changes in our cells, neurons, and brain, and eventually, we grow. Among all the learning that I thankfully have taken in my life, I can confidently say that it is during my Ph.D. program that I learned the most precious knowledge about **many things**. Especially, this dissertation carries all the lessons I have learned regarding my passion, doubt, faith, failure, and success throughout my academic journey. I am grateful for many people who granted me such precious opportunities for learning and encouraged me to keep growing while writing this dissertation.

First, I would like to send my deepest thanks to my advisor, Professor Chris Quintana, for his persistent support and invaluable advice throughout the years of my Ph.D. program. His profound knowledge and insights gave me many enlightening moments and kept my passion burning. I would also like to thank my other committee members, Professor Barry Fishman, Professor Nadine Sarter, and Professor Priti Shah, for providing me with priceless feedback and encouragement. Professor Fishman helped me think beyond the scope of my knowledge by sharing with me his great expertise in learning technologies and educational studies. Professor Sarter, with her expertise in human factors, provided me with detailed advice and novel perspectives, which I could not have envisioned before. Professor Shah guided me in framing and interpreting my research by giving me advice on how I could untangle the puzzles of my research in the context of cognitive and educational psychology. I am privileged and honored to have had these great scholars on my committee.

I am grateful for receiving support for this research from the teachers who participated in it – Ms. Amy Snapke, Ms. Barbara Buckwalter, and Ms. Barbara Karalash – and their students. In particular, the teachers' enthusiasm and pride as science educators significantly energized and inspired me. I also acknowledge my colleagues and the faculty of the School of Education of the University of Michigan. I especially thank Professor Joseph Krajcik for giving me the opportunity to join a team of great researchers and to carry out this exciting research. I also appreciate the support from the staff, Mr. John Moje, Ms. Eileen Brussolo, and Ms. Alfreda Fleming, who processed the grants that were used to conduct this dissertation research.

Both when I was confronted with uncertainty and when I was celebrating achievements, there were precious friends who always offered me comfort and cheer. I am blessed to have met Laurie, my spiritual sponsor, who shared her boundless love, kindness, and wisdom with me. She is a living saint. I cannot give enough thanks to Clara, who planted warmth and courage in my mind. She showed me how brightening a caring heart can be. Lastly, no words can describe how grateful I am for my parents and brother, for their infinite love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Supporting students in conceptualizing the range of imperceptibly small sizes (e.g., sizes of atoms or molecules) has been a challenging topic in science education. Commonly used macroscopic visual representations of imperceptible sizes have been unsuccessful, mainly because they perceptually contradict the definition of “imperceptibly small”. Research indicates that learners may benefit from a novel representation that incorporates a non-visual modality for conveying imperceptibly small sizes. To address this issue, an animated temporal-aural-visual representation (TAVR), which accumulates imperceptible objects across the diameter of a pinhead, was designed. In TAVR, the size of an imperceptible object is represented through its total accumulation duration (time necessary to span the pinhead), and the range of imperceptible sizes is conveyed by the range of the accumulation durations of different objects. Prior studies showed that seventh grade middle school students could understand what a TAVR represents and that they constructed more refined mental models of the range of imperceptible sizes after TAVR interactions. However, the roles and the influences of particular TAVR features, which aimed to augment learners’ temporal experiences, in students’ interpretations of the range of imperceptible sizes were unidentified.

In this context, this dissertation investigated the TAVR features in three different aspects: the effects of (1) different combinations of aural and visual modalities, (2) different accumulation intervals (ten objects/sec vs. one object/sec), and (3) perception of the passage of time (kinesthetic fast-forwarding vs. natural passage of time), with two hundred thirty-one 7th grade students. Multiple measures including surveys, pre- and post-instructional card-sorting tasks, students’ self-reported reflections, and focus group interviews were examined.

The results indicated that the students who interacted with TAVRs with the features that helped them more intensively perceive the durations of the different accumulation progressions (i.e., visual representation, slower accumulation interval, or natural perception of the passage of time) experienced vast range of temporal durations and, hence, generated the most refined mental models of the range of imperceptible sizes. Based on these findings, detailed discussions on the roles of each of the augmenting features in TAVR, possible scenarios for using temporal representations for learning, and future research topics are presented.

CHAPTER 1 INTRODUCTION

Representations are used in almost every science classroom. Representations such as figures, numbers, or symbols allow a learner to understand concepts that are absent in space and time and to access knowledge and skills that are beyond his or her cognitive capability. Such roles of representations become more critical when students learn about imperceptible phenomena. When students try to understand the size and behavior of subatomic particles or the distance between two planets, they need to think about things that humans can neither view nor touch. Since learners cannot directly perceive such imperceptible phenomena, they must depend on mediations of the phenomena – representations – to understand them. No one has ever directly seen what a hydrogen atom looks like, but its representation enables learners to visualize its shape and behavior.

Teaching and learning about the range of sizes of objects that are too small to see with human eyes[1] (objects such as cells, bacteria, viruses, DNA, molecules, and atoms), which I call imperceptible sizes, has been a challenging issue in science education (Tretter, Jones, Andre, Negishi, & Minogue, 2006; Tretter, Jones, & Minogue, 2006). Commonly used representations for conveying different imperceptible sizes normally provide learners with macroscopic depictions of imperceptible objects, which vary from static textbook graphics to interactive multimedia. The images of imperceptible objects in textbooks are usually consecutively aligned to convey their relative size differences.

[1] The limit of human vision lies at about 100 micrometers (μm) (*Encyclopedia Britannica*, 2012). One micrometer equals one-thousandth of one millimeter.

Interactive visual representations also present learners with certain types of macroscopic depictions of imperceptible objects using digital technology. For example, in a video called *Powers of Ten* (Eames Office, 1977), the graphics portraying imperceptible objects gradually become enlarged to a visible scale through automatic zooming-in animation. Similarly, the images of imperceptible objects in *Scale Ladder* (Nanoscale Informal Science Education, 2010) become visible through a click-to-zoom action by the learner. These representations are frequently accompanied by notations of relative size (e.g., "The diameter of a hydrogen atom is about 10,000,000 smaller than one millimeter") or absolute size (e.g., "The diameter a hydrogen atom is about 0.1 nanometer"[2]) of the represented objects to indicate that the sizes of the images are not the real sizes of the objects, and that in fact they are much smaller than they appear in the images.

However, it is known that many students construct naïve conceptions of imperceptible sizes, despite the use of such representations (Tretter, Jones, Andre, et al., 2006). While there exists a vast range of imperceptible sizes, research by Tretter, Jones, Andre, Negishi, & Minogue (2006) showed that middle school students tend to think that the sizes of imperceptible objects are similar to each other, even similar to the size of a small macroscopic object such as a fine grain of salt or a dust particle. For example, when middle school students were asked to classify several objects (e.g., atom, molecule, virus, bacterium, cell) by similar sizes, they classified the imperceptible objects into one "small" size group, while experts created at least three different size groups: sub-nano, nano, and microscopic sizes (Tretter, Jones, Andre, et al., 2006). Hence, as illustrated in Figure 1.1, middle school students tend to conceptualize the range of imperceptible sizes

[2] One nanometer (nm) is one million times smaller than one millimeter.

to be narrower than experts do. Moreover, it was found that some students misunderstood the meaning of “too small to see.” For example, in interviews that were conducted as a part of a pilot study for this dissertation, it was observed that some students even thought that bacteria and viruses (which they called “germs”) are as big as fine-grained salt, but that they are invisible because they are transparent.

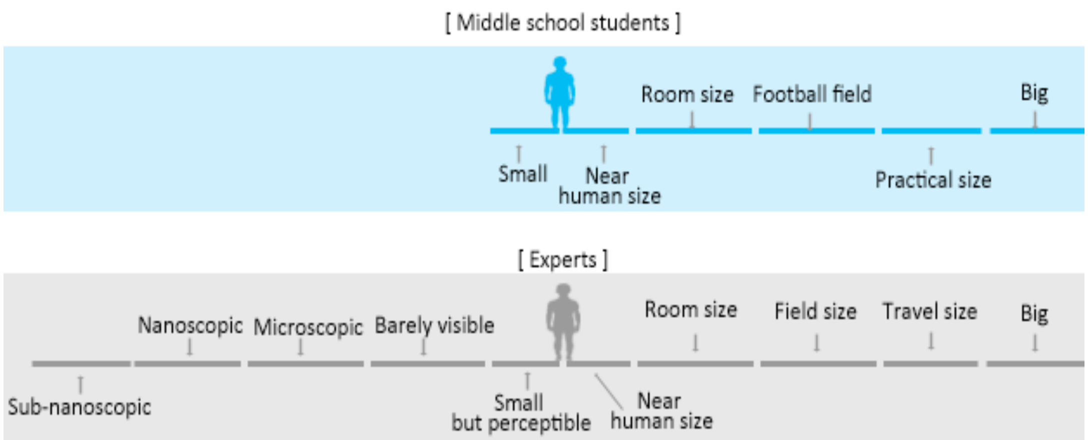


Figure 1.1. A visual representation of the size conceptions held by middle school students and experts.

These naïve conceptions of learners arise mainly because of the simple fact of nature that humans cannot see objects that are smaller than 100 micrometers. Tretter et al. (2006) discussed the point that directly seeing the entire body of an object is critical for precisely conceptualizing sizes. The students (from middle school to high school) who participated in their study exhibited the most accurate knowledge of the sizes of those objects for which they could have direct and holistic visual experiences. Consistent with this argument, studies on human spatial cognition emphasize the centrality of direct visual experience, arguing that spatial cognition is most frequently and pervasively based on what is perceived to exist and what has already been directly and visually experienced (Borst & Kosslyn, 2008; Kosslyn, Behrmann, & Jeannerod, 1995). Research on how people compare different sizes (Kosslyn, 1980, 1994; Kosslyn, Murphy, Bemesderfer, & Feinstein, 1977) also concludes that such visual memory of an object plays a critical role in effectively and accurately conceptualizing its size.

A critical practical challenge to learners is to try to understand imperceptible sizes using macroscopic visual representations of imperceptible objects. Studies indicate that learners' naïve conceptions are developed under the influence of big-enough-to-see images of imperceptible objects, which imply the exact opposite of "too small to see." According to Tretter and his colleagues (2006; 2006), frequent exposure to macroscopic depictions of imperceptible objects seems to cause students to overestimate the sizes of those objects. Although learners are told that the macroscopic visual representations are much bigger than the actual sizes of the objects, some students even tend to think that the size of the visual representation of an imperceptible object is the actual size of the object. Graphic images are very effective for illustrating shapes, features, movement of or physical relationships between certain imperceptible objects, as in the way they are frequently and effectively used to illustrate many scientific concepts; however, they do not directly convey how imperceptibly small they are. Since learners tend to focus on the perceptually dominating surface features of a representation rather than trying to decode its underlying meanings and theories (Kozma, Chin, Russell, & Marx, 2000), macroscopic depictions of imperceptible objects that require learners to visually imagine how small they are in order to conceptualize their sizes present a difficult task.

To summarize, learners have difficulty in conceptualizing how small is imperceptibly small and how varying imperceptible sizes are, mainly because of the limits in what they can visually perceive with their naked eyes. Macroscopic depictions of imperceptible objects are not considered to be useful because of the perceptually conflicting information they convey; the images are large enough to see with naked human eyes, while they claim the objects they represent are too small to see. Due to these challenges, learners tend to think that the sizes of imperceptibly small objects are roughly the same, even the same as the size of a fine grain of salt.

Since learners' perceptions of imperceptible phenomena are mediated only by representations, learners need an alternative way of conceptualizing how small and various imperceptible sizes are. Representations co-determine the very nature of the human cognitive task, and interaction with representations may enhance and transform human cognition (Hutchins, 1995; Pea, 1993; Salomon, Perkins, & Globerson, 1991; Zhang & Norman, 1994). Different representations that aim to represent the same concept can make thought processes less or more difficult. It is also known that a novel representation that directs learners to explore a phenomenon in different ways may help them recognize and revise their misconceptions of the phenomenon by revealing the inconsistencies between their mental models and the new representation (Ainsworth, 1999; Chi, 2005). A novel representation that does not employ macroscopic depictions of imperceptible objects may provide better support for learners in conceptualizing the range of imperceptible sizes.

With the advancement of computer technologies, the representations that are adopted in learning technologies can benefit from multimodal representations. Compared to static representations that are printed on paper, computer technologies can extend what learners are able to explore by providing them with an environment that incorporates interactive multimodal representations. The benefits of using multimodal learning technologies for science learning have been shown through several research projects (Buckley, 2000; Moreno & Mayer, 1999, 2007; Sweller, 2005a). Such technological power may address learning goals that cannot be met in any other ways.

In order to achieve this goal, an alternative representation, called a temporal-aural-visual representation (TAVR) was designed. It does not use a macroscopic depiction of an imperceptible object as its main vehicle for conveying imperceptible sizes. A TAVR takes the temporal sense as the main modality. The "temporal sense" here refers to the duration of a sequential accumulation that happens on the head of a pin. It sequentially places imperceptible objects across the head of a pin, which is 1 millimeter in diameter. This sequential accumulation of imperceptible objects is continued until the objects are fully lined up across the pinhead. Hence, the duration of the sequential action inferentially implies the size of the placed object via the inverse relationship between time and size; the smaller the object, the longer the accumulation time. The temporal representation is incorporated with two other modalities – aural and visual, which are adopted to augment learners' temporal experiences. The accumulation within the imperceptible scale is indicated by clicks, and visual representation (the red line) is added only after the length of the accumulation becomes macroscopic in scale. See Figure 1.2 for the illustration of how the accumulation in the TAVR progresses.

Prior work (Song & Quintana, 2009, 2010) investigated whether middle school students could understand TAVR and how their understanding of the range of imperceptible sizes changed over the learning activity with a TAVR. The results indicated that the middle school students could accurately interpret how a TAVR works and what it represents. Also, the students exhibited evidence of improvement in their knowledge of the range of imperceptible sizes. For example, at a card sorting task in

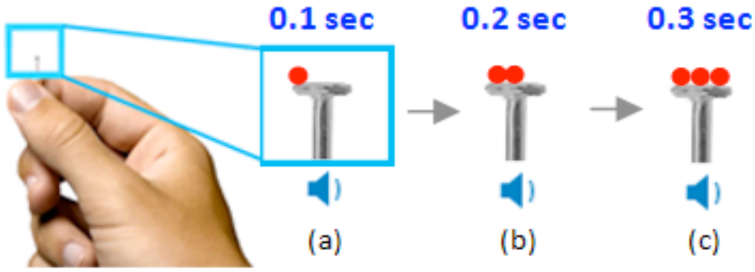


Figure 1.2. An illustration of the accumulation process in TAVR.
(a) The first object is placed on the pinhead, and one click is played. (b) The second object is placed on the pinhead next to the first one, and one click is played. (c) The process continues until the object spans the pinhead.

which the students classified imperceptible objects such as an atom, a molecule, and a cell according to size, the number of the size groups that they generated increased after the learning activity with TAVRs, indicating that the range of imperceptible sizes in their minds had expanded. These findings indicate that the temporal aspect of TAVR was useful for guiding students to recognize that the sizes of imperceptible objects are too small to see and that there exist a vast range of imperceptible sizes.

With these findings, more research topics for further investigation emerged. Although the positive impacts of TAVRs on students' conceptualization of imperceptible sizes were shown in the prior study, the roles and the influences of the individual features of TAVR in students' interpretations of their TAVR interactions were unidentified. Information on such topics will lead to the development of research-based guidelines for using temporal representations for teaching a concept that involves imperceptible sizes and their range. Hence, in this dissertation, I specifically explored three aspects that involve (1) a modality question: how to cognitively optimize students' temporal experiences with different combinations of the two supporting modalities – visual and aural, (2) a time perception question: how to maximize students' conceptual range of imperceptible sizes by using different accumulation intervals (i.e., accumulation rapidity 1 object/sec vs. 10 objects/sec), and (3) a temporal manipulation/interaction question: how to augment students' temporal experiences with the different time manipulation features. In the following I discuss each research question and corresponding hypotheses.

Research Question 1

How do the combinations of the supporting modalities support learners to construct the mental model of the range of imperceptible sizes?

A TAVR is composed of a temporal representation and two other representations of different modalities, aural and visual, which are adopted to provide support for learners' comprehension of their temporal experiences of imperceptible sizes. However, it is not guaranteed that learners would use and benefit from the two supporting modalities, because it cannot always be assumed that people will interact multimodally with a multimodal system (Oviatt, 1999). This issue brings up the need for an investigation of whether students utilize each augmenting modality, and if so, how they interpret and synthesize what was conveyed via the augmenting modalities.

Additionally, it will also provide information as to whether the modalities in TAVR overload students' cognitive capacity. According to the cognitive load theory, a learner should be facilitated in using his/her limited working memory efficiently because human working memory is limited with respect to the amount of information it can hold and the number of operations it can perform on that information (Pass, Renkl, & Sweller, 2004; Sweller, 1988). If the resources of a learner's working memory are exceeded due to the difficulty of the task or the amount of given information, his or her learning is likely to be ineffective. If all three modalities, or certain combinations of them, in the temporal representation overwhelmed students' cognitive capacity, this would be revealed in the students' reflections on their interaction with the temporal representation. Although theories on human information process (see Paivio, 1986; Baddeley, 1986) and working memory structure theory imply that visual and aural information are processed in parallel in human working memory and, in consequence, effectively comprise one's comprehension of the target information, it cannot be guaranteed that the students will easily integrate and process all information when the temporal modality is added to the aural and the visual modalities in one representation.

I hypothesize that students who experience both supporting modalities will develop more refined knowledge of the range of imperceptible sizes than students who use a single supporting modality (either aural or visual) because having more information about the progress of accumulation will help students to comprehend what the temporal representation represents and how the accumulation is progressing, despite the possibility of being cognitively overwhelmed by stimuli from two different modalities.

Research Question 2

How do different intervals of temporal experiences influence learners' conceptualization of the range of imperceptible sizes?

In the prior study (Song & Quintana, 2010), it was observed that 7th grade students made meaningful interpretations of the accumulation durations of imperceptible objects that were shown in TAVRs, especially when the sizes of the objects were submicroscopic (smaller than one micrometer). The accumulation durations of those objects were longer than one day (i.e., about eleven days for hydrogen atoms and three days for water molecules), and the students thought that the durations longer than one day were considerably long, and they inferred that the objects were extremely small. For example, the students responded that the objects were "extremely smaller than they thought," because most of them did not expect that the accumulations would take such long time, particularly when compared to other imperceptible objects, such as e coli bacteria (fifty seconds) or red blood cells (sixteen seconds).

This observation implies that the radical differences between the total durations of the object accumulations in TAVRs might direct learners to conceptualize an even more expansive range of imperceptible sizes. The accumulation durations become dramatically longer when the accumulation interval is set to one object per second. In this case, the relative differences between the durations of accumulation of the imperceptible objects become greater, while the numbers of the accumulated objects are the same. For example, about ten million hydrogen atoms can be placed on the diameter of a pinhead. With the accumulation interval of one object per one second, the accumulation duration of hydrogen atoms becomes about 115 days (eleven days at the interval of ten objects in one second), and that of a relatively bigger object, a water molecule, for instance, becomes twenty-three days (two days at the interval of ten objects in one second). Therefore the difference between the accumulation durations of these two objects becomes about ninety-two days at the interval of one object in one second. Students who interact with the TAVRs with slower accumulation interval may think the sizes of these objects are drastically different than the students who interacted with the TAVRs with the faster accumulation interval. Hence, with the second research question of this dissertation, the effect of different accumulation intervals on students' conceptualization of the range of imperceptible sizes was examined.

I hypothesize that the exaggerated differences between the total accumulation durations of imperceptible objects may direct students to conceptualize more a expansive range of imperceptible sizes. The students who interact with the TAVRs with the slower accumulation interval (one object/sec, which I will call the "extended condition") will interpret the sizes of imperceptible objects to be more discrete and distinguishable than the students who interact with the TAVRs with the faster accumulation interval (ten objects/sec, the "compressed condition"). As a result, the extended condition students will construct mental models of a more expansive range of imperceptible sizes.

Research Question 3

How do active temporal manipulations vs. passive observation by students influence their perceptions of the durations of the temporal experiences with TAVRs and the conceptualization of the range of imperceptible sizes?

Research question three concerns the impacts of the students' temporal manipulation methods on their conceptualizations of the range of imperceptible sizes. The impact of one feature, called *Skip ahead* buttons, which seemed to be influential to the students' interpretation of their temporal experiences in the pilot study, was not investigated in the prior work. *Skip ahead* buttons are embedded in the TAVRs of all imperceptible objects in order to enable students to observe the accumulations of all imperceptible objects being completed within one class. When students use the buttons while the accumulations are progressing, the total accumulation durations of each imperceptible object that students actually perceive become much shorter than those that appear in TAVRs. For example, it takes about eleven days for the accumulation of hydrogen atoms; however, by clicking the *Skip ahead* button of the hydrogen atom

TAVR, a student can accelerate the accumulation and observe its completion in less than ten minutes, depending on how frequently he clicks, without having to physically wait for eleven days to witness the completion of the accumulation.

In the prior study, it was observed that most of the students understood the use of *Skip ahead* buttons. They actively used *Skip ahead* buttons almost from the beginning of the accumulations because they wanted to accelerate the accumulations "*to figure out which one is the smallest.*" They utilized the buttons to figure out which object in fact takes the longest time to accumulate and hence is the smallest. Many of the students seemed to be excited about the repetitive and laborious button pressing action. Some students even raced with peers to compete to be the first person who completed the accumulation, particularly for the smallest object (hydrogen atom). The students remembered such arduous kinesthetic experiences and used the memory when trying to interpret their temporal experiences. Knowing whether and what impacts the kinesthetic manipulations have on the students' conceptualizations of the range of imperceptible sizes will permit the development of design suggestions regarding how to augment a learning activity that employs temporal representations.

To investigate this matter, three different student groups were formed, and their achievements were compared. To distinguish whether the effect of the kinesthetic manipulation was specifically due to the kinesthetic input from students, a group of the students who interacted with the TAVRs with the *Skip ahead* buttons (whom I call "IM - interactive manipulation" group) was compared with the group of students who interacted with the TAVRs that automatically accelerate the accumulations ("AM – automatic manipulation" group). Furthermore, to identify whether the manipulation of the temporal experience is ever more effective for constructing more refined mental models in students, a group of students interacted with the TAVRs that provided no manipulating feature, which forced the students to experience the natural passage of time ("NM – no manipulation" group). Their achievements were compared with the other two groups as well.

I hypothesize that the interactive kinesthetic manipulation of the temporal experience will work more powerfully than the natural passage of time (no manipulation) because the kinesthetic manipulation will reinforce students' development of another layer of meaning from their learning activity with the TAVRs. Students will remember the sensation that they felt in the muscles of their fingers and wrists while pressing the *Skip ahead* button, which gradually became irritating after pressing the button for a certain period of time. The TAVRs with no manipulation feature may have less impact on students' conceptualizations of the range of imperceptible sizes than the TAVRs with the interactive manipulation condition because the perceptually long temporal experiences can color the way students interpret the total accumulation durations of the TAVRs; for example, they may interpret ten minutes to be extremely long if the waiting was boring to them. Lastly, I predict that the TAVRs with the automatic acceleration feature will have the least impact on student conceptualizations of the range of imperceptible sizes because the students who use this type of TAVR will have to perceive the accumulation durations that are shorter than the real passage of time without having any opportunity to enrich their abbreviated temporal experiences.

Significance of This Study

Unlike visual or aural modalities, the temporal modality has not been explored as a form of representation for conveying an abstract concept in the field of learning technologies. In addition to showing how interacting with a novel form of multimodal representation can alter the ways students think about an abstract concept and consequently improve the comprehension of it, I expect that the results of this dissertation will inform the community of learning technology researchers and designers about how a temporal modality can be used to expand the potentials of learning technologies by expanding what has been available in conventional multimedia that most commonly have adopted visual and aural representations.

Specifically, this dissertation shows cases of how a temporal representation can be used to enhance students' conceptualization of the range of imperceptible sizes. This study may provide science educators with information regarding how to support students' conceptualizations of imperceptible and abstract spatial information, which has been a challenging topic in science education (Tretter, Jones, Andre, et al., 2006), with a representation that utilizes the concept of time.

Further, while this study is important for designers of learning technologies, it has further implications for human-computer interaction (HCI) research as well. It points to the potential role of a non-typical modality in expanding our experience of the world. It may also inform as to how the interaction with a novel form of technology can alter the ways people think about an abstract concept and, consequently, improve the comprehension of it.

Overview of the Dissertation

In this chapter, I provided a rationale and argument structure for the need of this study and introduced research questions, correspondent hypotheses, and significance of this study. In Chapter 2, the literature on how people perceive and conceptualize sizes, the in-depth details of the challenges to learners, and the theories on representations that led to the design of TAVRs are discussed. Chapter 3 explains the details of the features of TAVR with the theoretical rationale. It also introduces a Flash™ application named *Wow, It Is Small! (WIIS),* a learning environment in which students can interact with TAVRs, and brief findings from the prior work. In Chapter 4, the research design, the construct, and the research context are elaborated with discussions of what data were collected and how they were analyzed. In Chapter 5, the results of this study are presented. I report on the results from data analyses and discuss whether the hypotheses were supported. In the last chapter, Chapter 6, I formulate possible alternative explanations for any inconsistencies between the hypotheses and results, and discuss how the findings respond to the literature. Then I provide a suggestion as to how to effectively implement and use a temporal representation in a learning technology, how to effectively exploit it in association with other learning materials. Finally I discuss potential future studies, which focus on issues that remain unanswered or unexplored due to the scope and limits of this study.

CHAPTER 2 LITERATURE REVIEW

In this chapter, I place the present study in the context of previous research. I begin by discussing the aspects that may constitute the knowledge of the range of imperceptible sizes. Then I analyze why it is difficult for learners to conceptualize imperceptible sizes and their ranges, and why they develop misconceptions of such sizes, while, in contrast, they have concrete and accurate mental models of perceptible sizes. Then I review the research on the use of representations in science education, and particularly for teaching about imperceptible sizes. Finally I discuss the use of a multimodal representation in science learning and its potential roles for teaching imperceptible sizes.

Student Understanding of Imperceptible Sizes

Mental Model

An individual constructs in his/her mind a mental model, which is also called "knowledge structure" or "memory structure," in the process of understanding incoming information (Wiley & Ash, 2005) from external stimuli. A mental model represents perceptual and conceptual features of a system, object, situation, event, story, etc, but is not an exact replica of them (Barsalou, 1999). Rather, they are abstract representations that store the spatial, physical, and conceptual features of the external stimuli, and they are retrieved from one's memory in order to be used in problem solving, inference generation, and decision making (Rapp, 2005). Hence, one's mental model of a phenomenon is composed of information from one's preexisting knowledge and from the external world. A mental model can be used to generate hypotheses, solve problems, and transfer knowledge to new domains (Rapp, 2005). In this way, mental models are used for understanding information that is conveyed by a representation or for constructing a more profound comprehension of target information.

There are two factors that mainly influence the content of a mental model: the *referent* and the *external representation*. A referent is an actual phenomenon that exists in reality with which a learner can have a direct perceptible experience. A referent in a classroom normally is a scientific phenomenon (e.g., chemical reactions) that learners must perceive and conceptualize. An external representation, which I simply call a "representation" in this dissertation, is a re-created model or theory of the referent, which in most cases does not exactly replicate the referent's physical attributes. External representations involve graphics, symbols, rules, constraints, and relations embedded in figures, such as spatial relations of written digits, visual and spatial layouts of diagrams, physical constraints in abacuses, etc. Such external representations give students access to knowledge and skills that are not available in their minds (Rapp & Kurby, 2008) and allow students to work with events and things that are absent in space and time (Norman, 1993).

Theories suggest that external representations not only allow a learner to understand a concept that could be beyond his or her perceptual or cognitive capability, but also shape or give rise to a mental model of the concept (Buckley, 2000; Rapp & Kurby, 2008; Zhang & Norman, 1994). Zhang and Norman (1994) called this mental model an "internal representation" and explained this relationship between internal representations and external representations by the representational effect theory. The theory argues that internal representations are reconstructed through the interweaving process of the internal representation and external representation because external representations are not simply incoming stimuli to internal minds; rather, they mediate the formation or the elaboration of students' internal representations of a particular concept or phenomenon.

The theories imply that different external representations of the same target concept can have different impacts on the cognitive tasks with which learners must contend. Graphs and diagrams are commonly used examples that show how representations (external representations) can change the way people think about a phenomenon. Graphs are frequently used to convey quantitative information that would be more difficult to understand when described textually and numerically. The climate change during the past hundreds of years is much easier to understand when it is represented in a graph than in a series of numbers that requires readers to mentally calculate and interpret those numbers. Diagrams are particularly useful for conveying complex causal relationships or processes. A diagram of the water cycle would be much easier to understand than verbose textual descriptions of it.

However, improving one's mental model of a certain concept with the immediate information in a presented representation is not always guaranteed; rather, individuals tend to rely on their preexisting knowledge, often to a fault (Johnson & Seifert, 1999; van Oostendorp & Bonebakker, 1999). Since learners actively refer to their preexisting knowledge and their new knowledge is constructed upon it (Jonassen, 1994; Piaget, Gruber, & Vonèche, 1977; Wood, Bruner, & Ross, 1976), the way they perceive and conceptualize information is profoundly influenced by what they already know and what they have experienced. When a learner attempts to comprehend a representation, he or she retrieves the elements from his or her mental model and compares them with the representation to decode which components of the representation stand for which aspects of the referent (Buckley, 2000; Rohr & Reimann, 1998). Therefore, the weaker the student's preexisting knowledge about a concept, the more likely he or she is to inappropriately interpret its representation.

The information from either the referent or the representation comes in several different sensory modalities, such as vision, sound, touch, and smell, because the external stimuli, which include both the referent and the representation, reach a learner in multiple modalities. For example, a mental model of a spatial configuration (e.g., structure of a building), can be constructed of what one has perceived from direct visual, auditory, and haptic experiences (Schnotz, 2005). Such information in diverse modal forms that are stored in our mental models are called *mental images*. Mental images are a specific type of internal representation that is produced during perception (of either a referent or an external representation) and created from stored internal representations in memory (not directly from sensory input) (Kosslyn, 2005). Hence, mental images mimic the corresponding events in the world and can exist in the form of many different modalities other than visual, including kinesthetic, spatial (which includes size), auditory, and tactile, as researchers have proposed that mental models are "imagistic," but not inherently visual image-based (e.g., Kosslyn, 1994; Phylyshyn, 2002). For example, a mental model for celestial bodies in the Milky Way would not simply be a mental picture or video of the information, but an abstraction of the universe that conveys organized relationships between objects based on size, distance, and etcetera (perhaps through hierarchical organization or some other association-based system) (Schnotz, 2005).

Mental images in diverse modal forms are not the only component of a mental model. Theories on human working memory suggest that in addition to mental images that are processed nonverbally, there also exists verbal information. Nonverbal and verbal types of information are processed synchronously yet independently in parallel in separate cognitive modules in our working memory (Baddeley, 1986; Paivio, 1986). According to Baddeley (1986), the two parallel channels are: the phonological loop and the visuo-spatial sketchpad. Later he added a third process called the "episodic buffer," which is dedicated to linking and organizing information across domains to form integrated units of visual, spatial, and verbal information (Baddeley, 2000).

Similarly, Paivio (1986) suggested that human cognition can simultaneously deal with objects or events in both verbal and nonverbal forms; he called the verbal information a "*logogen*" and nonverbal information (mental image) an "*imagen.*" Both models argue that linguistic information, which is processed by the verbal processor, and mental images are related to each other. In other words, the same information is represented in different forms in both the verbal information and the mental images.

It has to be noted here that the modality of the input information does not always result in the creation of mental images with the same modality (Pineda & Garza, 2000; Zolna, 2008). It was found that information of the same modality can be expressed through difference senses (e.g., spoken and written language), and the same modalities can be used to perceive information of different modalities (i.e., written text and pictures are both interpreted through the visual channel).

Figure 2.1 is an illustration of these relationships and the iterative interactions that contribute to the formation of a mental model. To summarize briefly, a mental model can be composed of (1) mental images that are shaped in different modalities and (2) verbal information. Both mental images and verbal information, which become our preexisting knowledge and a mental model of a certain phenomenon, are created via interactions with referents or representations. In an iterative fashion, a mental model in turn influences how we perceive and interpret referents and representations. Since learners normally lack preexisting knowledge of scientific phenomena, they have poorly operating mental models of those phenomena. Hence, they may benefit from having fluent experiences with referents in order to learn more effectively with representations.

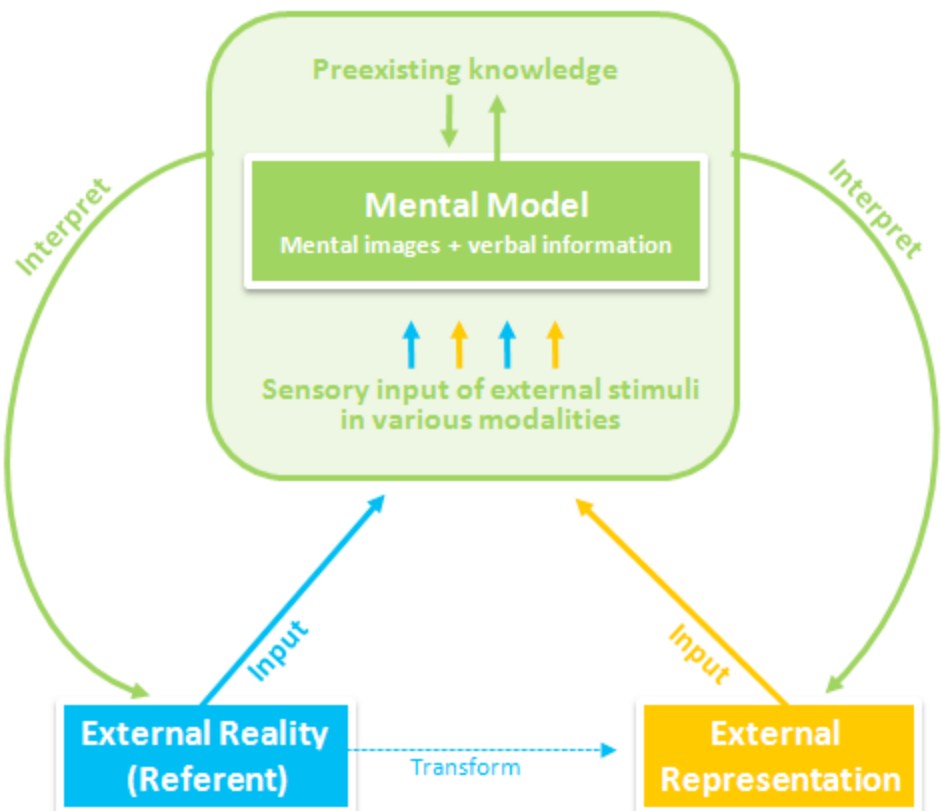


Figure 2.1. The iterative process of mental model formation.

The Components of the Mental Model of Sizes

The studies on human size perception and cognition (Kosslyn, 1980, 1994) indicate that a mental model of the size of an object can be composed of mental images and verbal information regarding the size of that object. The literature implies that mental images may involve visual or temporal-kinesthetic modalities, and verbal information can comprise conceptual size categories or propositional relationships between the sizes of objects. In the following, I introduce the details of these components. First I elaborate on both visual mental image and conceptual size categories, because these two components are the most commonly used resources when one tries to think about sizes. Then I discuss kinesthetic-temporal mental image of sizes, followed by the propositional relationship between different sizes. These resources are more commonly used for conceptualizing sizes of which one cannot have direct and holistic visual experiences.

Visual Mental Images and Conceptual Size Categories

A cognitive model of how people compare different sizes, suggested by Kosslyn (1980, 1994), gives the implications for the critical role of the visual mental images and conceptual size categories. A *visual mental image* is the mental invention or recreation of a visual experience that resembles the experience of perceiving an object or an event, either in conjunction with, or in the absence of, direct visual stimulation (Kosslyn, 1980). Visual mental images are important resources for scientific thinking, because learning-benefits, especially science learning, increase as a function of how easily a learner can develop a visual mental image of the to-be-studied information (Jenkins, 2010; Rapp & Kurby, 2008). Visual mental images are used as critical resources for size cognition. It was discovered that people recall and evaluate visual mental images in their memory, in the same way that they would evaluate what was imagined while actually seeing an object (Sims & Hegarty, 1997). A neuro-imaging study shows that the same parts of the brain are activated and used when people think of the image of an object, as when they were seeing the actual object (Roland & Gulyas, 1994). The development of visual mental images is highly influenced by direct and holistic visual experiences. Prior research on human spatial cognition emphasizes the centrality of personal experience, arguing that spatial cognition, which includes size cognition, is most frequently based on what has been directly perceived to exist (Wolpert, 1964).

Conceptual size categories are separable and distinct scale category tags that a person may have regarding the size of a certain object. For example, one may have a conceptual size category tag of "small" for a rat and "big" for an elephant. According to Kosslyn (1980, 1994), conceptual size categories are stored in our long-term memory when we see an object and are recalled when we compare the sizes of different objects. When a task of comparing the sizes of two different objects is given, people simultaneously access and use both visual mental images and conceptual size categories in parallel to determine which object is bigger or smaller. For example, when one attempts to compare the sizes of two objects with significantly different sizes (e.g., rats and elephants) to decide which one is bigger, one simultaneously begins to compare (1) the conceptual category tags of rat and elephant, and (2) the visual mental images of a rat and an elephant to develop an answer. In this case, the conceptual size category tags of the animals' sizes differ from each other (i.e., rat is "small" and elephant is "big"). In this case, where the conceptual size category tags are distinctively discrete, one can produce the conclusion that an elephant is bigger than a rat without having to recall the visual mental images of them. In contrast, when one attempts to compare the sizes of two objects with the same conceptual size category tags, rat and chick, for example, which are both “small,” one compares the visual mental images of them to decide which one is smaller than the other.

Kosslyn’s model implies that a visual mental image is the primary component that composes the mental model of a size. However, it is considered to be a resource only for the sizes of the objects that are appropriately large for humans to see, because visual mental images of objects are created when one has a direct visual experience of the objects. For example, in Tretter et al.’s study (2006), all the participants exhibited the most accurate knowledge of the sizes of objects that were available for direct and holistic visual experience, while their performances were relatively less accurate for the objects that were too small to see (imperceptible objects) or too big to see (e.g., the state of North Carolina or the Earth).

Kosslyn’s model also implies that conceptual size categories can become more useful components of the mental model of sizes than visual mental images because they can be retrieved and compared before visual mental images are completely recalled and compared (Kosslyn, 1994). This indicates that conceptual size categories can be particularly useful when one does not have a visual mental image of the object. The importance of conceptual size categories was also discussed in Tretter et al.’s (2006a) study. They used the term “unitizing” to describe how people use distinctive size categories to form a mental model of the range of sizes. Unitizing is a process of creating a new category unit from existing objects (Lamon, 1994), and it is a key capability of an expert that allows him or her to function using spatial knowledge, regarding size, volume, etc. The experts who participated in Tretter et al.'s study redefined inconvenient units when it was helpful for them to do so (i.e., converting 1/10000 meter to one millimeter). They said that rewriting the size in other units helped them better conceptualize the sizes than leaving everything in meters. This implies that a set of well-developed conceptual size categories would indicate a finely constructed mental model of size ranges, as also the research on expertise consistently shows that experts tend to have fine grained conceptual categories of the knowledge of their domain (Chi, Feltovich, & Glaser, 1981; Lakoff, 1987).

Temporal-Kinesthetic Mental Images

The human working memory model by Baddeley (1986) states that the tactile and kinesthetic (movement) senses, in addition to the visual and verbal senses, receive information and can help learners better understand incoming information. The kinesthetic or tactile senses provide additional "channels" that one can incorporate with (or use to complement) the information from verbal and visual channels in order to construct a more refined mental model of the information. This may mean that information about a size can also be stored and recalled if there exist kinesthetic experiences associated with that size.

Furthermore, kinesthetic mental images can comprise a mental model of sizes, in combination with temporal experiences, because the concepts of time and space are interwoven (Casasanto & Boroditsky, 2008; Droit-Volet, 2001; W. Friedman, 1979; Gentner, Imai, & Boroditsky, 2002; Jarman, 1979; McCormack & Hoerl, 2008; Tretter, Jones, Andre, et al., 2006; Vallesi, Binns, & Shallice, 2008). Providing evidence, Tretter et al. (2006) noticed that people conceptualized the sizes that are too big for the field of human vision by inferring them from the sequential physical actions (e.g., walking, running, or driving) that they made in relation to the sizes over a certain period of time. For example, the students who participated in their study compared the length of the time that it took to drive across a town and to drive across the state of North Carolina, in order to reason about which distance was longer and also to define how distinctively different the were from each other. Additionally, there exist examples of using the duration of a kinesthetic event to communicate about abstract spatial information in our daily lives. For example, the land size measurement unit called an "acre" has its origin in the unit that was used to communicate the amount of land that a man behind an ox could plow in one day (*Encyclopedia Britannica*, 2012). Astronomers use the unit called "light year" to communicate the extremely long distances between planets.

These examples imply that it is the temporal aspect of temporal-kinesthetic experiences that conveys the sizes that are much larger than the field of human vision. The duration of the temporal-kinesthetic event is used as a resource for inferring and judging such sizes. People are able to conceptualize the distance between two different locations when they know the total duration of the kinesthetic event involved in moving from one location to another. People normally can comprehend the lengths of time that are written in units of time, such as seconds, minutes, and hours, because their literacy of the passage of time is constructed throughout their daily lives, and they become fluent at understanding the length of time passage by the age of eleven on average (Acredolo, 1989).

Propositional Relationships Between Objects

Students sometimes make use of the propositional relationships between objects to judge the relative sizes of the objects, if they have such information. For example, some students who were interviewed for the pilot study of this dissertation answered that they knew that atoms were the smallest objects because they learned, *"atoms are the building blocks of every object,"* although they were not able to describe exactly how small atoms were. Some students could infer that molecules were bigger than atoms because molecules were the compounds of several atoms. A couple of students knew that viruses were smaller than a human cell, because they had heard that viruses enter human cells to make humans sick. These examples show that such propositional relationships between objects may give students some clues for inferring the relative sizes of the objects (i.e., which are bigger or smaller). However, this knowledge does not seem to be useful for enabling them to describe objectively how small or big the objects are. It only provides them with very basic propositions in verbal form, which can be converted into contextual information when they try to develop advanced comprehension of the sizes.

Students' Understanding of the Range of Imperceptible Sizes and the Issues

Documents of U.S. standards in science and mathematics define size and scale as concepts that encompass science and math, and which can be used to unify student learning across disciplines, topics, and grades; they are tools that help students understand the world (American Association for the Advancement of Science., 1993; National Academy & Research Council., 1996; National Council of Teachers of Mathematics., 2000). Furthermore, a "firm grasp on size and scale [is] a prerequisite for any further inquiry into nanoscale science and engineering" (Waldron, Sheppard, Spencer, & Batt, 2005). A recent article that identified the key concepts in nanoscience education states that "size and scale" is one of the core concepts of nanoscience (Stevens et al.,

2008) because knowing how small imperceptible objects are is critical for learning advanced scientific phenomena, such as size-dependent properties and behaviors.

Students' understanding of the range of imperceptible sizes is naïve and underdeveloped. Prior research shows that middle school students tend to overestimate the sizes of imperceptible objects and underestimate the differences between distinctive sizes of such objects. In their study, Tretter et al. (2006) interviewed people at different levels of expertise (1^{st} -12^{th} grade students and expert scientists) to investigate their knowledge of the range of sizes. The participants were provided with cards depicting thirty different objects with different sizes from the diameter of an atom to an inter-planetary distance (e.g., the distance between Earth and Moon). The cards showed macroscopic images and the names of the objects. The researchers asked them to order the objects from the smallest to the biggest and then classify the ordered objects by similar sizes. In the results, the biggest difference between experts and learners was found for the imperceptible sizes. For example, as shown in the upper graph in Figure 2.2, the middle school students categorized everything smaller than a human into one "Small" category that cannot be divided further. In contrast, experts (see the lower graph in Figure 2.2) formed more categories of sizes smaller than a person, creating one category for small but visible objects (Small), and two separate groups (Many Atoms and Atomic) for imperceptible objects.

Based on this observation, Tretter et al. found that students tended to overestimate the sizes of imperceptible objects and underestimate the differences between distinctive imperceptible sizes. As Figure 2.3 represents, the middle school students tended to believe that all objects that are too small to see with the naked eye were approximately the same size. Some students even thought that all imperceptible objects were similar in size, even similar to the sizes of small macroscopic objects (e.g., a grain of fine grained salt or a dust particle).

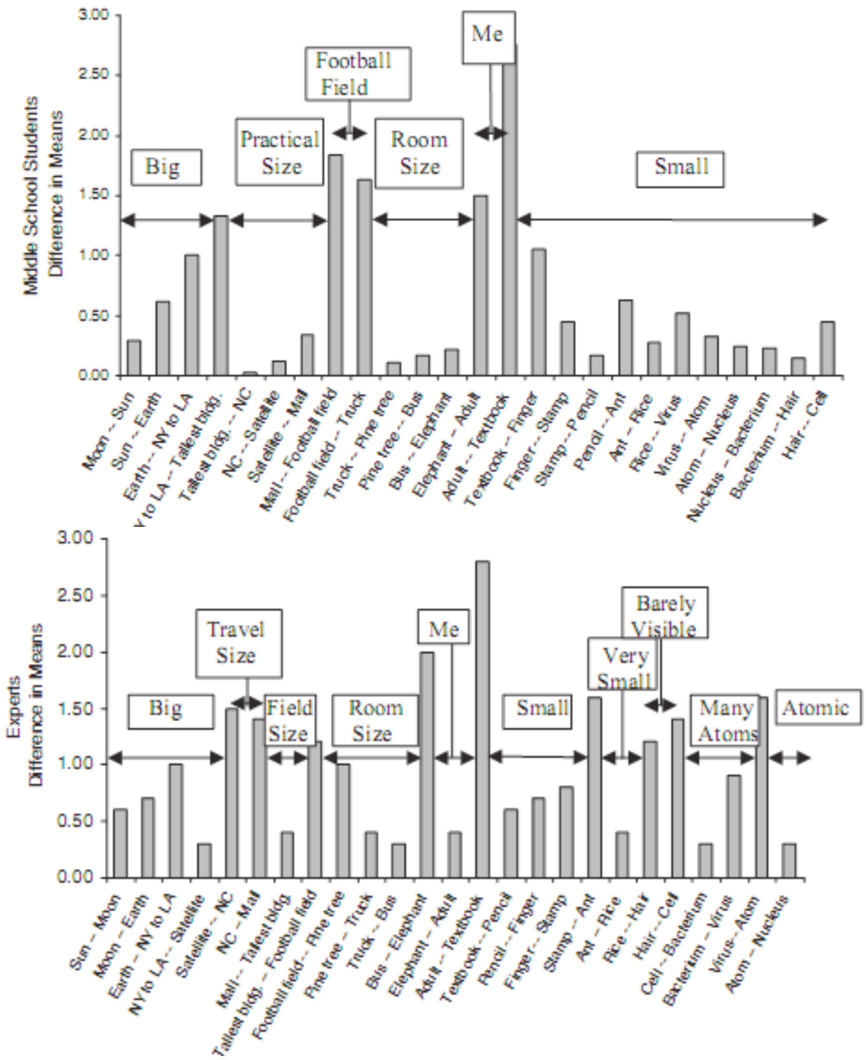


Figure 2.2. Conceptions of boundaries between distinctly different object sizes for middle school (upper) and experts (lower).

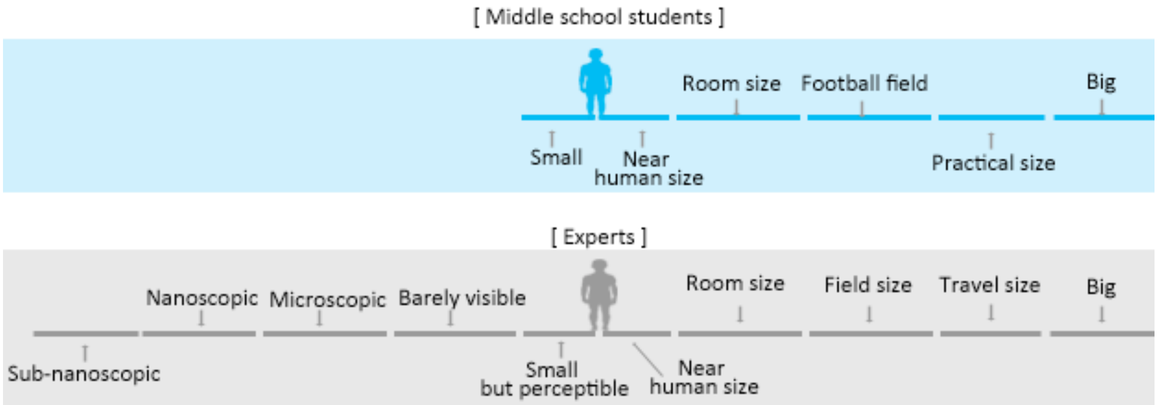


Figure 2.3. A visual representation of the size conceptions of middle school students and experts.

The development of such naïve conceptions in students is mainly due to the nature of imperceptible objects; they are too small to see. This means that the input from the referent is missing (see Figure 2.4), in contrast to normal perceptible phenomena (see Figure 2.1). The first challenge learners face is that they cannot form visual mental images of imperceptible sizes because they cannot have visual experiences of imperceptible objects; needless to say that there does not exist available experience in other modalities either. Moreover, learners have to depend on representations, which would not be easy to accurately understand for them, because of the absence of the referent and their poor preexisting knowledge. Rather, their mental images are prone to be molded from compounds of macroscopic visual depictions of imperceptible objects (Tretter, Jones, Andre, et al., 2006; Tretter, Jones, & Minogue, 2006), which contradicts the definition of "imperceptibleness." Viewing these big-enough-to-see visual representations of imperceptible objects is liable to inadvertently lead students to construct inaccurate mental models of the sizes of imperceptible objects. Without observable connections between a representation and its referent, the representation can become obscure and misleading to learners.

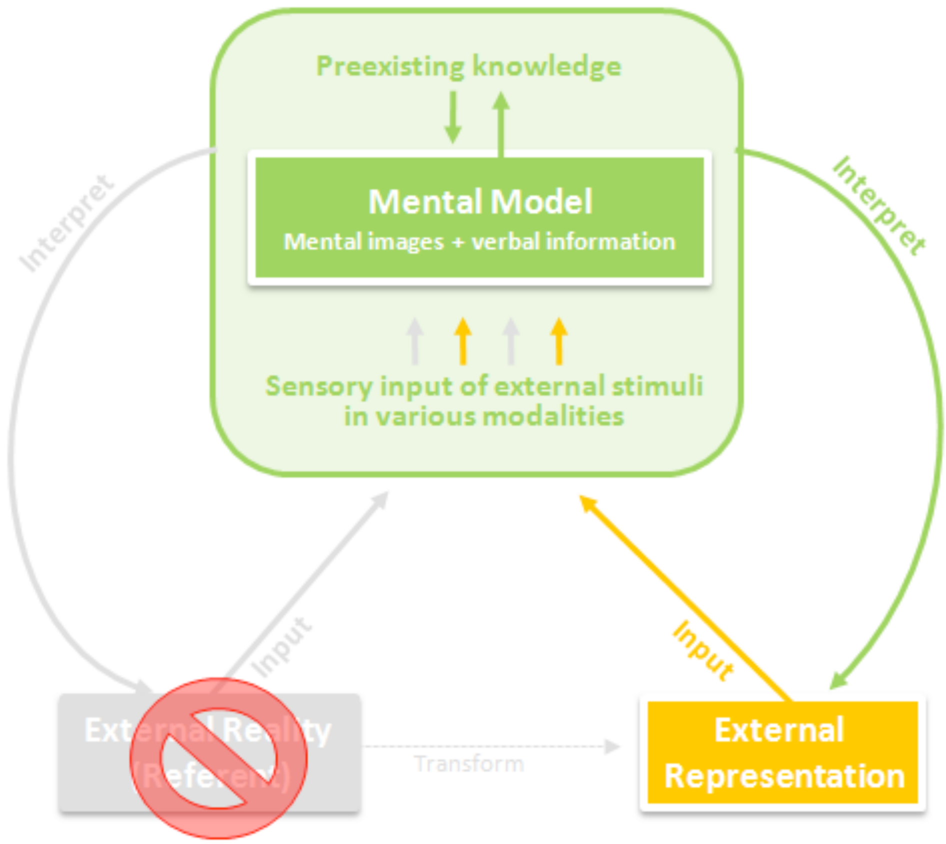


Figure 2.4. The iterative process of mental model formation for imperceptible phenomena, which lacks the input from 'external reality.'
Compare it with Figure 2.1.

Additionally, middle school students do not have well-developed conceptual categories of imperceptible sizes, while the role of the conceptual size categories becomes even more important with the absence of visual mental images. Misconceptions become even more difficult to adjust when students lack an alternative conceptual category to shift concepts into (Chi & Roscoe, 2002). Considering that category learning occurs mainly through interacting with phenomena in the world (Lakoff, 1987; Rosch & Mervis, 1975), it would be highly difficult for a learner to construct a set of conceptual categories for imperceptible sizes.

The limits in learners' proportional reasoning capacity also contribute to the construction of their naïve conceptual size categories. A strong foundation in proportional reasoning ability is considered to be a requirement for developing unitizing skills because the process of converting units involves proportional reasoning (Tretter, Jones, Andre, et al., 2006). However, conducting mental computations with a number with many digits (e.g., imagining the size of a hydrogen atom that is 10,000,000 smaller than one millimeter) is beyond their cognitive capacity (Tourniaire & Pulos, 1985).

The student misconceptions and challenges that were discussed here are created mainly because of the absence of the referent. Due to this problem, learners are not able to develop an appropriate mental model of imperceptible sizes that they can start constructing new knowledge upon. Consequently, students misinterpret macroscopic visual representations, and their naïve conceptions of imperceptible sizes continue to be naïve or incomplete. Meanwhile, it has to be noted that learners still have to depend on representations in order to perceive and conceptualize imperceptible sizes because of the absence of referents. This statement, hence, places a critical emphasis on the design of representations of imperceptible sizes.

Representations for Learning

The Roles of Representations in Learning

Vygotsky (1978) saw tools as a means to understand the world. He discussed a theory that interaction with tools enhances and alters human development because human's activities are "mediated" by tools of different types, such as "material tools" and "psychological tools." Material tools (e.g., pencil, ruler, and computer) are developed within cultures to enable people in order to accomplish tasks. Psychological tools (e.g., representations, language, mathematics, and chemical symbols) are socially created for the communication of thoughts with others. The role of these tools is important in education because the tools that learners use mediate what they perceive and conceptualize.

In more recent theoretical work, Salomon, Perkins, and Globerson (1991) discussed the “intellectual partnership” between a learner and a tool that activates higher order thinking skills. Hutchins (1995) also argued that cognitive processes may be distributed in the environment, in the way that the operation of the cognitive system involves coordination between internal and external information structures. Pea (1993) used the term “distributed intelligence” to explain the cognitive activities that are socially constructed via the interaction between people and the tools in their environment. He explained that human minds and tools reciprocally develop through the interactions. These scholars commonly suggest that humans’ minds are greatly influenced by the tools they use in order to perform cognitive tasks.

Among many forms of tools, representations have made critical contributions to the development and education of science. Representations are indispensable components for achieving advanced learning. Both scientists and students often utilize representations in their discourse of scientific investigation (Kozma et al., 2000). In particular, visual representations are the most commonly used type of representations in science education. Educational researchers have devoted efforts to the implementation and enhancement of visualization tools for science education because of their important role of supporting learners’ perceiving, understanding and problem-solving in many sciences (Stieff, Bateman, & Uttal, 2005). Visualizations convey information that is not directly observable, or is impossible to see, through symbolic cues, such as color, icons, or signs, in order to help students identify which elements are the keys to comprehending the underlying scientific issues (Tversky, Zacks, & Lee, 2004). Examples of scientific visual representations include dynamic multimedia that demonstrate the physics principles of how a pendulum swings, or gravitational force, or animated explanations that use vocal narrations to describe how a pump works, or direct manipulable three-dimensional graphics of the human brain that illustrate its subparts. They occasionally act as simulations that students can review, manipulate to test hypotheses, and potentially use to solve problem sets (Taylor, Renshaw, & Jensen, 1997).

Representations for Learning Imperceptible Sizes and the Issues

To support learners in perceiving and conceptualizing the range of imperceptible sizes, a number of learning technologies that adopted different types of imperceptible size representations, have been designed. For example, the measurement units such as micrometer[3] or nanometer[4] are used to express the absolute size of an object (i.e., "the diameter of a hydrogen atom is about 0.1 nanometers") of imperceptible objects, or, the relative sizes, such as "the diameter of a DNA helix is about 10,000,000,000 times smaller than one millimeter," were used in learning technologies, in association with the images of imperceptible objects. A logarithmic scale, in particular, aims to convey the range of different sizes. They are also often incorporated with the absolute sizes of

---

[3] One micrometer is one millionth of one meter (one thousandth of one centimeter).

[4] One nanometer is one thousandth of one micrometer.

imperceptible objects. Visual representations are the most commonly used type of representations in learning technologies. Macroscopic graphics of imperceptible objects are usually consecutively aligned for size comparisons, automatically zoomed-in on in a video (e.g., *Powers of Ten*), or interactively manipulated (e.g., *Cell Size and Scale, Nikon Universcale*, or *Scale Ladder*), for example, in a click-to-zoom fashion or slide-the-bar-to-zoom interface. These visual representations are frequently coupled with the relative sizes of the objects and require students to mentally visualize the sizes of the objects through proportional reasoning.

Types of representations, whether incorporated in a learning technology or not, influence students' learning, especially when learners have to depend only on those representations in order to comprehend imperceptible phenomena. There always is a possibility that learners do not benefit from all representations because representations may impose constraints and challenges on them. Domain-specific representations are uniformly understood among the experts of the domain, but are likely to be incorrectly or meaninglessly interpreted by learners because their preexisting knowledge is very poor and cognitive capabilities are limited. Although scientific representations are ubiquitous in classrooms, the factors that contribute to their popularity and implementation remain separate from evidence-based examinations of actual learning outcomes (Rapp, 2005). The representations introduced above are likely to burden students with cognitive challenges. For example, the absolute and relative size notations require a student to mentally visualize the size of an object that is one billion times smaller than one meter, through proportional reasoning, which is beyond a human's cognitive capability (Tourniaire & Pulos, 1985). In this manner, such representations are meaningless to students, and, in consequence, students' naïve conceptions usually remain unrepaired. The logarithmic scale also provides representational features that may bring unexpected learning results because it compresses differences between certain sizes and makes those differences look smaller than they actually are. Furthermore, in most cases, it is accompanied with absolute sizes or relative sizes that many learners have difficulty with interpreting. Learners with underdeveloped mathematical competence would not be able to benefit from it.

Visual representations, whether interactive or non-interactive, are particularly misconception-laden. In order to visualize the sizes of imperceptible objects, learners have to mentally scale the pictures of imperceptible objects in their heads, using the absolute sizes that are provided with each size. This is a highly challenging task for learners. Furthermore, exposure to macroscopic visual representations of imperceptible objects tends to inadvertently lead students to construct inaccurate mental models of the sizes of imperceptible objects. It seems that the visual representations take the role of the referent, which is incorrect, for constructing learners' mental models of imperceptible sizes. For example, in Tretter et al.'s study (2006), some of their participants answered, "*when I picture [microscopic objects] I see the drawing [in textbooks].*'' In an interview I conducted during a pilot study, a few 7th grade students even thought that bacteria and viruses ("*germs*" in their terms) were as big as dust or a particle of fine-grained salt, but they were only invisible because they are transparent, inferring this from what they have observed in an educational video. Some other students thought that the germs share the same size and would be visible under a simple magnifier, based on what they have seen in media such as a TV commercial advertising disinfectant sprays.

These observations imply that learners' misconceptions are likely to be developed under the significant influence of the big-enough-to-see images of imperceptible objects, implying that visual representations may not always be useful when one can't have direct visual experience of the target object. Visual representations are very effective for illustrating the shape, features, movement of or physical relationships between certain objects, as in the way they are frequently and effectively used to illustrate many scientific concepts. However, when it comes to the matter of representing imperceptible sizes, they tell learners the opposite of "too small to see" because the visual representations are big enough to see.

When representations stand between the target knowledge and the learners, they become intermediate hurdles that require the learners to be proficient in both interpreting the representations and constructing the knowledge, adding to the overall cognitive complexity for them. The more premises and premise-based inconsistencies, the more alternative mental models are constructed in learners. This makes the cognitive task more difficult (Rapp, 2005). There exist too many cognitive burdens that learners must overcome to conceptualize an imperceptible size using a macroscopic depiction of it. They must accept the macroscopic images in order to comprehend imperceptible sizes, but at the same time, they must cognitively resist the images in order to keep in mind that the depicted objects are not in fact seeable to human eyes. Then, furthermore, they must try to imagine how actually small the objects are through mental calculation, of which they are not capable.

Although diverse representations have been developed and used to teach imperceptible sizes, there still remains a need for a representation that tells learners, “This is what too small to see means, and this is how small they are” via non-visual modality and with fewer cognitive burdens. As diSessa (2004) pointed out, the quality of a representation should always be judged by its purpose. Conventional representations that depend heavily on visual modality are effectively used among scientists, and they have greatly contributed to the advancement of science. However, while students must learn and acknowledge that scientists highly value such representations, students must be able to benefit from representations that are designed for their cognitive capability and specific learning goals.

Multimodal Representations for Learning Imperceptible Sizes

The Roles of Multimodal Representations in Learning

Advances in technologies have enabled the design of representations that are composed of two or more modalities. There is a growing recognition, in science education and research, of the critical importance of understanding and integrating different representational modalities in learning science concepts and methods. Researchers (Ainsworth, 1999; Lemke, 2004; Moreno & Mayer, 2007) discuss their findings that, by providing different modalities (such as sight, sound, touch, smell, self-motion, and taste) that serve different needs of cognition and learning, students’ scientific inquiry and reasoning can be enhanced.

Multimodal representations particularly have potential for science education, where multimodal observations of scientific phenomena are frequently conducted by learners. The most commonly used multimodal representations are ones that incorporate visual and aural modalities. It has been suggested by researchers that animations that are combined with synchronous verbal narration can better support learning than the animation alone (e.g., Hasler, Kersten, & Sweller, 2007; Mayer & Anderson, 1991; Mousavi, Low, & Sweller, 1995; Nugent, 1982; Tindall-Ford, Chandler, & Sweller, 1997). Visual modality is also often combined with other modalities than narration or sound. For example, visual and haptic modalities have been incorporated to convey the phases of molecular bonding process (Chang, Quintana, & Krajcik, 2010). Sonification has been used in combination with visual representations to provide learners with dynamic computer simulations of an ecology microworld (see Pfeiffer, 2008), or to represent the electron probability amplitude of an atom (see Kuchera-Morin, 2010).

There also exist multimodal representations that employ less commonly used modalities. For example, a haptic interface that involves force and kinesthetic modality has been used to augment simulated physics principles (see Han & Black, 2011) and molecular interactions (See Gillet, Sanner, Stoffler, Goodsell, & Olson, 2004; Schonborn, Bivall, & Tibell, 2011). Touch and kinesthetic feedback have been combined with visual representations in order to support learners in conceptualizing virus morphology and the diversity of virus types (M. G. Jones, Minogue, Tretter, Negishi, & Taylor, 2006). Olfactory representations have additionally been used where traditional auditory and visual warnings do not function well (i.e., to warn miners of danger).

The literature suggests the benefits of using multimodal representations. The first benefit is the varying computational processes that multimodal representations provide for learners' comprehension of representations and target information. A representation that incorporates different modalities in order to convey target information can offer different inference processes for learners (Ainsworth, 1999; Oviatt et al., 2000; Rapp & Kurby, 2008; Sweller, 2005b; Zhang & Norman, 1994). Mental images that were created through different modal interactions, but with coherent information, can become associated with each other and provide meaningful learning experiences for learners by complementing the weaknesses of one modality with the strengths of another.

Through this “mutual disambiguation,” modalities are combined collectively in order to achieve a greater level of expressiveness (Oviatt & Cohen, 2000). For example, the robustness of speech recognition has been improved through mutual disambiguation via input from other modalities (Oviatt et al., 2000). In the study that studied the use of pictorial and verbal tools for conveying routes, Lee and Tverskty (1999) concluded that the existence of parallel depictions and descriptions for routes does not mean that both are equally effective in all situations. In many cases, a combination of the two was the most effective in portraying routes; these cases were able to simultaneously utilize the advantages of both methods. The descriptions were more appropriate for abstract information (“turn right”), while the depictions were more appropriate for information that is directly or metaphorically visualizable (“big oak tree”). Consistently, it was observed that middle school students’ achievement in science improved when they comprehended that no single modal representation is able to encompass the entire concept (diSessa, 2004). The students tried to include multimodal representations that gave minimal but sufficient information, were defined well, and were comprehensive to their rhetorical purpose.

The second benefit of using multimodal representations is their potential for creating natural cognitive mappings between a representation and its referent. As discussed earlier, a learner has a mental model of how to interpret a representation, and the representation also yields a conceptual model for how it can be interpreted. A cognitive mapping may create conceptual correspondences between the elements of the representation and the referent (Downs & Stea, 1973). When two models coincide, then there is a natural cognitive mapping between the representation and the mental model. The nature of human cognition is multimodal (Baddeley, 1986; Kosslyn, 1994; Paivio, 1986), and mental images mimic corresponding events in the world (Kosslyn, 2005). Multimodal representations can allow a learner to form mental images that are more natural and close to the referent in the real world than a uni-modal representation that only provides one mode of information, which, in many cases, could be in a different modality of the referent.

However, it must be highlighted that, although natural cognitive interaction with the world is multimodal most of the time, a multimodal representation is not necessarily a natural representation. The naturalness of the representation or the interface is not necessarily what makes a tool intuitive and effective (Anastopoulou, Sharples, & Baber, 2011). It is the representation designer's task to find the right balance between making a representation natural and making it uncomplicated for learners. In certain cases, natural interactivity might in fact be less suited for a particular interaction. The desire to make a system naturally interactive must not take precedence over allowing a learner to understand the learning goal.

The third benefit of using multimodal representation is its ability to increase motivation in learners (Barak, Ashkar, & Dori, 2011; Domagk, Schwartz, & Plass, 2010; Miller, Chang, Wang, Beier, & Klisch, 2011). It appears that information presented via multimodal media tends to be more stimulating than information presented in traditional materials. For example, a study that asked students to comprehend key concepts by using different representations found that the students were more motivated to learn with the multimodal materials than with the traditional uni-modal ones (Waldrip, Prain, & Carolan, 2010). Also, in addition to promoting students' comprehension and ability to narrate scientific concepts, the use of animations resulted the development of higher motivation to learn science, in terms of self-efficacy, interest, enjoyment, connection to daily life, and importance to their future, when compared to control group (Barak et al., 2011).

Based on these benefits of multimodal representation, I propose to design and use a multimodal representation to support learners in better conceptualizing the range of imperceptible sizes. The absence of accurate visual mental images, which is the main challenge for learners, could be complemented by other non-visual experiences. In this way, the learning experience regarding imperceptible sizes could be more "natural," because what the multimodal representation conveys will be consistent with the meaning of "too small to see." With the additional possibility of more active student engagement in the learning activity with the multimodal representation, the alternative multimodal representation could be effective.

Multimodal Representations for Learning Imperceptible Sizes and the Issues

It is known that a representation that directs learners to explore concepts in a novel way may help them realize and revise their misconceptions by revealing the inconsistencies between their mental models and the representation (Ainsworth, 1999; Chi, 2005; Hynd & Guzzetti, 1998). Previously held beliefs in learners are particularly resistant to change, but this resistance can be lessened if a learner is provided with an explanation that show why the flawed beliefs are incorrect. Based on this argument, I believe that a multimodal representation, which does not employ a visual modality as its main vehicle, may provide a natural perception and cognition process for learners, and it may consequently support the construction of more appropriate mental models of imperceptible sizes.

However, although the benefits of multimodal representations are becoming more widely acknowledged, there exists evidence that inappropriately designed multimodal representations can bring disadvantages in learning (Rapp, 2005, p. 53). Since learning scientific knowledge entails conceptually linking multimodal representations and scientific phenomena (Ainsworth, 1999, 2006; Lemke, 2004), merely exposing learners to rich multimodal representations does not automatically guarantee deep comprehension and learning. Some researchers point out that the structure of a multimodal representation can be too complex for learners (de Jong, 2010; R. E. Mayer & Moreno, 2003; Paas, Renkl, & Sweller, 2003). Oviatt (1999) emphasized the need for more research on how different modalities are combined and organized, based on her findings that indicated the existence of significant variability in how individuals integrate multimodal information.

Without proper designs, students are less likely to know what to attend to and what is being conveyed by a representation because learners, who possess weak representational literacy, tend to focus on the surface features of a representation (Kozma et al., 2000). As noted by Ainsworth (2006) and others, design researchers are beginning to struggle with many issues, such as the number, modal type, style, and sequence of representations that can maximize learning outcomes. Examining whether particular concepts are better matched to particular representational modes, and investigating the conditions when redundant information, in and across representations, enhances learning are issues shared among representation researchers (Ainsworth, 2006). Knowing how knowledge is represented in our mental models would enable a researcher to attempt to design appropriate multimodal representations (Rapp & Kurby, 2008).

In this chapter, I discussed the components of the mental model of sizes and, in relation to them, the challenges for learners in conceptualizing the range of imperceptible sizes. The challenges were mainly due to the impossibility of direct visual experiences, and the learning tools – representations – being confusing to learners. Then I emphasized the critical role of representations in the context of teaching and learning about imperceptible phenomena, and argued for the need for an alternative representation that does not employ a macroscopic depiction as the main vehicle for conveying imperceptible sizes. Subsequently, I highlighted the potential benefits of using a multimodal representation for supporting learners in better conceptualizing imperceptible phenomena. However, in research on representations for science education, the existing studies remain inconclusive regarding the question of how to design a non-visual representation that is more natural for learners for conceptualizing the range of imperceptible sizes. Moreover, the alternative modalities of the representation and their influence on students' mental models have not been explored either.

In the prior section, I discussed the components of the mental model of sizes. They are visual mental images, conceptual size categories, temporal-kinesthetic experience, and propositional relationships between objects. Since an imperceptibly small size cannot be represented visually, it may be more effective to design and use a representation that may help students better conceptualize the conceptual size categories. A representation that does not employ macroscopic visualization of an imperceptible object as a main vehicle for conveying size and that uses a modality, for example, temporal, may allow learners to easily perceive and make connections with the size.

In the following chapter, I introduce a temporal-aural-visual representation that was designed to support learners in better conceptualizing imperceptible sizes, along with findings from the pilot studies.

CHAPTER 3 TEMPORAL-AURAL-VISUAL REPRESENTATION (TAVR)

In Chapter 2, I explained that the mental model of sizes can be composed of verbal information (i.e., conceptual size categories) and non-verbal information (mental images that are either visual or temporal-kinesthetic). I have also argued that big-enough-to-see visual representations cannot effectively convey sizes that are too small to see; rather, they promote the creation of misconceptions in learners, concluding that a representation that does not employ visual modality as the main modality for conveying imperceptible sizes may be useful for supporting learners to better perceive and conceptualize the range of imperceptible sizes. On this basis, in this chapter, I propose a temporal modality as a main modality for representing imperceptible sizes. I introduce the design and rationale of a temporal-aural-visual representation (TAVR) and a Flash-based web application where students can interact with TAVRs and reflect on their learning experience. Then I present a summary of the findings from the pilot studies.

TAVR

In TAVR, a temporal representation is incorporated with visual and aural modalities, which are adopted to augment learners' temporal experience of imperceptible sizes. One object is put on the head of a pin every 0.1 seconds until the accumulation fully spans the diameter of the pinhead (1 millimeter). Hence, for example, when TAVR lines up strands of human hair, of which the thickness is about 100 micrometers, it may take 10 strands of hair to fully span the head of a pin. In this way, the size of an imperceptible object is represented by the inverse relationship between the size of the object and the duration of sequential accumulation. Therefore, the smaller the object, the longer the accumulation duration.

When one object is placed, an audio clip that sounds like a single click is played once (aural). See Figure 3.1 for an illustration of how the accumulation occurs in TAVR. The accumulation that is happening within the imperceptible scale is indicated via sound, and a red line (visual) appears when the mass of accumulated objects becomes macroscopic in scale (see Figure 3.2 for an example). The aural and the visual representations are the modalities used to convey the accumulation of objects on the pinhead. Because of the problem tied to the macroscopic depictions of imperceptible objects, visual representations are added only when the accumulation enters the macroscopic scale. Thus, the accumulation of objects within the imperceptible scale is represented via the sound.

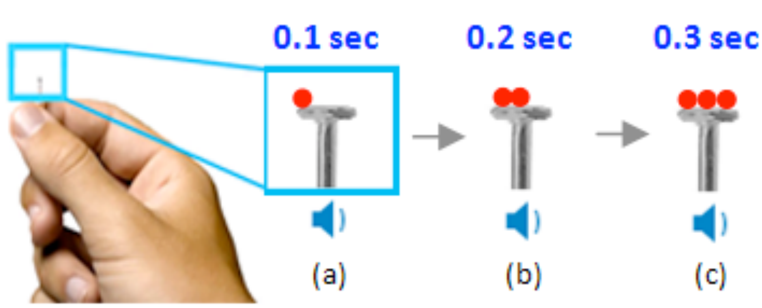
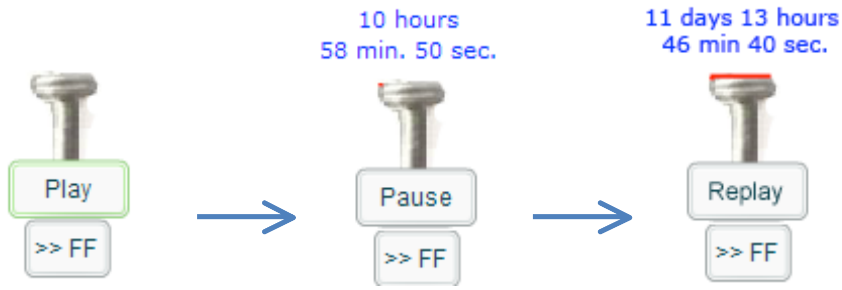


Figure 3.1. An illustration of the accumulation process in TAVR.
(a) The first object is placed on the pinhead and one click is played. (b) Then the second object is placed on the pinhead next to the first one, and one click is played. (c) The same process is repeated until the objects span the pinhead.



(a) (b) (c)

Figure 3.2. Screen captures of the TAVR of Hydrogen atoms at three different phases : (a) Default phase, (b) Interim phase. Note that the short red line appeared after about 11 hours. Until this moment, the accumulation was only indicated by the clicks. (c) The accumulation is completed. Students can fast forward the accumulation by clicking >>FF button below the Play button.

While interacting with TAVR, students are expected to observe the following events:

(1) The interval between object placements, which is the term between two clicks.

(2) The duration of the interval from the beginning of the accumulation to the moment when the accumulation becomes macroscopic (when the red line finally becomes visible after the sound-only period of accumulation).

(3) The duration of the interval from the first appearance of the red line to the completion of the accumulation.

To support learners in better focusing on observing the accumulation (i.e., to witness the moment when the accumulation enters the macroscopic scale), rather than keep watching a clock to measure the elapsed time, each TAVR has an embedded clock that shows the elapsed time (e.g., “1 hour 30 minutes”). Reading the embedded clock, students are expected to make active interpretations of the units of time (e.g., seconds, minutes, hours, days) in relation to the represented sizes of the objects. I anticipate that

TAVRs may provide a set of metaphorical conceptual size categories to learners. A conceptual metaphor is created by analogically extending the conceptual structure from richer, experience-based domains (i.e., units of time) to structure learners' understanding of relatively more abstract domains (i.e., imperceptible sizes) (Boroditsky, 1997; Boroditsky & Ramscar, 2002; Lakoff & Johnson, 1980b). From the experience of daily life, by the age of a middle school student, a person usually knows the different units of time (Acredolo, 1989). Therefore, it may be easier for them to understand the difference between the sizes represented in units of time (e.g., "it takes about 13 hours to span Rhinovirus across the head of a pin") than in notions of relative size (e.g., "the diameter of a Rhinovirus is about 40,000 times smaller than one millimeter"), or in measurement units (e.g., "a Rhinovirus is about 20 nanometers"). When the sizes are represented in units of time, students are not required to mentally calculate the numbers to get the sense of relative differences between two different sizes and overload their cognitive capacity. Rather, they simply need to refer to the biggest unit of time taken for the accumulation in TAVR because the units of time are more familiar to them.

Among different types of non-visual modalities (e.g., tactile) that might be useful for representing the accumulation in the imperceptible scale, I chose sound to represent imperceptible sizes based on the way our working memory processes two different channels of modality (Paivio, 1986), which proposes that information is processed through two separate but parallel channels - visual and auditory. According to this theory, effective learning happens when the information in two channels matches and interacts closely. Through this process, a learner would be able to integrate the mental representations from two channels.

The way the sound is used in TAVR is like sonification, which is the use of non-speech audio to convey information (Kramer et al., 1997). Sonification is considered effective for capturing temporally complex information, which involves the interactions of different types of information (Pfeiffer, 2008). In TAVR, the aural representation, as a sonification feature, delivers two different types of information to learners. The first is the information that the accumulation is in progress, and the second is the duration of the accumulation. Although learners cannot accurately perceive or infer the total accumulation durations, the aural representation also conveys an abstract sense of the passage of time.

Each TAVR is also accompanied by a ">>Skip ahead" button below a "Play" button (see Figure 3.2). This is added to allow students to literally skip ahead through the accumulation of a number of objects, depending on my decision about how many objects students may skip. This feature is designed with two different purposes. The first is to add kinesthetic experience to the temporal experience of imperceptible sizes. As discussed in Chapter 2, the concept of time is highly associated with kinesthetic experience, and experiencing redundant information in different modalities may result in more saturated mental model construction. The second purpose is to make the learning activity with TAVRs (using WIIS) practically feasible in classrooms. Considering that a science class in middle schools is normally fifty minutes long and, due to the logistics (e.g., attendance check, homework collection, seat assignment in a computer lab, introducing the learning activity for the day) that have to be taken care of for a class, there would only be 30 – 40 minutes allowed for the learning activity. Since four out of six objects that are used in WIIS have accumulation times longer than one hour, and the longest duration is nearly twelve days, students would not be able to observe all TAVRs in one class if the accumulation durations were longer than forty minutes. Moreover, almost no student would be able to endure the duration of “nothing is happening but the sound” phase any longer than five minutes. However, by pressing the “skip ahead” button, I believe students may be more engaged in the TAVR experience because they do not have to just sit and watch; rather, they actively manipulate their temporal experiences.

One may ask why I made the time and size inversely related because sequential temporal experiences, which are commonly used by people for conceptualizing long distances, usually are non-inverse in relationship (i.e., the longer the travel time, the longer the distance). When people conceptualize distinctive distances, the sizes that are similar to a human, such as a car, a bicycle, or a bus, or the segments of sequential kinesthetic movement, such as walking or bicycling, become the “size reference” that enables people to infer the sizes of the sizes too big to see. However, for sequential temporal experience with sizes that are too small to see, we cannot use an atom, for instance, for a size reference, because learners initially do not know the size of an atom. Learning happens when a learner can construct new knowledge based on their preexisting knowledge (Piaget et al., 1977). In order to help learners make connections with a size they already know, I decided to use the head of a pin as the size reference, although the relationship between the size and time must become inverse. Metaphorically speaking, using a non-inverse relationship without a familiar size reference in order to explain an imperceptible size would be like trying to explain the size of an alien planet by saying, “it is as big as one million of the smallest pebbles on that planet.” Among other small macroscopic objects (e.g., a grain of rice or an ant) that could be used as a reference size, I believe that the head of a pin makes a good size reference because the diameter of a pinhead is one millimeter, which can be convenient when learners have to make connections with measurement units (e.g., “one millimeter is one thousand times bigger than one micrometer”) or other advanced mathematical calculations.

Students’ Sense Making of Their TAVR Experience

While interacting with TAVRs of different imperceptible objects, students are expected to compare the total durations of the accumulations from the start to the completion of those accumulations. This concept of time, in other words, is the awareness of time passage. Although it is not explicitly perceivable through human sensing organs, such as eyes or ears, the literature indicates that the sense of time passage is a modality that we indeed perceive and conceptualize. Research that investigated the sense of time passage shows that people can perceive the passage of time and can approximately measure how much time it has passed using only their sense of time’s passage, without looking at clocks or other clues (e.g., the movement of the sun) (K. C. Friedman, 1944). Referring to such empirical evidence, Evans (2003) argued that our conception of temporality may ultimately be traceable to neurologically instantiated “temporal codes” underlying perceptual processing.

Studies indicate that temporal experience may be encoded in visuo-spatial mental images and verbal information in our memory. Many scholars (e.g., Belardinelli & Di Matteo, 2002; Frost, 2001; Kosslyn, Behrmann, & Jeannerod, 1995; Palmiero et al., 2009; Pylyshyn, 2002) argue that our modal experiences can be turned into mental images that are visuo-spatial, aural, olfactory, tactile, gustatory, somatic, and kinesthetic. They also state that the experiences in various modalities actually are turned into such mental images. However, no one has included “temporal mental images” in their considerations.

Hence, I hypothesize that the temporal experience may be encoded into verbal information, closely linked with visuo-spatial or kinesthetic mental images. I make this hypothesis based on the research that claims that it is virtually impossible to talk about time without invoking motion and spatial content of a temporal experience (Lakoff & Johnson, 1980a), and that temporal experiences are often stored in the form of verbal information in our mental models (Evans, 2003; Lakoff & Johnson, 1999). For example, many different temporal experiences in our daily lives are reconstructed in visuo-spatial mental images in our memory. The image of a clock is visual, but the reading of the clock is verbal. The movement across a certain distance is kinesthetic and (consequently becomes) visuo-spatial, but the duration of the temporal experience is remembered in verbal form (e.g., “It took one hour by car”). Units of time are closely associated with our daily lives, and individuals have their own way of interpreting different passages of time. It has been recognized by theorists that linguistic expressions for time utilize the same linguistic structure for kinesthetic events and for locations in three-dimensional space (e.g., Boroditsky, 1997; Casasanto & Boroditsky, 2008; Friedman, 1979; Gentner, Imai, & Boroditsky, 2002; Goldstone, Boardman, & Lhamon, 1958; Moore, 2006). In their view, motion and spatial concepts metaphorically structure temporal concepts. See Table 3.1 for the examples.

Taken together, I postulate that learners will assemble the information listed below, which will be perceived through their learning activities with TAVRs, in their mental model of imperceptible sizes in order to make sense of the range of imperceptible

Space	Time
at the corner	at noon
From here to there	From two o'clock to four o'clock
Through the tunnel	Through the night
He stood before the house	It happened before evening
He was running ahead of me	He arrived ahead of me

Table 3.1. Space-time correspondences in language, excerpted from Gentneret al. (2002)

sizes. To briefly summarize, I believe TAVR may effectively influence the creation of a refined mental model of imperceptible sizes by providing learners with the potential components of a mental model of imperceptible sizes that include verbal information encoded from the temporal, aural, visual (and kinesthetic) modal input and mental images of aural and visual (and kinesthetic) experiences with TAVR. These processes are graphically represented in Figure 3.3.

- *Temporal*: the concept of the passage of time (the duration of the accumulation), that is re-represented in the units of time, is the main information that makes distinctively different sizes distinguishable from each other. The temporal experience of observing the accumulation would result in the generation of verbal information. The units of time already exist in learner's preexisting knowledge. Learners would interpret the accumulation durations following their own interpretations of the units of time. Depending on the rapidity of the object placement, I assume that student perception and conceptualization of imperceptible sizes could vary, and, hence, the conceptual size categories for the sizes would also vary. For example, for hydrogen atoms, the total duration of accumulation becomes substantially longer if the accumulation interval is set to one object per second (the total accumulation duration becomes 115 days 17 hours 46 minutes 40 seconds) than ten objects per second (the total accumulation duration becomes 11 days 13 hours 46 minutes 40 seconds). In this case, a student who interacted with the hydrogen atom TAVR, with the accumulation interval of one object in one second, may construct a mental model of a smaller sized hydrogen atom than a student who interacted with a hydrogen atom TAVR with ten objects per second rapidity. The smaller the object, the longer a student must wait to see the completion of the accumulation. Then the student may give a smaller conceptual size category tag to the size.

The temporal experiences with TAVRs may also contribute to, in conjunction with the visual input from TAVR (the pinhead), generating a visuo-spatial mental image in a learner's mental model of imperceptible size; a visual mental image that is literally too small to see and, hence, un-seeable. For sub-nanoscopic objects (e.g., atoms and molecules) students have to wait for a long time, first to see the appearance of the red line, then to see the completion of the accumulation. If the wait were longer than the student had expected, he or she would not only have to keep revising the conceptual size category of the object, but would also need to keep the visual image of the pinhead "empty."

- *Aural*: The clicks in TAVR, which is used as the sonified indicator of the accumulation's progress and its duration, are expected to be stored in learners' memories as in the form of an aural mental image. It will inform learners of the interval between object placements when they attempt to recall their temporal experiences in order to infer the approximate number of the accumulated objects. The interval would influence the way students infer the number of objects that were placed on the pinhead. In other words, learners may differently interpret their TAVR experiences when different accumulation speeds are provided. For example, a study on sonification found that, for the tasks that involved greater stimulus complexity (e.g., a greater frequency of tones on the audiotape), the participants estimated greater durations (Ornstein, 1997). Along with an aural mental image (the clicks), the aural experience in TAVR would also lead to the creation of verbal information, such as "the object is being placed very rapidly" or "the sound was played for too long," in association with the temporal experience.

- *Visual*: The visual components in TAVRs are the size reference – the head of a pin – and the red dots that indicate that the accumulation has become macroscopic. The size of an imperceptible object itself is un-seeable. Learners will generate a visual mental image of an imperceptible object that is "too small to see"; in other words, a visual mental image of "nothing visible," on the head of a pin. In this regard, I expect that learners would generate verbal information, such as, "the object is much smaller than the head of a pin" or "the object is too small to see."
- *Kinesthetic* (optional): The kinesthetic experience in TAVR is the repetitive clicking of the ">> Skip ahead" button. It is optional because the repetitive button clicking would normally be applied for nanoscopic (e.g., DNA helix and Rhinovirus) or sub-nanoscopic objects (e.g., hydrogen atoms and water molecules) of which the accumulation duration lasts longer than one class hour. Also, it is up to the students' decision, regarding when they would start pressing the button, if they ever decided to use the button. However, if a student uses the button, the kinesthetic experiences with TAVR will add a kinesthetic mental image, which is the repetitive and labor-intensive button clicking that is sensed through the muscles in the fingers, hand, and arm, as

well as verbal information, such as, "I had to press the button so many times to see the accumulation completed for an atom."

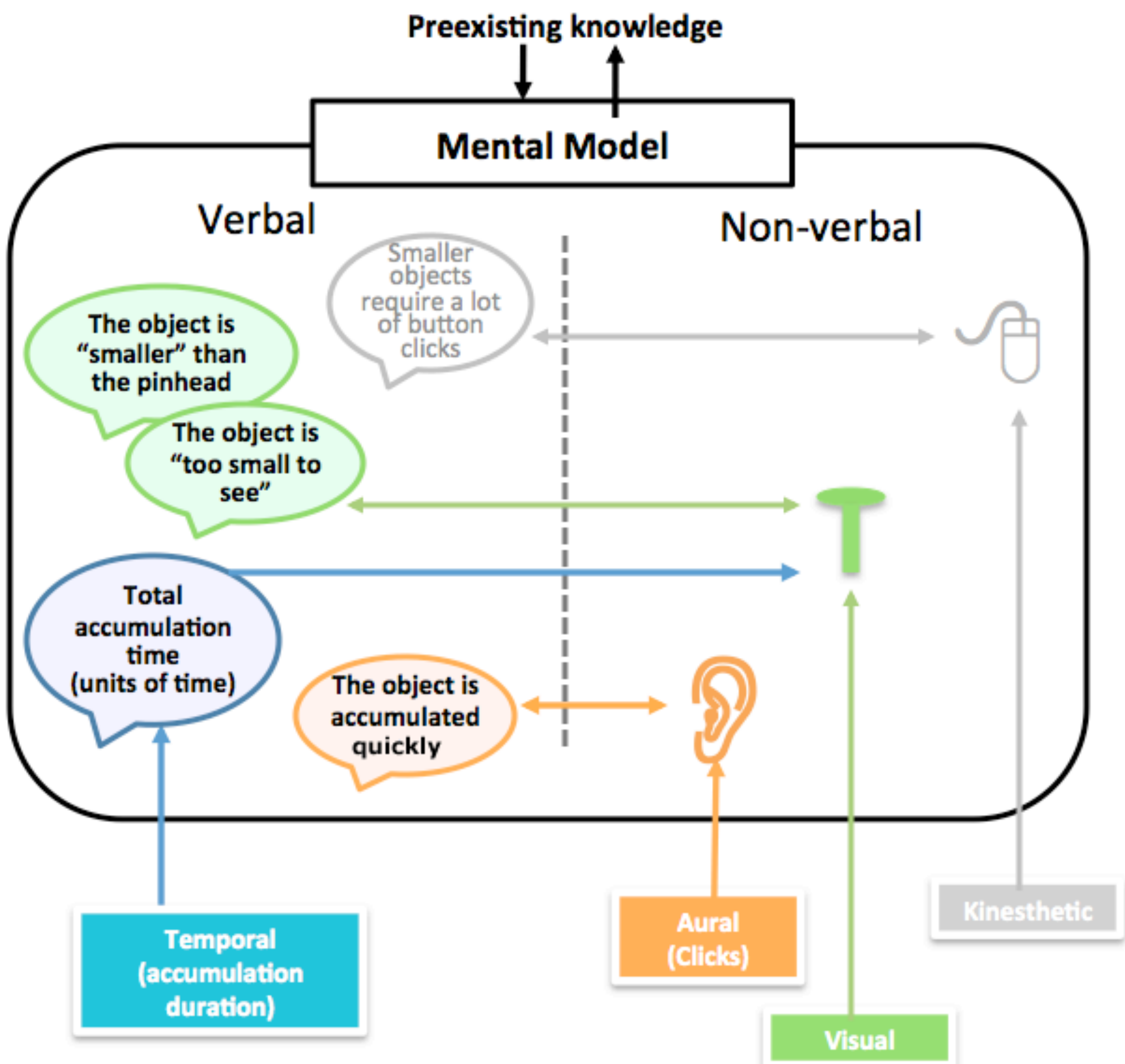


Figure 3.3. A hypothetical mental process model of TAVR interaction.

Application: Wow, It Is Small! (WIIS)

I designed *Wow, It Is Small* (WIIS), a Flash-based application that provides a learning environment where students can interact with TAVRs and keep track of changes in their ideas regarding the sizes of imperceptible objects (including the range of the sizes). The imperceptible objects that are used in WIIS and their durations of accumulation are summarized in Figure 3.4. In WIIS, the biggest units of time of the duration of sound matches with the scale category they belong to. It takes several seconds for microscopic objects, several hours for nanoscopic objects, and several days for sub-nanoscopic objects. To play WIIS (either in the full version or a simplified demo), please visit http://www.wow-it-is-small.info.

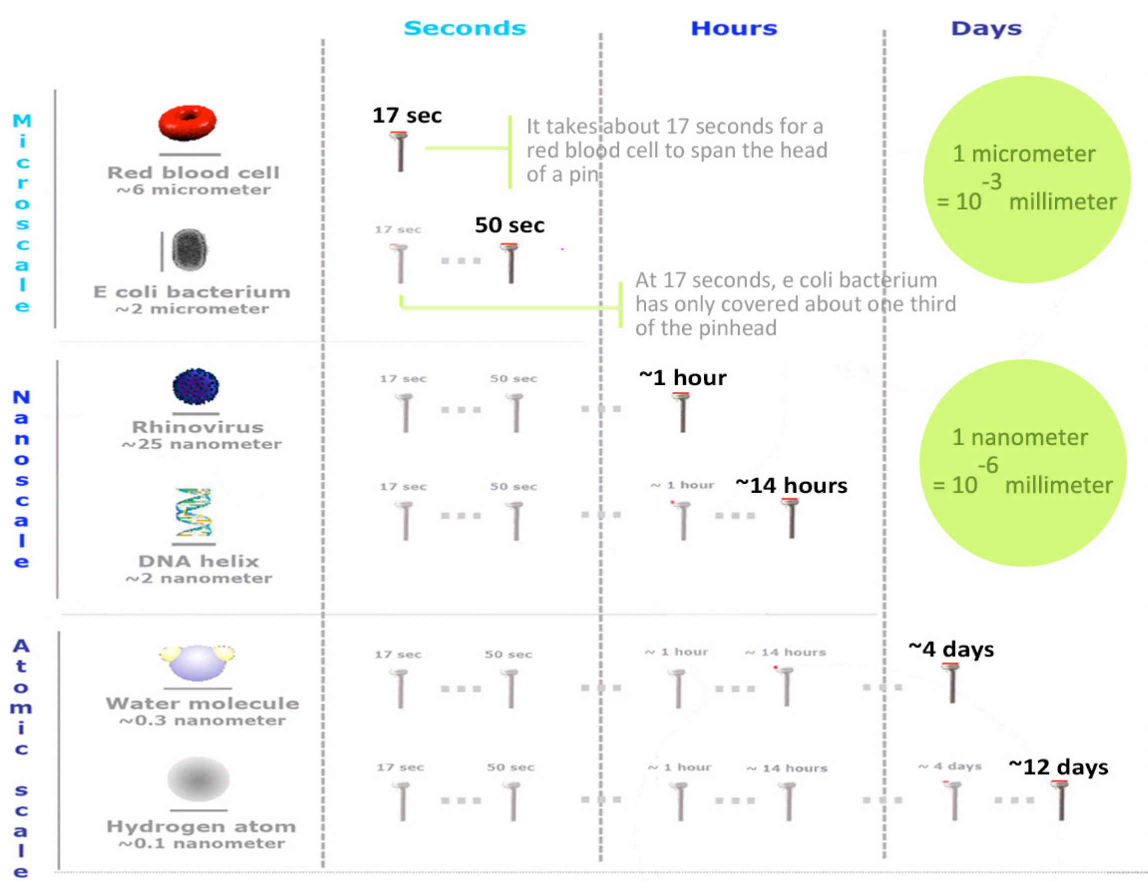


Figure 3.4. The sizes and the durations of accumulation of the imperceptible objects used in WIIS.

According to constructivist learning theory, conceptual changes happen most efficiently when learners are able to take new experiences and integrate them into their preexisting knowledge (Piaget et al., 1977). This implies that internal contradictions instigate the construction or reconstruction of knowledge, and conceptual conflict is necessary for learning. Quintana et al. (2004) called this process of cognitive construction "sense making." The authors stated that the process of sense making involves hypothesis development and collecting, analyzing, and interpreting data to test the hypothesis. They argued that providing learners with conceptual organizers with which students can generate and manipulate their own representations as they progress in learning a target concept is helpful for the learners' sense making. Following their design guideline for learners' sense making, in WIIS, I created a separate space for learners where they can represent their preexisting knowledge, compare their preexisting knowledge and newly learned information, and reflect on the difference between their preexisting knowledge and the new knowledge.

The sequence of learning activities in WIIS is summarized in Table 3.2. The learning activity in WIIS is divided into five different phases: (1) representing preexisting knowledge (step 1 – 2 and 5 in Table 3.2), (2) scaffolding (step 3 – 4), (3) interacting with TAVRs (step 6), (4) posttest (step 7), and (5) reflection (step 8). In the pretest, students complete a card sorting task that is designed to reveal their preexisting knowledge. During the scaffolding activity, students step through pages that are designed to introduce what a TAVR is and how it works, and then they predict the accumulation time of each object. After interacting with the TAVRs, in the posttest phase (step 7), students revise the arrangement of the cards they sorted in the pretest, viewing the accumulation durations of the TAVRs of each imperceptible object. As the final activity, in step 8, students are also asked to reflect on a seven-point bipolar Likert scale that ranges between “extremely smaller than I thought,” “similar with what I thought,” and “extremely bigger than I thought.”

Steps	
1	Students take a preexisting knowledge assessment test
2	Students complete a card sorting task.
3	Introduction to TAVR.
4	Students solve a quiz that aims to examine whether they understood what the representation represents and how it works.
5	Students represent their existing idea of the sizes of the imperceptible objects in the card sorting task.
6	Students explore the sizes of the imperceptible objects with TAVR.
7	Students revise their previous work.
8	Students reflect on the differences between the actual accumulation time and their predictions.

Table 3.2. The phases of the student tasks and collected data.

The card sorting task was composed of three sub-tasks:

(1) *Ordering:* Students order the imperceptible objects by size – from the smallest to the biggest.

(2) *Grouping:* Students then classify the objects by similar size and divide them into several groups. The number of groups could vary from 1 to 6.

(3) *Labeling:* Students give names to the groups they created. Students were provided with prompts (e.g., a set of adjectives – “small,” “tiny,” “mini,” “teeny,” and a set of adverbs – “extremely,” “very,” “strikingly,” “surprisingly”). Students could also use any additional vocabulary they could come up with.

The card sorting task was adopted as a main student assessment task because, as discussed in Chapter 2, conceptual size categories play a critical role in making sense of what learners experience (Chi & Roscoe, 2002; Lakoff, 1987) and constructing a mental model of sizes (Kosslyn, 1980). Having more conceptual category tags means that a person has a more refined mental model of the range of sizes and can make a shift in conceptual categories in order to repair misconceptions (Chi & Roscoe, 2002). Therefore, an increase in the number of the size groups and the descriptive words in the labels in the card sorting task will imply positive effects of TAVRs on students' mental model refinement.

Scenario of a Learning Activity Using WIIS

The instruction is designed to be conducted in a computer lab, and every student is to be assigned to an individual computer. The students access WIIS via web browsers. During the instruction, students are prohibited from going back to the previous learning activities in WIIS to change their answers. They are told to move on to the next learning activity all together at the same time, following the instruction. They are allowed to discuss with peers what the tasks are, how TAVRs work, and how to use the interfaces, but they must generate the answers by themselves. The instructor must make sure that all the students understand the inverse relationship between the accumulation time and the size of an object before the students predict the accumulation time for each imperceptible object. Each student is provided with a real pin for the size reference. During the instruction, students must be continuously reminded of the fact that the accumulation is happening on the head of the real pin in front of them (the pin is planted in a piece of cork for safety). Looking at it during the learning activity will help them be more clearly aware of the fact that the sizes of the objects represented in TAVRs are much smaller than the pinhead.

The following figures (from Figure 3.5 to **Error! Reference source not found.**) are screen captures of each step of the learning activity in WIIS. Each figure is followed by a description of the student activities shown on the screen.

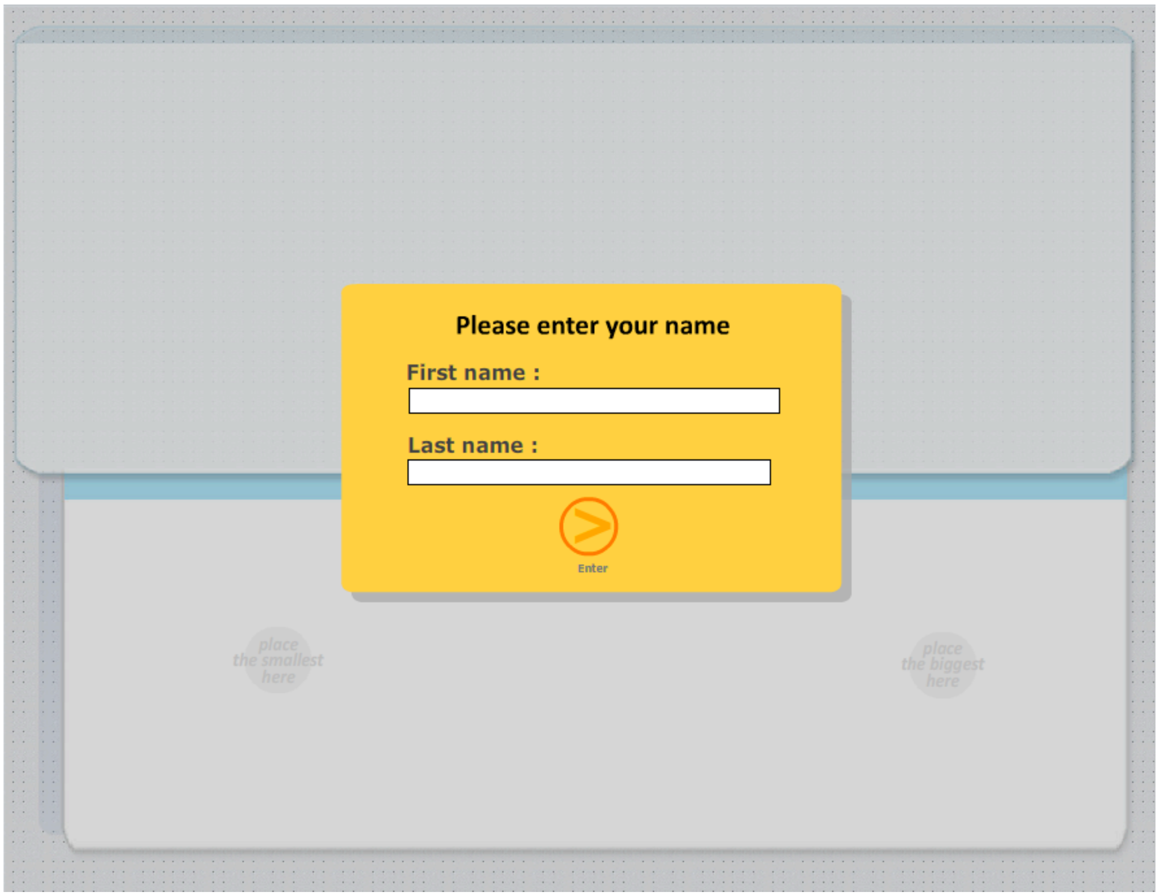


Figure 3.5. The first screen of WIIS.
Students enter their names to log in.

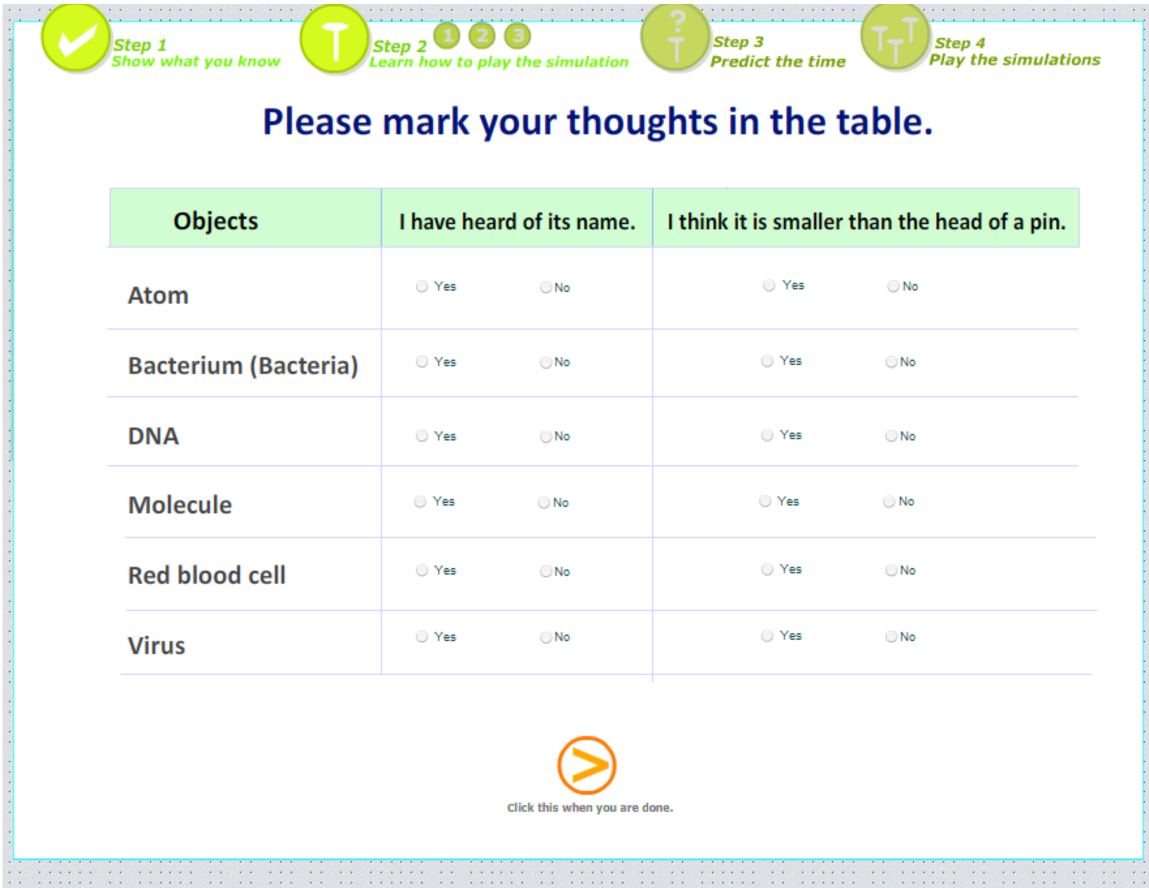


Figure 3.6. Students answer a survey.
It attempts to probe their basic background knowledge – knowing that the objects are imperceptible.

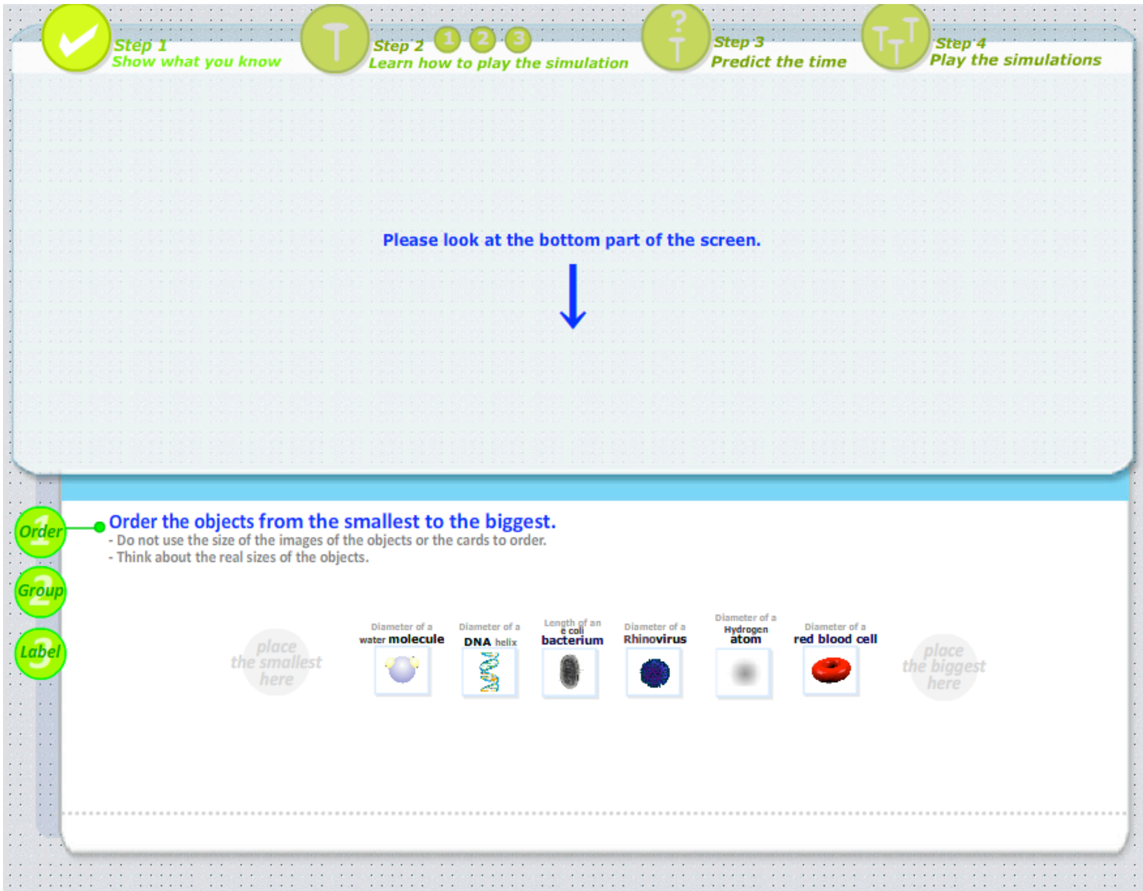


Figure 3.7. Card sorting task 1 – Ordering.

Students work on the card sorting task in the bottom part of the screen (white background) that is designated for representing their ideas. At this step students order the cards from the smallest to the biggest. The buttons in the upper parts (where they will interact with TAVRs later) are disabled so that they can move forward without instructions.

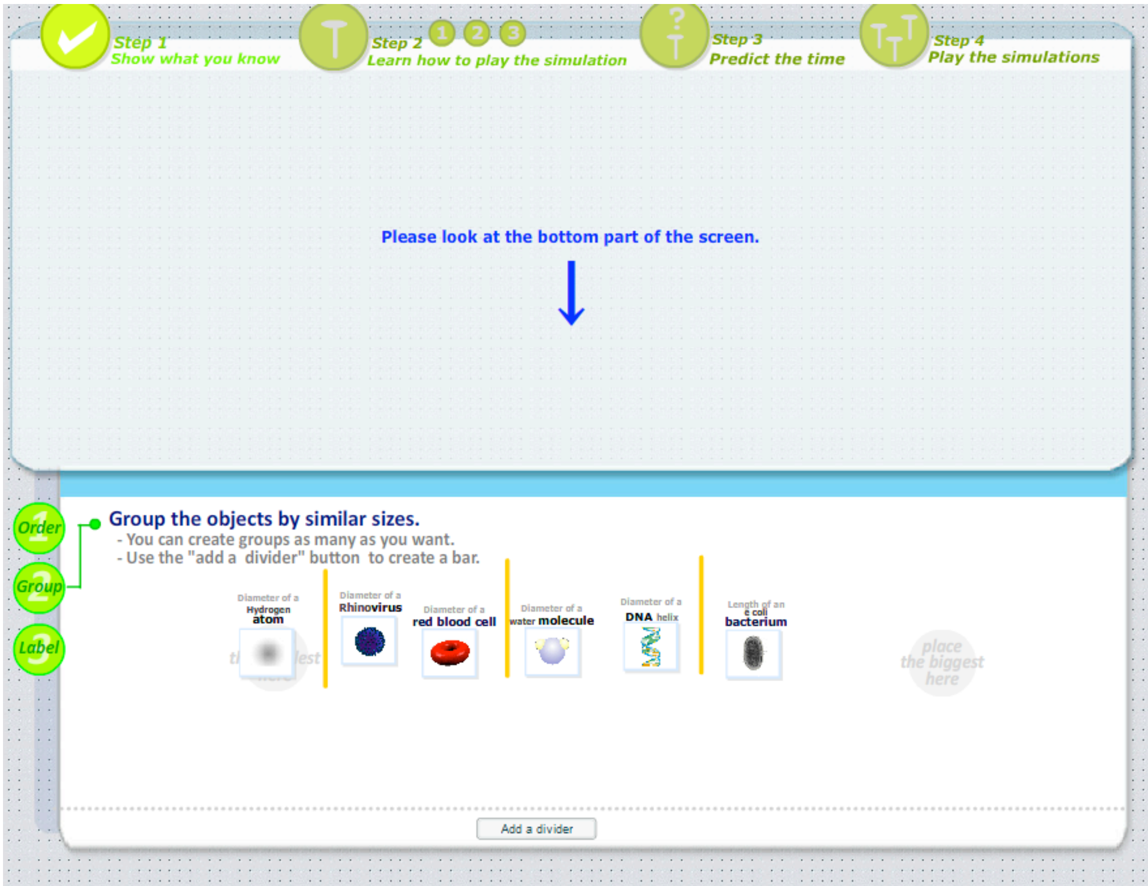


Figure 3.8. Card sorting task 2 – Grouping.
Students group the objects by similar sizes. They can create the size groups by placing the “dividers” (yellow sticks) between the groups. Students can add more dividers by pressing the “Add a divider” button on the bottom of the screen. Students are allowed to create between one and six groups. Please note that this image does not show an appropriate sorting of the cards. It just demonstrates how a student would do it.

Figure 3.9. Card sorting task 3 – Labeling.
Students give names to the size groups that they have created. The vocabularies they can use are written on the board in the computer lab.

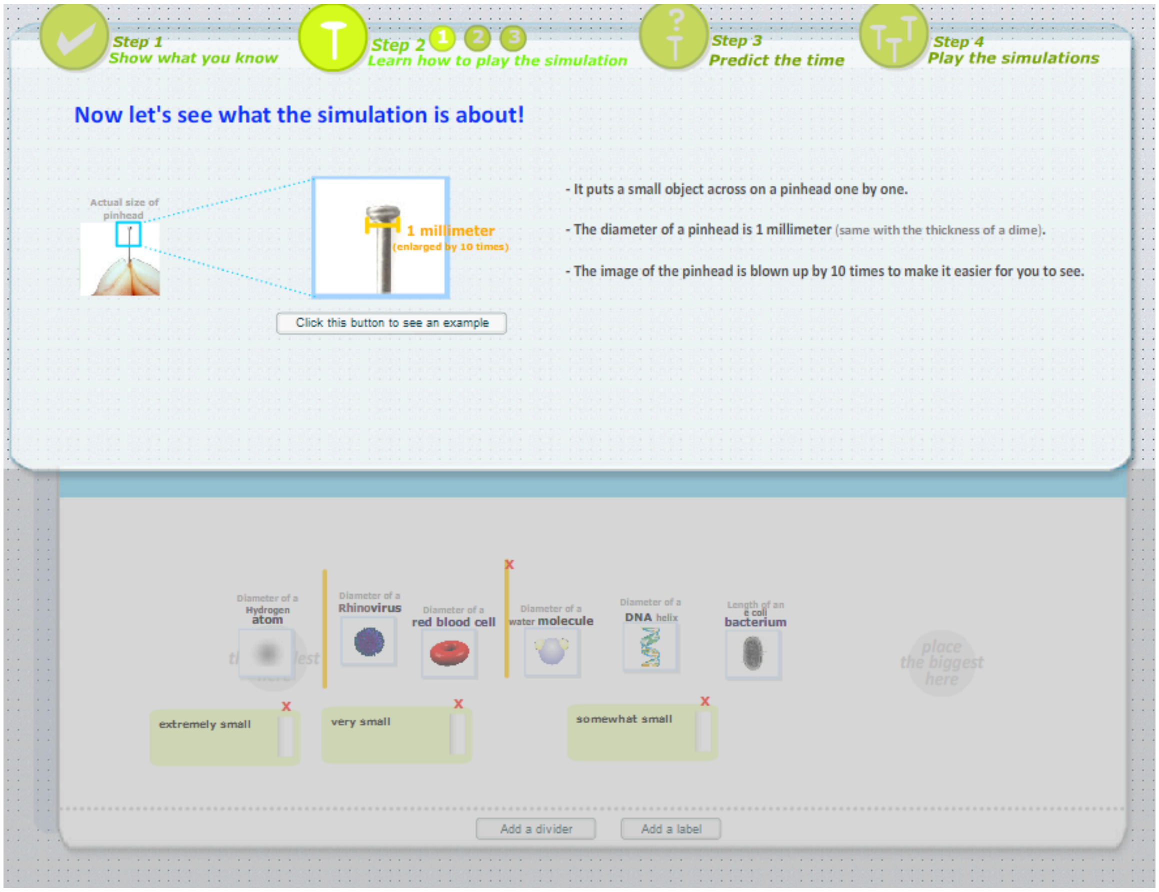


Figure 3.10. Introduction to TAVR (part 1).
Students play a very simple version of TAVR. It puts five dots (about 0.2 millimeters in size) across the head of pin. During this activity, the instructor tells the students that the accumulation is happening on the head of the pin that is provided to them. Notice that the student workspace in the bottom part of the screen is deactivated in order to prevent students from randomly playing around with the cards, until they set up hypotheses for the total accumulation time for each object. The navigation buttons (in the upper part of the screen) are not activated except the one that allows the students to move forward to the next step.

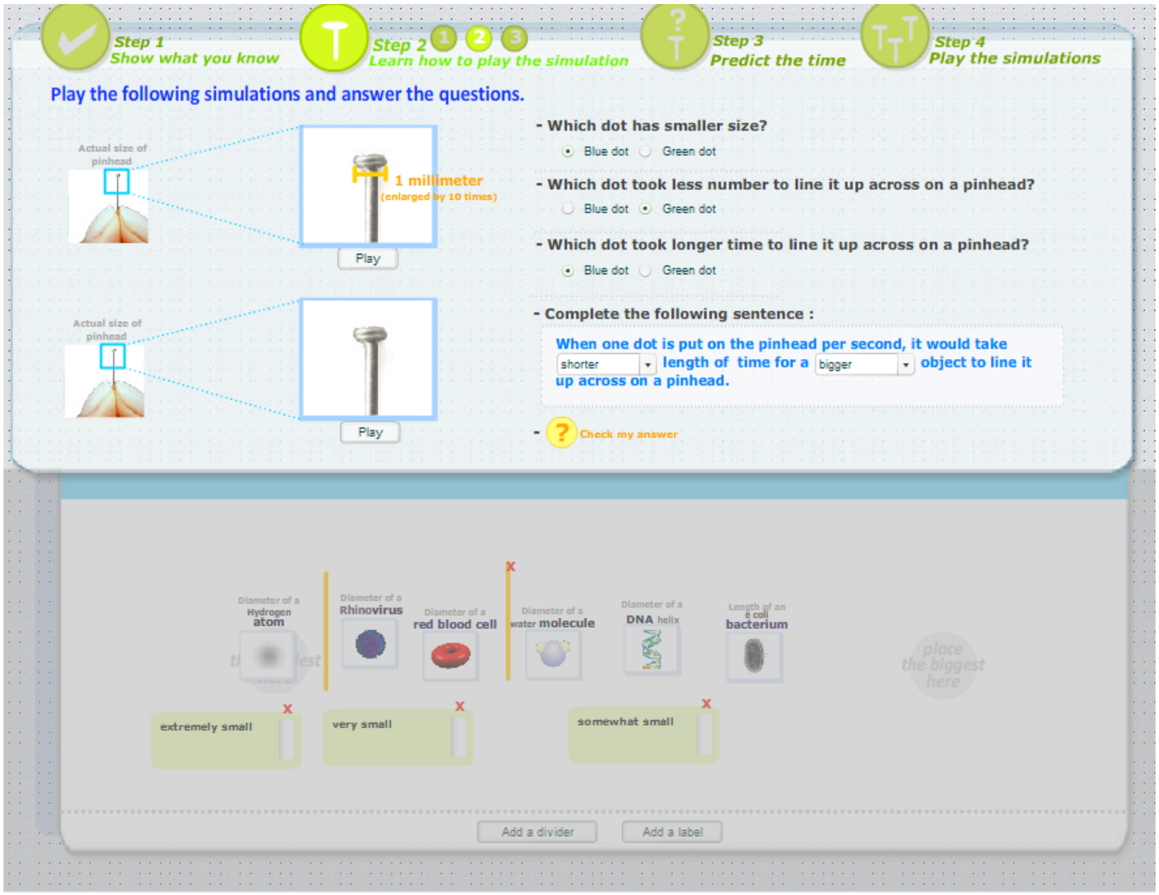


Figure 3.11. Introduction to TAVR (part 2).
Students play two sample TAVRs of two objects with different sizes, and solve a set of quizzes that direct them to think about the inverse relationship between the accumulation time and the object size. Students must give correct answers to become able to move to the next step. Again, students are told that the accumulation takes place on the tiny little pinhead in front of them.

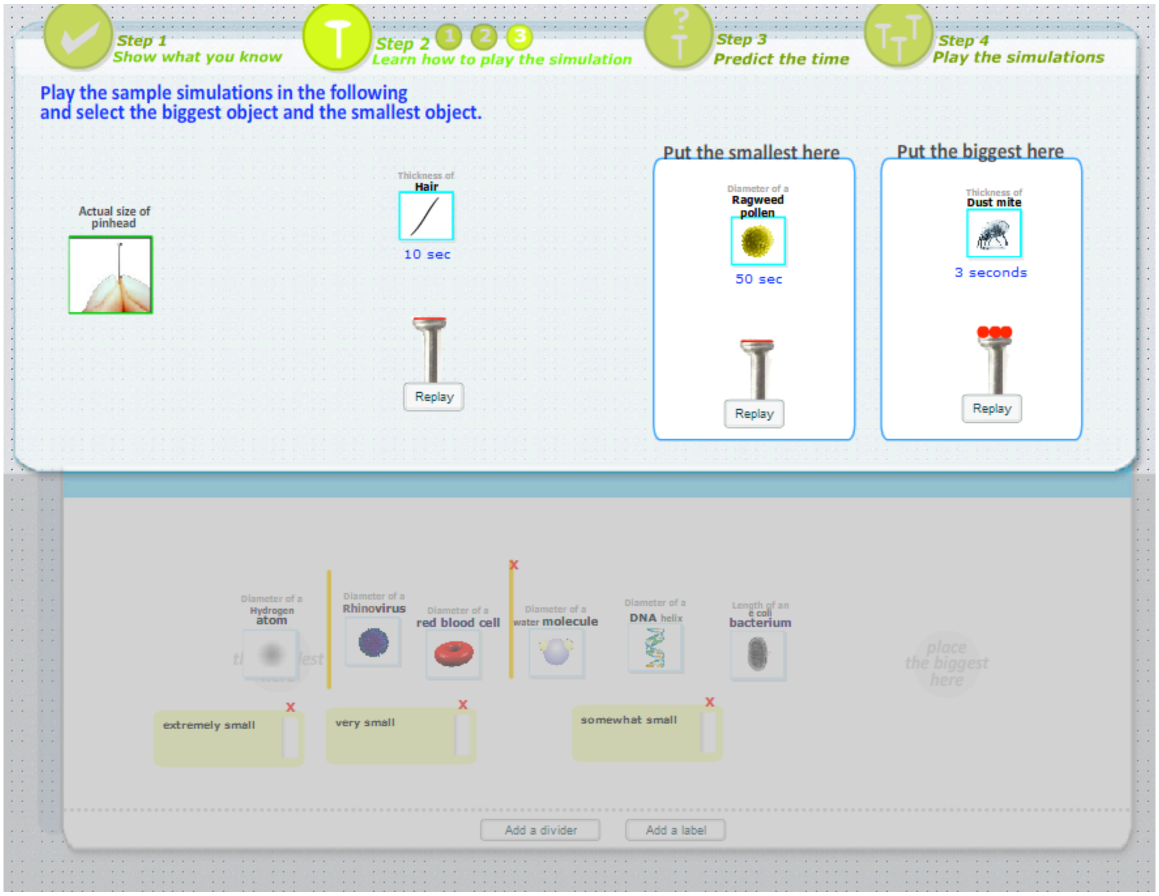


Figure 3.12. Introduction to TAVR 3 (part 3).
Students answer another quiz. They are asked to play three different sample TAVRs in order to select the biggest and the smallest. Among the objects, hair is macroscopic (in addition to the dust mite) and an object that students are familiar with. Students are told to remember how long it took for the hair to be completely accumulated. In the next step, they will be asked to use this information to predict the accumulation time for the imperceptible objects. Also in this phase, students are told, again, that the accumulation takes place on the tiny little pinhead in front of them.

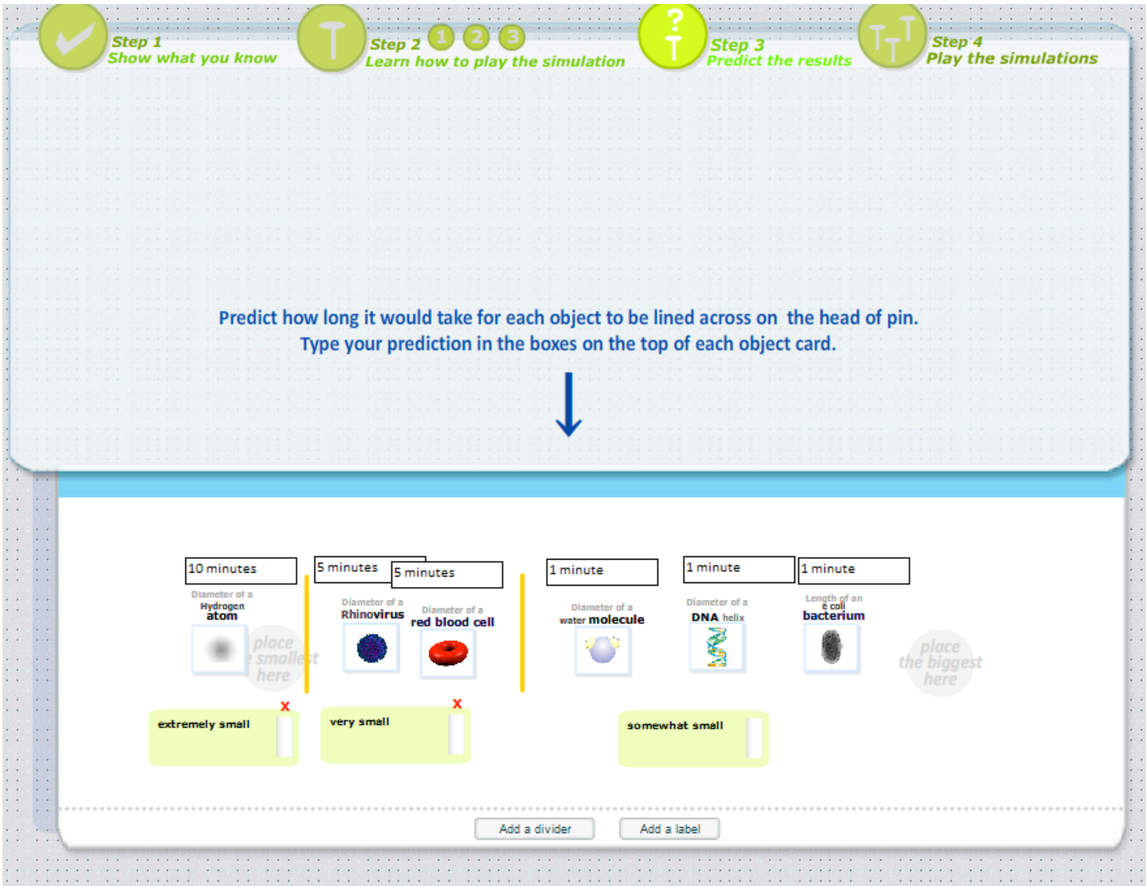


Figure 3.13. Predict the accumulation time for each object.
In the work space, students predict the accumulation time of each imperceptible object based on their hypotheses of the sizes of each object.

Figure 3.14. Students finally play all TAVRs and revise their card sorting work. Students can start playing the TAVRs by pressing the "Play all" button on the top right corner of the screen. They are allowed to use the ">> Skip ahead" button if they want to. When the accumulations are complete (or in the middle while the accumulations are progressing), students are asked to revise their card sorting work (all of the ordering, grouping, and labeling) in the work space by looking at the actual TAVR results.

Figure 3.15. Students reflect on the difference between their prediction and the actual results.
In this screen, students are provided with their predictions and the actual results. After pondering the different between the two, they are asked to reflect using a seven point Likert scale that offered the options of "extremely smaller than I thought, much smaller than I thought, somewhat smaller than I thought, similar with what I thought, somewhat bigger than I thought, much bigger than I thought, and extremely bigger than I thought."

Prior work

The prior work introduced in this section is composed of the summary of two pilot studies (Song & Quintana, 2009, 2010) that were conducted in order to explore whether middle school students could interpret what TAVR represents, and also in order to anchor more advanced research questions. For the pilot studies, I framed the research questions into three themes: (1) examining whether students can appropriately interpret

TAVRs (research question 1 in the following), (2) investigating whether and how they interpret their temporal experience and reconstruct their idea of imperceptible sizes (research question 2), and (3) identifying how students interpret units of time (research question 3).

Methods

Forty-five middle school students participated in the study for research questions 1 and 2, and thirty-five middle school students participated for research question 3. There was no overlap in the participants between the two groups because research question 3 was formed and studied after research questions 1 and 2 were pursued. In total, eighty middle school students from local public and private schools located in the Midwest region of the United States participated in the study.

For the pilot studies, WIIS did not ask the students to label size groups during the card sorting tasks. The labeling task was designed and added after the pilot studies were completed. Other than skipping the labeling task, the rest of the student learning activities (including the imperceptible objects) were the same as those introduced in the scenario in the previous section in this chapter. The data were collected via direct recordings of the students' activities on the computer screen (including pre- and posttest card sorting tasks), voice recordings of interviews, student-generated representations (using pencil and paper), and surveys. To analyze the data, coding schemes were developed to quantify them, following Chi's (1997) finding that quantifying subjective or qualitative verbal utterances is useful for objectively assessing learners' knowledge.

Results

Research Question 1: Do students understand what TAVR represents?

In students' answers to the questions that aimed to assess their understanding of the mechanism of TAVRs, 91% of the participants showed evidence that they had correctly comprehended what TAVRs represent. In detail, 41 out of 45 students accurately stated that (1) one click corresponds to one object placed on the head of a pin, (2) there only exists sound (the clicks) while the accumulation is processing within the imperceptible scale, and (3) the longer the accumulation, the smaller the size. This implies that TAVR and the inverse relationship between the accumulation duration and the size of an object may not be a difficult concept for middle school students to understand.

Research Question 2: Does TAVR help learners refine their mental model of the range of imperceptible sizes?

All participants correctly sorted the objects by size using the temporal information given in the TAVRs, implying that they understood the inverse relationship between the time and size. 84 % of the students exhibited evidence that showed that TAVR was useful for them for refining their mental models of imperceptible sizes; the number of the size groups increased in the posttest, although there did not exist a unified number of the size groups. Three out of eleven participants classified the objects into three groups, simply by looking at the biggest time units of the total accumulation durations. Some students overly amplified the differences between the sizes of the imperceptible objects. These students classified the sizes into more than three groups because they thought the differences between the accumulation durations of some imperceptible objects were large enough to be separated into different size groups. It appeared that this difference occurred mainly because of the difference in the participants' interpretations of the durations of time, and this observation led me to generate research question 3, which explored the pattern of middle school students' interpretations of different durations of time.

82% (37 out of 45) of the students stated that all of the given imperceptible objects turned out to be smaller than they thought. 18% (8 students) answered that the sizes of red blood cell was similar to their initial beliefs and the rest of the objects were smaller than they thought. In particular, the smaller the size of the objects, the more dramatic the reactions the students had, and this may imply that TAVR is particularly effective for conveying sizes that are extremely small.

During the focus group interviews, I noticed that (1) all students actively referred to the difference between their prediction and the actual result to respond to the Likert scale question, (2) the responses in the Likert scale were consistent with the difference between the prediction and result they noticed, and (3) their responses in the scale were consistent with the severity of the difference between the prediction and result. For example, a student who predicted that it would take less than one hour for an atom to span the pinhead (it takes about twelve days) answered that the size of an atom was "extremely smaller than I thought" in the scale. The same student responded that the size of a red blood cell was similar with what he thought because his prediction was ten seconds, which was only four seconds short of the actual duration (it takes about fourteen seconds for a red blood cell). This consistency in the justification was noticed in all participants' responses.

Research question 3: How do students classify the different durations of time into groups?

Many students (80%) grouped the durations of time just using the same biggest units. They classified the time by seconds, minutes, hours, days, and years. The rest of the students (20%, 9 students) grouped the lengths of time using other strategies. For example, one student grouped time longer than one day into one "very long time" group that even included one year. Another student converted thirty-six days into one month and five days and classified it as an independent "month" group.

Implications of the Findings and Remaining Issues

The findings from the pilot study indicate that average middle school students may not have difficulty in interpreting what is represented in TAVR. All of the participating students understood the inverse relationship between time and size after introductory learning activities. In the post-instruction card sorting task, the students accurately ordered the given imperceptible objects by size, using the accumulation durations that were shown in TAVRs, implying that they were able to correctly understand the inverse relationship between the size and accumulation duration. The students' performances on the grouping task showed that they constructed more refined mental models of the range of imperceptible sizes after the learning activity with TAVRs. It was clear that time was a meaningful concept for learners upon which they could analogically achieve better comprehension of abstract spatial knowledge.

The findings from the prior work provide a foundation for using TAVRs for learning imperceptible sizes. They show that the students could understand what a TAVR represents and make meaningful interpretations of their learning with TAVRs depending on their knowledge of the durations of time (in association with the units of time). However, these findings do not provide insights into how to effectively use temporal representation in association with other augmenting modalities, or how to best augment the temporal experiences for teaching and learning imperceptible phenomena. In the following chapter, I discuss the research questions for this dissertation in depth.

CHAPTER 4 RESEARCH QUESTIONS AND METHODS

Research Question 1

How do the combinations of the supporting modalities support learners to construct the mental model of the range of imperceptible sizes?

TAVR is a multimodal representation. It is composed of a temporal representation and other two supporting representations, aural and visual, which are adopted to augment learners' temporal experiences with imperceptible sizes. It is known that the nature of human cognition is multimodal and learners have a preference for interacting multimodally rather than unimodally (Oviatt, 1999; Waldrip et al., 2010). However, one cannot guarantee that learners will make use of all three modalities and integrate them into their learning. It cannot always be assumed that people will interact multimodally with a multimodal system just because they tend to prefer such systems (Oviatt, 1999).

This issue brings up a need for an investigation into whether students make use of each augmenting modality, aural and visual, and if so, how they synthesize what is conveyed via the augmenting modalities. Examining which combination of the augmenting modalities results in the greatest learning benefits will provide the insight into this. By contrasting the learning outcomes from the students' learning activities with the TAVRs with varying modality combinations, I will be able to identify the roles of each augmenting modality during students' temporal experience with imperceptible sizes, and how they are encoded in learners' mental models.

Additionally, comparing the effects of different modality combinations will also tell me whether the modalities in TAVR are overloading students' cognitive capacity. According to the cognitive load theory, a learner should be facilitated in using his or her limited working memory efficiently, because human working memory is limited with respect to the amount of information it can hold and the number of operations it can perform on that information (Pass et al., 2004; Sweller, 1988). If the resources of a learner's working memory are exceeded due to the difficulty of a task or the amount of given information, learning will be ineffective. If the three modalities (or certain combinations of them) in TAVRs are too much for students, the results will be seen in low student performances. Although both dual coding theory (Paivio, 1986) and working memory structure theory (Baddeley, 1986) imply that visual and aural information are conveniently processed in parallel, I cannot be certain that the students will easily integrate and process the information with the co-presence of a temporal modality.

In order to compare the effect of different modality combinations, different student groups for each combination of the modalities were formed, as summarized in Table 4.1. I discuss the details of each combination after the table.

Groups	Modality		
	Temporal	Aural	Visual
TAVR	●	●	●
TVR	●	-	●
TAR	●	●	-
TR	●	-	-

Table 4.1. The combinations of modalities for each student group.
(• indicates that the corresponding modality will be included and - means the opposite.)

- *TAVR*: The students in the TAVR group interacted with original TAVRs that have all three modalities. The students in this group were expected to show the best performance, meaning that they would correctly order the cards, create more size groups, and use more descriptive words in the labels of the size groups.

- *TVR*: The students in this group were given TVRs, the TAVRs without the aural component (the clicks). Without the aural representation, which was thought to be particularly useful while the accumulation is progressing within the imperceptible scale, students would have to virtually keep track of the accumulation only by using the clock (embedded in each TAVR) that shows the elapsed time. I thought that the students in this group might lose the sense of accumulation speed and eventually think that the accumulation was happening at either a faster or slower interval. Then they may develop a different approximation of the total number of accumulated objects.

  The comparison between the student performance in this group and the TAVR group is expected to help me identify whether and how students make use of aural representation in the TAVR. If the aural component plays a critical role in supporting learners to perceive and conceptualize the temporal experience of an imperceptible size, the students in the TVR group will show significantly different (lower) performances versus the students in the TAVR group.

- *TAR*: In this group, the students interacted with TARs (the TAVR without the visual – the red line) throughout the learning activity in WIIS. In TAR, the placement of an imperceptible object is represented via the repeating clicks (aural) and the clock. Hence the students would have to imagine the accumulation in their heads, relying on the clicks, even after the accumulation had entered the macroscopic scale. The results of the student performance in this group will help us see the role of the visual component in TAVR. If the visual component plays a critical role in supporting learners' ability to perceive and conceptualize the temporal experience of an imperceptible size, the students in the TAR group will show significantly lower performances than the students in the TAVR group.

There is another reason for testing this combination. Although the red line works only as an indicator of accumulation, there is a possibility that students think that they can actually see the accumulated objects with their naked eyes when they are accumulated in a single line, which is a scientifically incorrect idea. The imperceptible objects that are used in WIIS are very small, so it is impossible for us to see them when they are aligned in a single line. Although the length of the line of the accumulated objects is macroscopic, the thickness of the accumulated line must still be too thin to see in principle. If students can still appropriately perceive and conceptualize the temporal experience of imperceptible sizes without the visual representation, I may be able to prevent the construction of a misconception by not providing the visual representation.

- *TR*: The students in this group used the temporal representation without any supporting modalities (hence, TR). There was no explicit aural or visual representation on the pinhead; there were only the head of the pin and the clock that showed the elapsed time. To be fair with the other treatment groups, the TR group was also provided with a ">> Skip ahead" button for each TAVR. The completion of accumulation in each TAVR was indicated only by the stopping of the clock and a blue square that appears above the image of the pinhead. The result of the student performance in this group will inform me as to what extent a temporal representation can be used as a standalone representation for representing imperceptible sizes.

Hypothesis

I hypothesize that the learning effect will be relative to the number of the "constituents" of the mental model of sizes. By the term "constituents," I mean the verbal or non-verbal information that may contribute to the formation of a mental model of an imperceptible size. See Figure 3.3 and the corresponding discussions in Chapter 3 for the potential constituents of the mental model that may be constructed from TAVR experiences. I posit that more constituents will result in a more saturated and refined mental model. Considering that a mental model of a size can be comprised of interconnected verbal and non-verbal information, having constituents of these two types will be more advantageous for learning than having constituents only of a single type. Besides, it is known that experts tend to have more fragments of knowledge than novices, and that types of information are tightly linked with each other (Chi, Glaser, & Farr, 1988). Learning experiences that are conveyed in different forms of information may result in the construction of more complex mental models. Therefore, I expect that the students in the TAVR group will achieve the highest scores on their posttest (and also the largest increases from the pretest) because their temporal experiences with the imperceptible accumulation will be supported by both visual and aural representations, if dealing with three modalities is not overwhelming work for the students.

The students in the TVR group and the TAR group are predicted to score the next highest because they have only one supplemental modality (hence, fewer constituents) for experiencing the accumulation over the passage of time. Finally, the TR group is expected to rank the lowest because they do not have any additional support for inferring the accumulation of the imperceptible objects other than the clock. See Table 4.2 for a summary of the potential verbal and non-verbal constituents of the mental model of the size of an imperceptible object that may be constructed by interacting with the TAVR (and TVR, TAR, and TR). In Table 4.2 the kinesthetic experience was put in parenthesis because, as discussed in Chapter 3, the “Skip ahead” button was optional; students could choose or not choose to use the button, and even if they started using it, they could stop pressing the button any time they wanted to.

Research Question 2

How do different intervals of temporal experiences influence learners' conceptualization of the range of imperceptible sizes?

The first version of TAVR places ten objects in one second and was designed for practical in-class use. If the simulation takes too long, teachers will not be able to use WIIS in classrooms that allow only a limited period of instruction time, although students could fast forward the accumulation process with the “skip ahead” button. However, the initial prototype was designed to place one object in one second. I thought, in this way, it would be easier for learners to approximately grasp the number of the objects that were placed on the pin during the elapsed time. Additionally, as Table 4.3 and Figure 4.1 show, the total durations of accumulation become dramatically longer when the interval of accumulation is one object in one second, and hence the relative differences between the accumulation durations of different objects become even greater.

TAVR type	Constituents		Hypothesis
	Verbal information	Non-verbal information (mental images)	
TAVR	▪Verbal interpretations of the temporal experience. ▪Verbal interpretations of the visual mental image of the head of a pin. ▪Verbal interpretations of the visual mental image of the accumulation progression. ▪Verbal interpretations of the aural experience. ▪Verbal interpretations of the kinesthetic experience.	▪Visual mental image of the head of a pin. ▪Visual mental image of the accumulation progression (red line). ▪Auditory mental image of the clicks. ▪(Kinesthetic experience of button clicking.)	Highest
TVR	▪Verbal interpretations of the temporal experience. ▪Verbal interpretations of the visual mental image of the head of a pin. ▪Verbal interpretations of the visual mental image of the accumulation progression. ▪Verbal interpretations of the kinesthetic experience.	▪Visual mental image of the head of a pin. ▪Visual mental image of the accumulation (red line). ▪(Kinesthetic experience of button clicking.)	Middle
TAR	▪Verbal interpretations of the temporal experience. ▪Verbal interpretations of the aural experience. ▪Verbal interpretations of the kinesthetic experience.	▪Visual mental image of the head of a pin. ▪Auditory mental image of the clicks. ▪(Kinesthetic experience of button clicking.)	Middle
TR	▪Verbal interpretations of the temporal experience. ▪Verbal interpretations of the visual mental image of the head of a pin.	▪Visual mental image of the head of a pin. ▪ (Kinesthetic experience of button clicking.)	Lowest

Table 4.2. The comparison of the hypotheses regarding the constituents of each manipulation type.

In the prior study, I observed that learners tend to make meaningful interpretations of the differences between the accumulation durations of each imperceptible object, especially when the sizes of the objects were submicroscopic (smaller than one micrometer), which results in accumulation durations that are longer than one day. Based on this observation, I expect that radical differences between the total durations of accumulation may direct them to form even more distinct conceptual categories of imperceptible sizes (i.e., higher number of groups and more elaborated labels).

Object	Accumulation intervals	
	10 objects / sec (compressed)	**1 object / sec (extended)**
Hydrogen atom	11 d 13 h 46 min 40 sec	115 d 17 h 46 min 40 sec
Water molecule	2 d 7 h 33 min 35 sec	23 d 3 h 30 min 35 sec
DNA helix	13 h 53 min 20 sec	5 d 18h 53 min 20 sec
Rhino virus	1 h 15 min 45 sec	12 h 37 min 30 sec
e coli bacterium	50 sec	8 min 20 sec
Red blood cell	16 sec	2 min 40 sec

Table 4.3. Comparison between two different velocities of accumulation in TAVR.

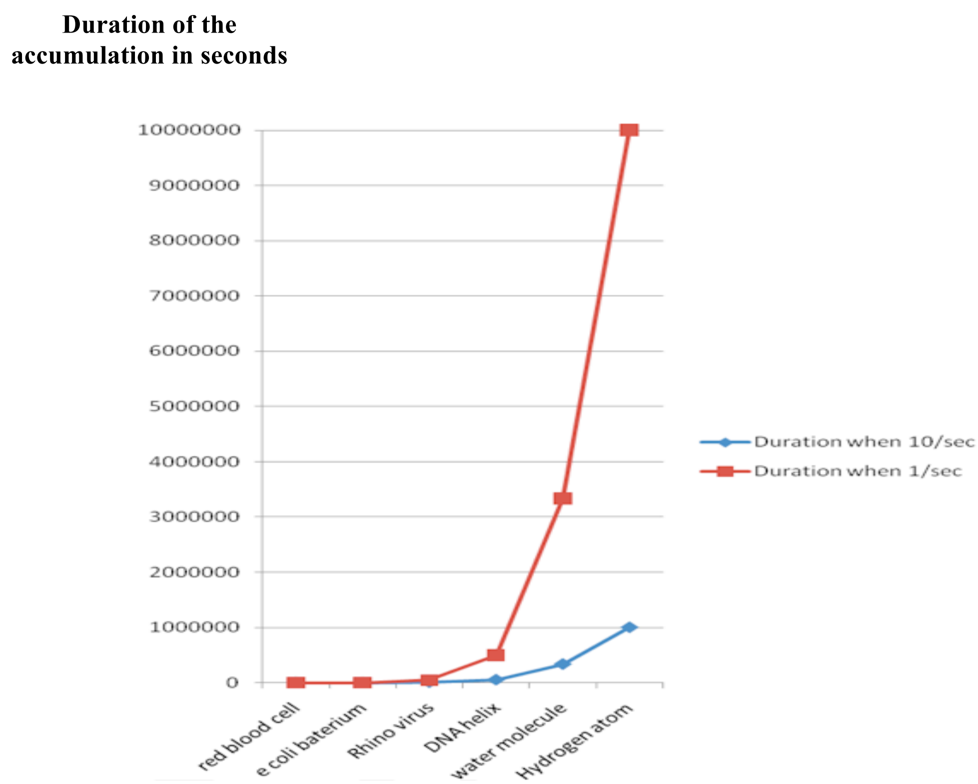


Figure 4.1. A graph that shows the differences in the durations of object accumulation at different accumulation rapidities.

Hypothesis

I hypothesize that the students who interacted with the TAVRs with the slower accumulation rapidity ("extended," one object/sec) would experience more discrete and distinguished imperceptible sizes, and therefore, would construct more refined imperceptible size categories than the students who interacted with the TAVRs with faster accumulation rapidity ("compressed," ten objects/sec). Since the extended temporal experience results in more drastic differences between the accumulation durations of the objects, learners might create more size groups and use more descriptive words to label the size groups. The results from this research question will let me discuss potential design suggestions for a temporal representation for abstract spatial information that requires learners to construct fine-grained conceptual categories of that spatial information.

Research Question 3

How do active temporal manipulations vs. passive observation by students influence their perceptions of the durations of the temporal experiences with TAVRs and the conceptualization of the range of imperceptible sizes?

The first version of TAVR, which was explored in the prior work, successfully supported the students in constructing a more refined conceptual range of imperceptible sizes. However, one influential feature of TAVR was not included in the investigation of the impact of TAVR on students' learning: the "Skip ahead" button. When students use the button, the differences between accumulation durations are experienced as being perceptually much shorter than the actual difference that one may feel in reality. For example, a student must wait for nine days to observe the completion of the accumulation of hydrogen atoms, the smallest among the given objects, after watching the second smallest object (a water molecule) be completely accumulated across the head of a pin in about two days. However, he or she can accelerate the passage of time by clicking the "*>>Skip ahead"* button for the hydrogen atom TAVR several times. Depending on how arduously and patiently the student clicks the button (and how I set up the number of the objects to be skipped at each button pressing), it would take about ten minutes at the maximum, and five minutes at the minimum.

This feature was disregarded in the prior studies, because the focus of the prior work was solely on examining whether the students could understand TAVR and how they made use of the concept of the passage of time in order to conceptualize the range of imperceptible sizes. Another reason was the inconsistency between the possible ways of utilizing the button. For example, some students might decide to observe the accumulation for a long period of time before they come to a conclusion that they had better start pressing the button, while other students might start pressing the button right away.

However, interestingly, most of the students whom I observed in the prior studies started using the "Skip ahead" button almost from the beginning of the accumulation once they understood its use because they "*wanted to see the total time as soon as possible.*" They pressed the button, switching between different objects, in order to figure out which object takes the longest time to accumulate (and hence is the smallest). The students in fact seemed to be enjoying the repetitive and laborious button pressing. Some students even raced (to see the accumulation completing) with peers, particularly for the smallest object (hydrogen atom). In the interviews in the pilot study, I noticed that the students made connections between this arduous kinesthetic experience and the durations, unexpectedly more than I thought they would, in addition to the "*surprisingly long*" temporal experience (the accumulation duration) itself. These observations brought up the need to investigate research question 3.

Knowing whether and how the kinesthetic experience impacts the students' mental models of imperceptible sizes will enable the generation of design suggestions regarding how to enhance a learning activity that employs temporal representation. In order to explicitly differentiate the effects of the kinesthetic experiences on students' learning, these effects must be compared with those of the TAVRs that automatically accelerate. Furthermore, an issue regarding whether the manipulation of the temporal experience is ever more effective for constructing more refined mental models than a non-manipulative temporal experience should also be looked into. Taken together, I propose three different treatment groups for research question 3, as shown in Table 4.4.

Group	TAVR type	Description
1	**No manipulation**: Unmodified perceptual temporal experience without any manipulation	Students simply observe the accumulation in TAVRs that do not provide any temporal manipulation features. The temporal experiences will perceptually match the accumulation durations that appear in TAVRs.
2	**Automatic manipulation**: Automatic acceleration of the accumulation (automatic skipping ahead)	Students observe the TAVRs that automatically skip through the accumulation. Students watch the TAVRs accumulate the imperceptible objects at an accelerated accumulation velocity.
3	**Interactive kinesthetic manipulation**: Labor-intensive acceleration of the accumulation by students (repetitive "Skip ahead" button pressing)	Students can repetitively press the "Skip ahead" button to accelerate the accumulation. The perceived accumulation duration will be much shorter than the accumulation duration that will appear in TAVRs.

Table 4.4. Summary of the three different temporal manipulation conditions.

- *Group 1 – No manipulation (NM)*: The students in this group use the TAVRs with neither a “Skip ahead” button nor an automatic acceleration feature. They are supposed to “just watch” the imperceptible objects being accumulated across the head of a pin. Therefore, the duration of their temporal experience will be the actual duration of the accumulation.
- *Group 2 – Automatic manipulation (AM)*: This group watches TAVRs that automatically skip through the accumulation, without having any input into the accumulation process. The occurrence of the automatic acceleration is announced to students by a change in color of the border of the object image. The total accumulation durations that students physically perceive are much shorter than the actual passage of time in reality.
- *Group 3 – Interactive manipulation (IM)*: The students in this group are supposed to interactively accelerate the accumulations in TAVRs by repetitively pressing the “>> Skip ahead” button. For example, students have to click the button hundreds times to finish the accumulation of the smallest objects.

Hypothesis

I expect that, since the temporal experiences are to be re-represented into verbal information (e.g., the units of time and the student’s verbal interpretation of the passage of time) in learners’ mental models, their impact will depend on how the student interpret their temporal experiences; what was perceived either in manipulated or non-manipulated fashion. If the students who used the TAVR with no temporal manipulation developed interpretations that are similar to those of the students who used the TAVR with the interactive temporal manipulation feature (e.g., similar number of size groups), it would mean that the physical perception of the passage of time was not particularly influential. It would also imply that the duration that appears on a clock was more influential and reliable information to students than the perceived passage of time through their temporal modality. In contrast, if students developed significantly different interpretations – for example, the no temporal manipulation group created significantly more size groups – it would imply that the physical perception of the passage of time actually influenced the way learners interpreted the passage of time that is shown in the clock.

I predict that the interactive manipulation of the temporal experience will be more influential than the natural passage of time, not only because the kinesthetic experience is tightly linked to the temporal experience, but also because it is supposed to generate two constituents for a mental model: a kinesthetic mental image and the verbal interpretation of the kinesthetic experience, which are tightly interwoven with each other. For example, students would remember the sensation they felt in the muscles of their fingers and wrists while pressing the "Skip ahead" button (kinesthetic mental image), and they would also remember that the sensation gradually became irritating after pressing the button for a certain period of time (kinesthetic mental image with temporal dimension). They would also make verbal interpretations of these perceptual experiences into phrases, such as "*my fingers started to feel uncomfortable because I had to press the button so many times for a long time,*" as one of the students who participated in the prior study commented. These constituents are in two different formats, verbal and non-verbal, and are expected to be tightly coupled in learners' mental models. These are richer than verbal information alone.

Integrating all together, I hypothesize that group 3, the IM group, would score the highest in their card sorting task and offer more intense reflections (e.g., "extremely smaller than I thought") on the seven point Likert scale because the temporal and kinesthetic experiences are expected to be re-represented in both the kinesthetic mental image and the verbal information, interlinked. Group 1, the NM group, would perform the second highest, because the impressively long temporal experience, which was actually perceived through the channel of temporal modality, might influence the way students generate verbal interpretations of their temporal experiences with TAVRs. The physically perceived experience of enforced waiting for a significantly long time to observe the accumulation of hydrogen atoms being completed may color their translation of the passage of time. Lastly, group 2, the AM group, would score the lowest. The students in this group have to perceive accumulation durations that are shorter than the real passage of time without any opportunity to enrich their abbreviated temporal experiences with other verbal or non-verbal information.

See Table 4.5 for the list of the constituents in each temporal manipulation condition and corresponding prediction. Although I did not discuss every TAVR condition in the above, I included the visual and auditory mental images (non-verbal information) and the verbal interpretations (verbal information) as the constituents of the mental models that may be generated after the learning activities with each condition.

Methods

Overview

The initiation of this study was motivated by acknowledging the difficulties that middle school students have in conceptualizing the range of imperceptible sizes, which has long been a challenging topic in science education, and I attempted to address the challenge by designing and providing an innovation. This study is a design experiment, which differs from a traditional empirical approach, in that design research aims to engineer and enact the designs with hypothesized learning processes, whereas traditional empirical research focuses on studying participants' behavioral and cognitive actions in a setting with no particular interventions (Brown, 1992). Design research is conducted by stepping through iterative cycles of (1) analysis of practical problems, (2) development of solutions, (3) iterative cycles of testing and refinement of solutions in practice, and (4) reflection to produce design principles and enhance solution implementations (The Design-Based Research Collective., 2003). This study is framed by following this cycle, and this dissertation is situated in the second cycle of the iteration of the design.

The main goal of design research is to generate innovative forms of learning and theoretical development by investigating the complex and dynamic interactions between learners and educational improvement through iterative design and implementation processes (Brown, 1992; diSessa & Cobb, 2004; The Design-Based Research Collective., 2003). Hence, design research must not only document the success or failure of a designed system but also focus on interactions that refine our understanding of the learning and thinking processes involved. In order to meet these requirements of design research, I employed a mixed method approach, not only because it provides a more complete view of the research topic than a single type of method, but also because it allows a study to flexibly adopt different specific methods for different phases of the intervention.

Group	TAVR type	Constituents		Prediction
		Verbal information	**Non-verbal information (mental images)**	
1	No manipulation	･Verbal interpretations of the unmodified realistic temporal experience through temporal modality. ･Verbal interpretations of the visual mental image of the head of a pin. ･Verbal interpretations of the aural mental image of the clicks.	･Visual mental image of the head of a pin. ･Visual mental image of the accumulation on the pin. ･Auditory mental image of the clicks.	Middle
2	Non-interactive acceleration	･Verbal interpretations of the manipulated temporal experience through temporal modality. ･Verbal interpretations of the visual mental image of the head of a pin. ･Verbal interpretations of the aural mental image of the clicks.	･Visual mental image of the head of a pin ･Visual mental image of the accumulation on the pin. ･Auditory mental image of the clicks	Lowest
3	Interactive kinesthetic manipulation	･Verbal interpretations of the manipulated temporal experience through temporal modality. · Verbal interpretations of the kinesthetic experience. ･Verbal interpretations of the visual mental image of the head of a pin. ･Verbal interpretations of the aural mental image of the clicks.	･Visual mental image of the head of a pin. ･Visual mental image of the accumulation on the pin. · Kinesthetic mental image of the arduous button pressing. ･Auditory mental image of the clicks.	Highest

Table 4.5. The comparison of the hypotheses regarding the constituents of each manipulation type.

I collected and synthesized the data from student surveys, the card sorting tasks, the seven-point Likert scale, verbal comments, and focus group interviews. The student tasks for this study were designed to reveal the properties of their mental models of the range of imperceptible sizes. Students' hypotheses, revisions, and reflections were the primary data that were used to address the research questions. Student background surveys helped me understand the context of the findings from the data. The data collected from student performances on the tasks were quantified and statistically analyzed. The student responses for the focus group interviews were first checked for logical consistency, and then were used as qualitative data.

The data were collected and analyzed using a quasi-experimental design. Quasi-experimental designs are similar to normal experimental designs except that the participants are not randomly assigned to groups. Since the participating school's cooperation is essential for collecting data, and assigning the participants to tasks is heavily influenced by the school's schedule, I was unable to randomly assign students to treatment groups. In research that compares different treatment groups, a researcher has to eliminate the threats to internal validity, which are alternative causes other than the treatment that may be responsible for differences in observed outcomes. Quasi-experimental research is prone to a threat called "selection bias," which refers to the differences between groups that may interact with independent variables, and thus influence observed outcomes. Selection bias can make it difficult for a researcher to determine whether the discrepancy between the groups is due to the independent variable or subject-related variables. In order to address the possibility selection bias, I examined correlations between exogenous variables and a treatment indicator. To validate my interpretation of students' performances, I conducted focus group interviews to assess whether it was truly the treatment that influenced the students' mental model refinements and whether what their performances show in fact reflect their mental models.

External validity concerns the degree to which conclusions can be generalized beyond the sample. If participating students were recruited from private schools, for instance, the findings would not be able to be applied to the general student population of the United States. Therefore, in order to maximize generalizability, I recruited the students from public schools. Furthermore, I collected background information on the participants that included their gender, age, achievements in science, and experiences with interactive computer programs to set the boundaries of the generalizations that I can make conclusions within.

In the following, I discuss the details of the student tasks, and the corresponding data collection and analysis.

Data Collection

Participants

I collected multiple sources of data from nine seventh grade science classes, a total of two hundred thirty-one students, in collaboration with three science teachers at local public middle schools. The teachers voluntarily offered their support for student participation and their instructional support during data collection as well. The student groups were divided, first by their teachers for three different research questions (i.e., the students of teacher A participated in the data collection for research question 1, teacher B for research question 2, and teacher C for research question 3), and then by their class hours for each treatment of the conforming research question. To collect data, I went into each group's science class and instructed the students for one class period (50 minutes). I met with the teachers days prior to giving the instructions in order to check whether the teachers felt comfortable with the setting of the learning activities and the content.

The instruction was held in a computer lab, and every student was assigned to an individual computer. Their teachers were present in the computer lab throughout the class hour and supported the instruction. I was introduced to the students, by their teachers, as a researcher from the University of Michigan who was in need of their help for designing and developing a computer program for middle school students of the United States. I told them what they were going to do was not a test that would be reported as part of their GPA; rather, I emphasized that they were going to help me find out if the computer program was good or bad for the students like them. Then the teachers told the students that they would still look at how mindfully the students worked on the tasks.

The students accessed WIIS via web browsers. They were provided with headsets for listening to the clicks in TAVRs. The students who used the no-sound TVR (TAVR without the aural representation) did not use headsets. In the following, I introduce the participants for each research question in detail.

Research Question 1

Teacher A offered me five class hours total, and I decided to use her students for research question 1, which required the participation of four student groups. I used the first class as a rehearsal (the students received the same instruction with the TARs). Hence, there were four valid student groups who had learning activities with different forms of the TAVRs (TAVR, TAR, TVR, and TR). As a result, 104 seventh grade students participated in this study. See Table 4.6 for the composition of the participating students. The original number of the participants was bigger than the number of students in Table 4.6 because the students who did not finish the learning activity, who had a learning disorder, or whose first language was not English (these students needed an interpreter throughout the instruction) were eliminated from the data for the analysis.

Groups	Participants			Focus group interview
	male	female	**Total**	
TAVR	13	14	**27**	5
TVR	14	12	**26**	4
TAR	14	12	**26**	4
TR	12	13	**25**	4
Total	**53**	**51**	**104**	**17**

Table 4.6. The number of participants for each treatment group for research question 1.

The students had already been briefly introduced to the imperceptible objects during their previous science classes by their science teachers. For example, most of them already knew that Rhinovirus was a germ that causes the common flu, atoms are the basic building blocks of every object in the world, and DNA is *"in our body and has something to do with genes."* The gender distribution was almost even for all student groups. The ethnicities of the students were also evenly distributed between African-American and Caucasian. There were zero or one Asian/Pacific Islander and one or two Hispanic students in each class. The five participants for the focus group interviews for each TAVR condition were randomly selected according to their seat assignments. For instance, the individual computers in the room were numbered, and I had pre-selected the numbers of the computers using certain intervals (e.g., 1, 5, 10, 15, 20, 25, 30) before the instruction started, so that my selection of the students was not influenced (either consciously or subconsciously) by my experience with the students during the instruction. However, focus group interview participants had to be replaced if an initial student did not finish the learning activity, did not speak English as his or her first language, or had a learning disorder. In these cases, I interviewed the student who was sitting in the next seat. Before the beginning of the instruction, I announced to the class that a few of the students would be "randomly" selected for the interviews.

Research Question 2

Teacher B allowed me to come into three of her science classes. However, as a result, I could only instruct two and a half classes, because one class had to be intermittently ceased because the teacher felt that the students were becoming out of control (i.e., chatting too much with peers and not cooperating with the instructor) and wanted them to leave the computer room. In consequence, fifty-seven seventh grade students participated for research question 2 (see Table 4.7). The students were divided into two groups, according to their class hours. The students had already been briefly introduced to the imperceptible objects during their previous science classes. The ethnicities of the students were evenly distributed between African-American and Caucasian. A few of the students were of Asian or Hispanic origin. The numbers of the female and male students were also almost even. As in research question 1, the students who had a learning disorder, whose first language was not English, or who did not finish the tasks were excluded from the data. I used the same strategy as above for selecting the students for the focus group interviews.

Groups	Participants			Focus group interview
	Male	Female	**Total**	
1/sec (Extended)	16	13	**29**	5
10/sec (Compressed)	15	13	**28**	3
Total	31	26	**57**	8

Table 4.7. The number of the participants for each treatment group for research question 2.

Research Question 3

Teacher C invited me to four of her science classes, and I instructed all four classes. I used one class as a rehearsal for the instruction with “no temporal manipulation" TAVRs. A total of seventy seventh grade students participated for research question 2. The students were divided into three groups, according to their class hour. As Table 4.8 shows, the male and the female students were almost evenly distributed in all groups. The ethnicity of the students was mainly Caucasian (*N*=56, 80% of the students). Two to four students in each student group had an Asian or African ethnicity. Every student finished all student tasks and they all spoke English as their first language. None of them had a learning disorder. I used the same strategy as above for selecting the students for the focus group interviews.

Group	TAVR type	Participants			Focus group interview
		Male	Female	**Total**	
1	No Manipulation	10	13	23	4
2	Automatic Manipulation	11	12	23	4
3	Interactive kinesthetic acceleration	12	12	24	4
Total		**33**	**37**	**70**	12

Table 4.8. The number of participants for each treatment group for research question 3.

Student Tasks

The research questions concern the changes in learners' mental models that do not involve mastering complex problem solving skills; they focus on detecting the changes in students' mental models of sizes. A useful method for assessing the changes in learners' mental models is to observe the changes in the way they represent their knowledge (Ainsworth, Bibby, & Wood, 2002; Rowe & Cooke, 1995; Vosniadou & Brewer, 1992, 1994). Taking this approach for all research questions, I sought to examine how the students re-represented their knowledge of the range of imperceptible sizes after the learning activity with the TAVRs. I gave the students tasks that specifically asked them to represent their preexisting knowledge first, and then to revise after the learning activity. The collected data included student performances on a card sorting task, both in their hypotheses setup and revision (which later were compared to each other), focus group interviews, post-instructional long-term effect surveys, and student background surveys. The same student tasks were applied to collect data for all research questions. Most of the tasks were embedded in WIIS, except for the focus group interview and the post-instruction long-term effect test. See Table 4.9 for a summary of the target data to be collected and corresponding student tasks. In the following, I discuss each student task in detail in the same order in which they were carried out during the instruction.

Preexisting Knowledge Survey

At the beginning of the instruction, the students were asked to answer a survey that aimed to check how much they already knew about the sizes of imperceptible objects (see Figure 3.6 in Chapter 3) before beginning the learning activity. The purpose was to see if the students were at least familiar with the objects and if they knew the objects were smaller than the size of a pin. I asked the students to raise their hands if they answered "no" (meaning they either had not heard of the name of the object or thought it was bigger than the size of the pinhead) to any one of the questions in the survey. If someone answered "no," I was going to provide additional instruction about the object to the students without giving any other information regarding the size than the fact that it is smaller than the head of a pin. This survey was conducted in order to attain information regarding the individual students' performances. The students' answers to the survey questions were automatically coded by a hidden program in WIIS and were transferred to a database via the Internet.

Data to collect	Student task	Description
To track how the students' mental models of the range of imperceptible sizes changed.	Card sorting task 1 & 2	*Card sorting task 1:* Card sorting task 1 is for representing the students' preexisting knowledge that has not yet been influenced by the learning activity with the TAVRS. Before starting the learning activity with the TAVRS, students first order a set of imperceptible objects by size from the smallest to the biggest. Then they classify the ordered objects into groups of objects with similar sizes, maintaining the order of the objects. Finally students are asked to give names to the groups they have created. The labels of the groups must be about the size, not the shape, color, or roles. *Card sorting task 2:* Later, for card sorting task 2, students observe the accumulation durations that are shown in the TAVRs (and other relevant multimodal experiences that are embedded in the TAVRs) and make changes to their work in card sorting task 1.
To understand how the students interpreted their temporal experiences and constructed the mental models.	Prediction and reflection	*Prediction:* Students predict the accumulation durations of each imperceptible object in WIIS. *Reflection:* Students later compare their predictions and the actual accumulation durations that are shown in each TAVR. Then they reflect their thoughts on a seven-point Likert scale that ranges from "extremely smaller than I thought" to "extremely bigger than I thought."
To gain deeper comprehension of the students' TAVR experiences.	Focus group interviews	Randomly selected students are interviewed regarding the logic behind their work on card sorting task 1 and 2 and their prediction/reflections.
To understand the context of the student performances.	Preexisting knowledge survey (and student background information)	Students answer a brief survey that asks if they have heard of the names of the imperceptible objects that are presented in WIIS
To examine what kinds of long-term effects the learning activity with TAVRs have had on the students' mental models.	Post-instruction long-term effect test (card sorting task)	Students perform another round of the card sorting task.

Table 4.9. Summary of the data to be collected and corresponding student tasks.

Card Sorting Task 1 and 2

The students conducted the card sorting tasks twice; once before they started the learning activity with TAVRs (card sorting task 1) and once after they interacted with TAVRs (card sorting task 2):

- *Card sorting task 1*: In WIIS, after the preexisting knowledge survey, but before they were introduced to TAVRs, the students were asked to sort the cards based on what already they knew about the sizes of the objects.
- *Card sorting task 2*: After the students were introduced to what TAVRs are and interacted with the TAVRs of the imperceptible objects, they modified their card sorting work from card sorting task 1. They were presented with their previous work, in WIIS, and could rearrange the cards in drag-and-drop fashion. They were allowed to change the labels that they originally gave for each object group.

The student performances on card sorting tasks 1 and 2 were automatically codified by the hidden codes in WIIS. The coordinates of the cards, the number of the student-created object groups, and what the students typed for the labels of each group were automatically transferred to a database over the Internet and stored under each student's name. The card sorting task was embedded in WIIS. A brief description of the card sorting task was provided in Chapter 3. Here I present it again, with additional discussion of how it can be useful for assessing the changes in the students' mental models. The imperceptible objects were the same ones that were introduced in Chapter 3. See Table 4.10 for a list of the objects, their sizes, and the accumulation durations.

Scale	Object	Size	Total accumulation time
Sub-nano	Hydrogen atom	~0.1 nm	11 d 13 h 46 m 40 s
	Water molecule	~0.3 nm	2 d 7 h 33 m 35 s
Nano	DNA helix	~2 nm	13 h 53 m 20 s
	Rhino virus	~25 nm	1 h 15 m 45 s
Micro	E coli bacterium	~2 μm	50 s
	Red blood cell	~6 μm	16 s

Table 4.10. The imperceptible objects that were used for this study.

The card sorting task was composed of the following three sub-tasks:

(1) *Ordering:* Students order the imperceptible objects by size – from the smallest to the biggest. This sub-task would let me know if the students correctly understood the most fundamental information of the TAVR – the inverse relationship between the object size and the accumulation duration. Also, ordering the objects from the smallest to the biggest must be done first before students can start classifying the objects by similar sizes (the grouping task).

(2) *Grouping:* After ordering the objects, students classify the objects by similar sizes and create groups, maintaining the order of the objects unchanged. The number of the size groups can vary from 1 to 6. As discussed in Chapter 2, conceptual size categories are one of the key components of the verbal information of a mental model of size. The way students form groups of the objects with similar sizes will reveal their mental model of the range of imperceptible sizes. More size groups will mean a more refined mental model of the range of imperceptible sizes. I contextualized the grouping task for the students by telling them, *"If there were a grain of rice, an ant, a truck, and a bus, how would you group these objects by similar size? [Wait a while and listen to the participant's answer] Yes. You would*

group the rice and ant into one group of objects with similar sizes and the truck and bus into a group of objects with similar sizes. This is what I mean by grouping by similar sizes." However, the number of the size groups can only show the number of the conceptual size categories, not the names of the conceptual size categories. Hence, the following sub-task, labeling, was carried out to reveal the conceptual size category "tags."

(3) *Labeling:* Students give names to the size groups they created. For this task, students are provided with prompts (e.g., a set of adjectives – "small," "tiny," "mini," "teeny," and a set of adverbs – "extremely," "very," "strikingly," "surprisingly"). Students can also use any additional vocabulary that they can recall. Earlier in Chapter 2, I argued that temporal experience is perceived through temporal modality, but it does not seem to be re-represented into a mental image, because the literature on mental images does not include it as a mental image. Rather, according to researchers, a temporal experience is interpreted verbally, and it is also encoded into verbal information in mental models, in addition to conceptual size categories. Furthermore, there is a possibility that the number of the conceptual size categories stays the same but the "tags" become more refined. Likewise, I cannot assume that the number of the size groups will show how the students' conceptual size categories are constructed. Therefore, the labeling task will reveal whether the students in fact have refined their mental models. The student-generated labels for each student-generated group will reveal the conceptual range of imperceptible sizes – the actual gap between each group – in the students' mental models.

Prediction and Reflection

Before the students began playing all TAVRs of the imperceptible objects in WIIS, they were asked to predict the accumulation durations for each imperceptible object (see Figure 3.13 in Chapter 3), and then later they were asked to compare the differences between their predictions and the actual results (the reflection task). To support the students in making their predictions, I provided them with a sample TAVR of a macroscopic object – hair. Hair is an object whose size is familiar to students; they have their own visual mental image of hair and can comfortably describe the size of hair. The thickness of hair is about one hundred micrometers, which is ten times smaller than the diameter of the head of a pin. Therefore it takes one second to accumulate strands of hair across a pinhead at the interval of ten objects per second, and ten seconds at the interval of one object per second.

However, predicting the accumulation durations is, of course, beyond the students' cognitive capacity because the students do not know the sizes of the objects, and even if they knew the sizes (absolute or relative), they would not be able to mentally calculate the accumulation durations. However, if a student can infer that the objects are too small to see and, hence, smaller than the thickness of hair, he or she will expect that the accumulation durations of the objects will take longer than the time it takes for hair. Then, depending on their concept of the range of imperceptible sizes, they are supposed give the longest prediction that they can conjure for the smallest object that they hypothesized. During the instruction, I specifically told the students to make the predictions using many units of time, not using seconds only. After the students interacted with TAVRs, they had to compare their predictions and the actual accumulation durations of each object. The students were asked to reflect on a seven-point Likert scale that ranged from “extremely smaller than I thought” to “extremely bigger than I thought.” See Figure 3.15 in Chapter 3 for a screen capture of the page that students used for this task. See Figure 4.2 for an excerpted example of the seven-point Likert scale.

Hydrogen atom

Extremely smaller than I thought	much smaller	somewhat smaller	similar with what I thought	somewhat bigger	Much bigger	Extremely bigger than I thought

|--------------|---------------|---------------|---------------|---------------|---------------|

Figure 4.2. An example of seven-point Likert scale that was used for the reflection task.

The purpose of this predict-and-reflect task was to observe how the students reflected on the differences between their predictions and the actual accumulation durations of the TAVRs. The students’ reflections will show the way the students interpret and reflect on the differences. I also intended to help them contextualize their learning experience by making connections between their prior knowledge and new learning because students learn most effectively when they can construct new knowledge upon the preexisting knowledge in their memory (Piaget et al., 1977). The student predictions for each object were automatically sent to and saved in a database.

Focus Group Interviews

Focus group interviews with randomly selected students were conducted. Due to the characteristics of the data collection environment, where I was not able to personally interview individual students one after another[5], I collected masses of data from groups of students and then conducted focus group interviews for each research question, in order

[5] The students were recruited through their science teachers. A normal learning activity session with WIIS takes about 30 minutes. If the students were met for individual interviews and observations, they would have to miss two thirds of one science class.

to gain additional in-depth insights regarding the logic behind the students' interactions with TAVRs.

The focus group students were randomly selected right after the instruction in the manner that was explained in the previous section, where I introduced the participants of this study. The instruction was designed to be finished about 7-10 minutes before the end of the class. During the leftover time, I played a video that shows the range of size and scale of the universe (which is the opposite phenomena from the imperceptible scale) called *Powers of Ten* for all of the students, and then went to the selected students' seats to individually and quickly ask the focus group interview questions while the rest of the students were watching the video.

The questions were asked using the stimulated recall method. Stimulated recall is a research method that allows the investigation of cognitive processes by re-inviting participants to recall their concurrent thinking during an event when prompted by visual recall. It is a subset of introspective research methods that accesses participants' reflections on mental processes (Lyle, 2003). During the stimulated recall, the student's work was still shown on the computer; therefore, I could easily move between the student task pages in WIIS while asking questions to the student. I wrote down the student responses with the students' names. I could not record the audio of the interview because of the loud sound of the video that was being played in the room. I emphasized to the students that the questions were irrelevant to the correctness or incorrectness of their responses in the tasks, and I was not judging them.

The questions were designed following the framework of construct-centered design (Pellegrino, Krajcik, Stevens, & al., 2008). Construct-centered design allows one to develop both assessments that embody desired student achievements and instructional materials that aim at achieving these outcomes. To follow the construct-centered design framework, one has to create a set of claims, evidence, and tasks that support the expectations regarding student achievements. First, the claim is the cognition that students should apply to the content under consideration. Second, evidence is what the researcher will accept as a demonstration that the claim has been satisfied. Third, the task is the instantiation of the evidence statement. See Table 4.11, Table 4.12, and Table 4.13 for the claims, evidence, and tasks that were constructed for the card sorting, prediction, and reflection tasks. Since the main focus of the research questions is on the changes in the students' mental models in response to different types of TAVRs, the questionnaires (tasks) were designed to address these changes. A focus group interview could not be conducted for the post-instruction long-term effect test due to the limits of the allowed time.

Additionally, I asked the students a question that aimed to probe the influence of the "nothing visible on the pinhead" period. I wanted to check whether the students developed a visual mental image of "the un-seeable because it is too small to see." I asked them, "When you think about the SIZE of an atom, NOT the color or shape of it, what pops up that you see in your head?" However, in some cases, not all questions were covered due to the limits in the allowed time.

Card sorting task		
Claim	**Evidence**	**Task**
Students are able to describe their previous idea (preexisting knowledge) of the range of imperceptible sizes that are shown in card sorting task 1 and contrast it with their learning experience to explain how they revised their work in card sorting task 2.	When provided with their work on the card sorting task (in both the hypotheses setup and revision), students can clearly explain the rationale behind their work in a logically consistent manner, making connections with the temporal experiences that they had with the TAVRs.	Students are asked to explain the rationale behind the way they ordered, grouped, labeled the objects. *Questionnaires:* "For what reasons did you order (group, label) the object this way?" "What do these names of the groups mean?" "Please explain to me why you used [the words] for this group but not for the other groups."

Table 4.11. The claim, evidence, and task construct for the card sorting task.

Prediction		
Claim	**Evidence**	**Task**
Students are able to explain the rationale behind the way they made predictions of the accumulation durations for each imperceptible object.	When shown to their predictions of the accumulation durations for each imperceptible object, students can provide logically consistent rationale for their predictions, particularly focusing on the passage of time for the accumulations.	Students are asked to explain the logic of their predictions.

Table 4.12. The claim, evidence, and task construct for the student predictions for the accumulation durations of each imperceptible object.

Reflection		
Claim	**Evidence**	**Task**
Students are able to explain the rationale behind the way they reflected on the differences between their predictions and the actual accumulation durations on a seven-point Likert scale.	When re-presented with their reflections, students can provide consistent explanations of the logic behind their predictions by referring to the temporal experience they had with the TAVRs.	Students are asked to explain based on what criteria they have made the reflections on the scale.

Table 4.13. The claim, evidence, and task construct for the student reflection on the differences between their predictions and the actual accumulation durations.

Post-Instruction Long-Term Effect Test

Lastly, I met the students again about three to six weeks after the instruction to conduct surveys on how much they remembered the learning activities. If a representation were effective in helping students to build a mental model for a scientific phenomena, one would expect students to retain this information beyond a short period of delay (Rapp, 2005). An analysis of student knowledge months after the lesson would reveal how they encoded the representations that they interacted with, as a function of the learning experience.

The initial intention was to go back to the classes after five or six months; however, this seemed logistically unfeasible. With the start of a new school year, one of the teachers was going to take a maternity leave and the other two teachers were not going to teach the same students. Hence, I had to conduct the test within a limited window of time. With the cooperation of the science teachers, I once again went into the science classes when the students were using computers in a computer lab.

A Flash-based test application that can be accessed using a web browser for this test was created. It required the students to enter their names and complete the card sorting activity with the same imperceptible objects. It also asked them to recall and write the total accumulation durations of each object, as they remembered. See Figure 4.3 for a screen capture of the post-instruction long-term effect test. The student responses were automatically transferred to the database and saved.

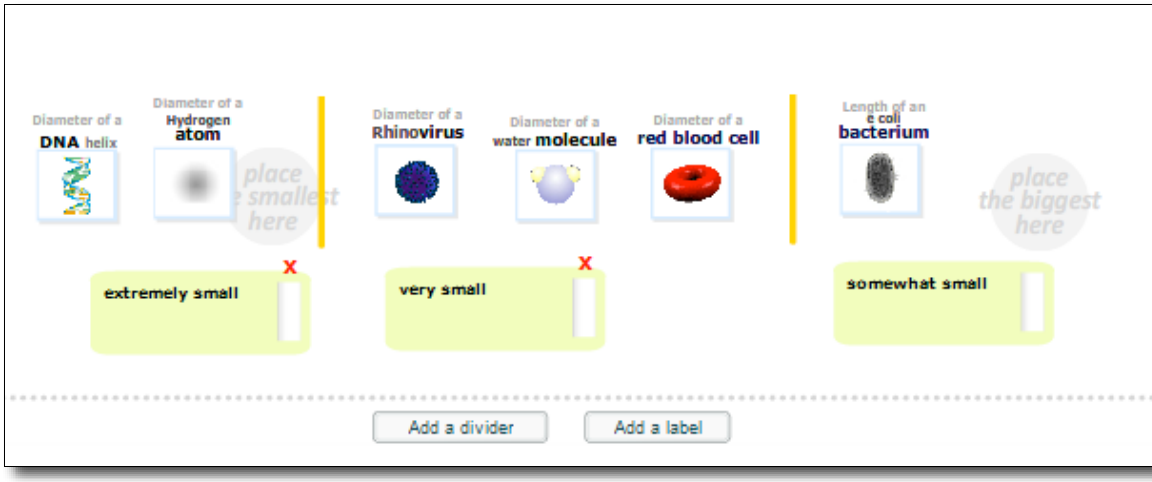


Figure 4.3. The screen capture of WIIS that was used for the post-instruction long-term effect test.

Summary

See Table 4.14 for a summary of the student tasks, the corresponding purposes, and the forms of the data collected. I also created a diagram (see Figure 4.4) of the flow of the student tasks to help readers more conveniently comprehend the tasks.

Order	Task	Purpose	Data collected
1	Take a preexisting knowledge survey	To check if students know that the provided objects are smaller than the size of the head of a pin.	Quantitative: Student yes/no responses, codified, transferred to and saved in a database. Qualitative:
Card Sorting Task 1: Hypotheses setup			
2	Represent preexisting knowledge through card sorting task	To track the changes in students' mental model (specifically, conceptual size categories) of the range of imperceptible sizes.	Student performances on the card sorting task, codified, transferred to and saved in a database.
(Between 2 and 3, students are introduced to sample TAVRs)			
3	Predicting the accumulation durations	To see later (at 5) how students reflect on their temporal experience with TAVRs by comparing their predictions and the actual accumulation durations.	Student predictions written in units of time, codified, transferred to and saved in a database.
(Between 3 and 4, students are introduced to sample TAVRs)			
Card Sorting Task 2: Revision			
4	Revising the prior card sorting task based on the TAVR results	To assess how students' mental model (the conceptual size categories) of the range of imperceptible sizes changed after the learning activity with TAVRs.	Student revisions on the card sorting task, codified, transferred to and saved in a database.
Reflection			
5	Reflecting on the difference between the predictions and actual accumulation durations	To reveal how interpret the passage of time. And also to see if their card sorting task was carried out in a consistent manner regarding the way they interpret the temporal experiences.	Student reflection on the seven-point Likert scale, codified, transferred to and saved in a database.
6	Focus group interview (with randomly selected students)	To gain insights into the logic behind the students' work and their learning experiences.	Student responses to the interview questions and their names written on a note.
7	Post-instruction long-term effect test (with reduced number of students)	To see how much influence the TAVRs had on the student mental models of the range of imperceptible sizes.	Student names and their responses to a card sorting task, codified, transferred to and saved in a database.

Table 4.14. The order of the student tasks and the kinds of data collected.

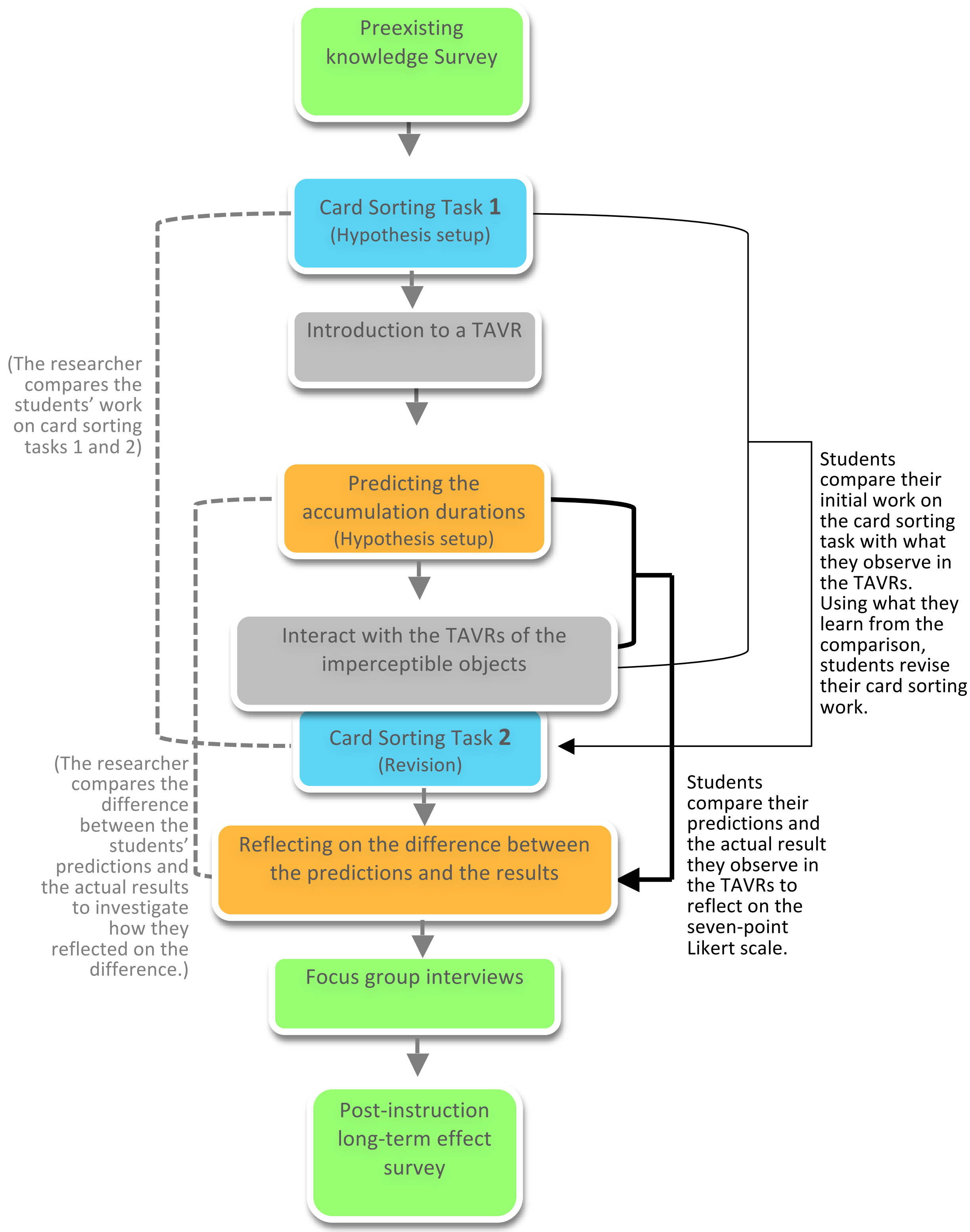


Figure 4.4. A diagram of the phases of the instruction with additional information about what the student did, what data the researcher compared in the analysis.

In Figure 4.4, the boxes with non-gray colors represent the phases where the data were collected. Specifically, the blue boxes and orange boxes represent the main test methods for assessing the changes in the students' mental models: card sorting tasks 1 and 2 and the prediction/reflection. Green boxes represent the data that were collected to attain additional context or clues for interpreting student answers. Gray boxes represent the phases during which data collection did not take place.

In the following, I explain each TAVR that was modified for the treatment conditions for each research question.

Research Question 1

The primary goal of this research question is to identify how to best augment the temporal experience by investigating the changes in the students' mental models of the range of imperceptible sizes through the learning activity with temporal representations with different augmentation features. I created four different types of modified TAVR: TAVR (original), TVR (temporal-visual), TAR (temporal-aural), and TR (temporal only).

- *TAVR*: The students in the TAVR group interacted with original TAVRs that have all three modalities.
- *TVR*: The students in this group were given TVRs, TAVRs without the aural component (the clicks).
- *TAR*: In this group, students interacted with TARs (TAVR without the visual – the red line) throughout the learning activity.
- *TR*: The students in this group used the temporal representation without any supporting modalities (hence, TR). There was no explicit aural or visual representation on the pinhead; there were only the head of a pin and the clock that

showed the elapsed time. To make it fair with the other groups, the TR group also was provided with ">> Skip ahead" button for each TAVR. The completions of the accumulations in each TAVR were indicated both by the stopping of the clock and a blue square that appears above the pinhead.

Research Question 2

Research question 2 concerns the influence of the length of the interval of the sequential action (the accumulation) on the way the students interpret their temporal experiences.

- *Extended*: The students in this group were provided with TAVRs that placed one object every one second.
- *Compressed*: In this group, TAVRs accumulated one object every 0.1 seconds.

Research Question 3

This research question aimed to compare three different manipulations of the temporal experience.

- *No-manipulation (NM)*: For the no-manipulation condition, the students were supposed to sit in front of the computers and "just watch" the manipulation progressing (at the rate of one object in 0.1 seconds). I initially planned to set up a computer station in the back of the science classroom, but it could not be carried out due to inconveniences (i.e., students might get distracted by the computer during other classes) and other concerns, such as the security of the computer. Then I attempted to visit the school every day with a laptop computer to show the students how the accumulations had been progressing. This plan, however, also turned out to be impractical when I realized that the students could not observe the accumulation of four objects out of six being accomplished because they were going to happen when the students were not in the science class or in the middle of the night, which meant that the students were going to be verbally informed of the completion of the accumulations the day after the completion of the accumulation. Then, the students' temporal experiences would not precisely be "natural."

It was difficult to find imperceptible objects that the students might be familiar with or find easy to understand after a brief introduction. Hence, the imperceptible objects that were used for this research question are one object less than six objects. They are: skin cell, red blood cell, e-coli bacterium, mitochondria, and influenza virus ("flu virus"). See Table 4.15 for their sizes and accumulation durations. These objects were used for all of the student groups that were formed for this research question. The teacher informed me that the only object the students were not familiar with was mitochondria. Hence, I gave brief lecture on the objects to the students before the learning activity with WIIS. After the learning activity with these objects, just in order to provide the students with a learning experience about the fuller range of imperceptible sizes, I introduced them to the accumulation durations of the other objects that were originally included in WIIS: hydrogen atom, water molecule, DNA helix, and Rhinovirus.

Object	Size	Accumulation duration
Skin cell	30 micrometers	3 seconds
Red blood cell	8 micrometers	13 seconds
e-coli bacterium	3 micrometers	33 seconds
Mitochondria	1 micrometer	1 minute 40 seconds
flu virus	130 nanometers	12 minutes 49 seconds

Table 4.15. The objects that were used for research question 3 and their sizes and accumulation durations.

- *Automatic manipulation (AM)*: The accumulations in the TAVRs for this group were automatically skipped ahead (by which I mean "accelerated"). The students watched the TAVRs accumulate the objects without having any interactive input into the accumulations. There appeared a red dot above the image of the object, in the TAVR, whenever the acceleration had happened. The objects that the NM group students used were provided to the AM group students as well.
- *Interactive manipulation (IM)*: For this research question, the number of the skipped objects was set to ten objects at one button press because the longest accumulation duration was only about thirteen minutes (flu virus). If it had been set to 1,000 objects, as in other TAVR versions, the students would have finished the accumulation of influenza virus by pressing the "Skip ahead" button only a couple of times. The objects that the NM group students used were provided to the IM group students as well.

Data Analysis

As I collected multiple forms of data, both qualitative and the quantitative, I associated the qualitative analysis with quantitative measures. The data from the students' work on the card sorting tasks (1, 2, and 3 - the post-instruction long-term effect test) and the reflections were quantified by coding rubrics. The focus group interviews were both qualitatively and quantitatively analyzed. The students' preexisting knowledge survey and background information were used to gain a deeper comprehension of their answers. I elaborate on the details of the analysis of the data in the following.

Card Sorting Tasks

The main goal of having the students to do the card sorting tasks was to assess the changes in the students' mental models. For this, I looked into (1) the students' responses in card sorting task 1 during their hypothesis setup, (2) their work on card sorting task 2, which was their revision of their card sorting task 1 work, and (3) the difference between (1) and (2). In this way I could assess (a) which student group performed the best after the learning interventions, and (b) which student group had the greatest increase in performance, which indicates that the refinement of their mental models of the range of imperceptible sizes has occurred.

Although the students who participated for each research question were recruited from the same school and the same teacher, I could not assume that the students in each treatment group shared the same level of preexisting knowledge. Then the highest scores of the card sorting tasks could not guarantee the most increase in the card sorting task scores. If the best group from analysis (a) and (b) were the same, it would mean that the type of TAVR that the group used was effective. If the result showed a disparity between (a) and (b), it would call for another layer of analysis. For example, if the group who ranked the highest in the analysis (a) did not ranked the highest in (b), I would have to see if the group scored significantly higher than the other groups on card sorting task 1. These analyses, which excluded card sorting task 1, were conducted for all three sub-tasks of card sorting task: ordering, grouping, and labeling. See Table 4.16 for a summary of the quantifying scheme for the data from each card sorting sub-task and the statistical analysis methods employed.

To analyze **ordering** tasks 1 and 2 (in card sorting tasks 1 and 2), I developed a coding rubric. See Table 4.17 for the coding rubric I constructed for the ordering tasks. The score varied between 0 and 3. After coding the data from both card sorting task 1 and task 2, I statistically analyzed them following the hypotheses.

Sub-task	**Quantifying scheme**	**Statistical analysis**		
		RQ 1 (four groups)	RQ 2 (two groups)	RQ 3 (three groups)
Ordering	Used a coding rubric. See Table 4.17.	One-Way between subjects ANOVA with a contrast subcommand (a priori contrast)	One-Way between subjects ANOVA	One-Way between subjects ANOVA with a contrast subcommand (a priori contrast)
Grouping	Counted the number of the student-created object groups.			
Labeling	Developed a coding scheme to quantify the student responses.			

Table 4.16. The summary of the quantifying scheme for the data that were collected from each sub-task of the card sorting task and the employed statistical analysis.

Novice	**Apprentice**	**Semi-veteran**	**Veteran**
Ordered all objects incorrectly.	Ordered 3-5 objects incorrectly	Ordered 1-2 objects incorrectly	Correct order of the objects
(0 pt.)	(1 pts.)	(2 pts.)	(3 pts.)

Table 4.17. The coding rubric constructed for the ordering task.

The students' object **groupings** were measured by counting the number of the student-generated size groups because creating more groups indicates having a more refined mental model of the range of imperceptible sizes. If the number of the student-generated size groups became bigger after the learning intervention with the TAVRs, it would mean that the conceptual size categories in the student's mental model had been refined to a finer set.

The **labeling** tasks were assessed by statistical analysis of the data that were quantified by a coding rubric. I not only compared the changes in the number of the descriptive words that were used to represent the smallest group in card sorting tasks 1 and 2, but also compared the changes in the difference between the smallest group and the biggest group. I specifically asked the students to use expressions such as "very small" or "extremely small," not simply "the smallest," or, in particular, "biggest." If the students used such words they would not have to use descriptive words (e.g., "very," "extremely") in order to represent the sizes of the groups that they formed. However, in the pilot test, I noticed that about a half of each student group used the word "smallest" or "biggest." Based on this observation, in every real instruction session, I heavily emphasized to the students that they must not use "smallest" or "biggest." However, unfortunately, some students still used it. Hence, I developed a coding rubric to objectively and quantitatively analyze their responses (see Table 4.19).

To develop coding rubrics, I followed the coding rubric guidelines suggested by Stix (1996):

- Decide whether the rubric addresses the most important aspects of student performance.
- Decide whether or not the rubric addresses the instructional outcome(s) to be measured.
- Decide whether the rubric includes anything extraneous. If so, change the rubric or use a different one.

- Don't pay too much attention to the rubric's stated grade level. It may be usable at other grades with little or no modification
- See if a rubric from a different subject area can be adapted to fit your needs.
- Make sure the rubric is clear.
- Try the rubric out on some actual samples of student work.
- See if you and your colleagues can usually arrive at consensus about what scores to assign a piece of student work (Stix, 1996).

In addition to these principles, I also adopted Stix's coding rubric development table (see Table 4.18), which allows a researcher to judge whether the rubric addresses the most important aspects of student performance and the instructional outcome(s) to be measured.

Criteria	Novice	Apprentice	Veteran
Logic	Vague and unclear (0 pt.).	Some focus, but not organized enough (1 pt.).	Well organized and clearly presented (2 pts.).
Content	Incorrect or few facts, hardly any detail (0 pt.).	Some facts are accurate, some detail (1 pt.).	Substantial amount of facts, good amount of detail (2 pts.).

Table 4.18. Framework of coding scheme proposed by Stix (1996).

The coding rubric for the labeling task examined two aspects of the student responses: logic and content (see Table 4.19). The logic aspect inspected the students' intention of explicitly distinguishing different size groups in the labels. The content aspect addressed the quantity of the descriptive words that individual students used, which reflected the verbal information of the students' mental models of imperceptible sizes. As the coding rubric examines the students' explicit intention of discriminating between different size groups and how they described the smallest and biggest groups, the scores are naturally lower for the students who created a smaller number of size groups. In the analysis, I noticed that the students used other methods to differentiate the sizes of the groups by choosing different descriptive words. For example, some students labeled their biggest object group as “somewhat small” and smallest object group as, “extremely small.” In another case, a student labeled the biggest as “*big daddy donuts*” and the smallest group as “*cute baby nuggets*.” This shows that although they did not use descriptive words additively as I instructed, they clearly intended to represent the difference between the smallest and the biggest. See Table 4.20 for examples of various student responses in the labeling task and how I coded them.

Criteria	0 point	1 point	2 point
Logic	No intention of expressing the difference between size groups.	Exhibits an intention of expressing the difference between size, but the labeling scheme is NOT consistent across the size groups.	Exhibits an intention of expressing the difference between sizes, and the labeling scheme is consistent across the size groups.
Content	• **+1 point** for each extra decoration: - additive adverbs - exclamation mark - capital letters (when the intention was obvious) • These extra points were counted when: - Explicitly more decorative words were used for the smallest group than for the biggest group.		

Table 4.19. A coding rubric for quantifying the student responses to the labeling task. The scores from the logic aspect and content aspect are combined to produce the total score.

These data were statistically analyzed to test the corresponding hypotheses. As summarized in Table 4.16, the data for research questions 1 and 3, which had three or more treatment groups, were analyzed by One-Way between subjects ANOVA with a contrast subcommand following my hypotheses, which produced a t-test version of the planned comparison. The data for research question 2 was analyzed by One-Way between subjects ANOVA. In the next section, I discuss the specifics of the analysis methods I employed for each research question.

Examples				Score		
Student (N. of size groups)	**Smallest group**	**Interim group**	**Biggest group**	**Logic**	**Content**	**Total score**
A (3)	very very small	very small	small	This student explicitly used a different number of descriptive words for each size group that he created. The represented sizes of each group are consistently distinguishable by the number of uses of the word "very." 2 points	"very" (+1) "very" (+1) 2 points	4
B (3)	SMAAAAAAALLL!	SMALL	small	Student B actively used capital letters for the medium group and the smallest group. For the smallest group, she even added more letters to the word and an exclamation mark at the end. The intention of distinguishing different size groups is clear and consistent. 2 points	The use of the capital letter (+1) and an exclamation mark (+1) 2 points	4
C (2)	cute baby nuggets	-	big daddy donuts	This student exclusively came up with his own terms for representing his size groups. Although it is metaphorical, the intention of discriminating two different size groups is clear and consistent. 2 points	"cute baby" (+1) 1 point	3
D (3)	very tiny mini cuties	tiny mini cuties	mini cuties	Student D also used unique words, rather than the vocabulary set that I provided, for representing different size groups. 2 points	"very" (+1) "tiny" (+1) 2 points	4
E (1)	tiny	-	-	Student E generated only one group, meaning that he thought that all objects had similar sizes. There was no intention of discriminating between different sizes. 0 point	"tiny" (+1) 1 point	1

Table 4.20. Examples of the student responses for the labeling task and how they were coded.

- *Research Question 1*: Previously, in the discussion of research question 1, I hypothesized that the students in the TAVR group would achieve the highest scores in card sorting task 2 (and also the largest increase from card sorting task 1) because their temporal experiences of the imperceptible accumulation is supported by both visual and aural representations. The students in the TVR group and the TAR group were predicted to score the next highest for they have only one supplemental modality for visualizing the accumulation in their working memory. Finally, the TR group was expected to rank the lowest because they did not have any support for visualizing the accumulation of the imperceptible objects other than the clock. To test this specific one-tailed hypothesis, I conducted an a priori contrast analysis; (1) TAVR > TVR, (2) TAVR > TAR, (3) TVR > TR, and (4) TAR > TR. To obtain this contrast, I ran a One-Way between subjects ANOVA with a contrast subcommand, which produced a t-test version of the planned comparison using SPSS 19.
- *Research Question 2*: Previously, I hypothesized that the extended group would outperform the compressed group because the extended temporal experience results in more drastic differences between the accumulation durations of the objects, and learners might create more groups of objects that are classified by similar sizes. In order to test this hypothesis, the data were quantitatively analyzed using One-Way ANOVA to test for the statistical significance between two groups using SPSS 19.
- *Research Question 3:* I hypothesized that the group of students who interacted with the TAVRs with an interactive manipulation feature (IM group) would score the highest on card sorting task 2 because the temporal and kinesthetic experiences are expected to be re-represented in a kinesthetic mental image, which can be

"interwoven" with verbal information. Then I expected that the students who were not provided with the feature for manipulating the passage of time – no manipulation feature (NM group) - would rank the second on card sorting task 2. Lastly, the group of students who interacted with the TAVRs with an automatic manipulation feature (AM group) would perform the poorest. Hence, the hypothesis can be summarized as, IM > NM > AM. To test this specific one-tailed hypothesis, I ran an a priori contrast analysis: (1) IM > NM and (2) NM > AM. To obtain this contrast, I conducted a One-Way between subjects ANOVA with a contrast subcommand, which produced a t-test version of the planned comparison using SPSS 19.

Seven-Point Likert

The students' reflections on the seven-point Likert scale were turned into numbers, ranging from 1 being "extremely bigger than I thought" to 7 being "extremely smaller than I thought."

Focus Group Interviews

The data from the focus group interviews were analyzed by incorporating both quantitative and qualitative methods. In addition to examining how the students performed, it was necessary to check whether their logic was consistent across their answers. If the answers were created on the spot, it would be meaningless to look closely into their answers. To analyze the consistency in the students' answers in relation to their performances, I referred to the coding rubric framework suggested by Stix (1996) again. See Table 4.21, Table 4.22, and Table 4.23 for the coding rubrics for the three sub-tasks of the card sorting task.

Criteria	Novice	Apprentice	Semi-veteran	Veteran
Content	Ordered all objects incorrectly. (0 pt.)	Ordered 3-5 objects incorrectly. (1 pts.)	Ordered 1-2 objects incorrectly. (2pts.)	Correctly ordered all six objects. (3 pts.)
Logic	Uses irrelevant parts of the TAVRs to support the rationale. (0 pt.)	Inconsistently uses the closely relevant components of the TAVRs or not-so-relevant components of the TAVRs to support the rationale. (1 pts.)	Inconsistently uses the closely relevant components of the TAVRs or not-so-relevant components of the TAVRs to support the rationale. (1 pts.)	Refers to the learning experience with the closely relevant components of the TAVRs to justify their rationale. (2 pts.)

Table 4.21. The coding rubric for focus group interviews regarding the ordering task.

Criteria	Novice	Apprentice	Veteran
Content	Created only one group. (0 pt.)	Created two groups. (1 pts.)	Created more than three groups. (2 pts.)
Logic	Uses irrelevant parts of the TAVRs to support the rationale. (0 pt.)	Inconsistently uses the closely relevant components of the TAVRs or not-so-relevant components of the TAVRs to support the rationale. (1 pts.)	Refers to the learning experience with the closely relevant components of the TAVRs to justify their rationale. (2 pts.)

Table 4.22. The coding rubric for focus group interviews regarding the grouping task.

Criteria	Novice	Apprentice	Veteran
Content	Used only one descriptive word to name the smallest group. (0 pt.)	Used two descriptive words to name the smallest group. (1 pts.)	Used three or more descriptive words to name the smallest group. (2 pts.)
Logic	Uses irrelevant parts of the TAVRs to support the rationale. (0 pt.)	Inconsistently uses the closely relevant components of the TAVRs or not-so-relevant components of the TAVRs to support the rationale. (1 pts.)	Refers to the learning experience with the closely relevant components of the TAVRs to justify their rationale. (2 pts.)

Table 4.23. The coding rubric for focus group interview regarding the labeling task.

Post-Instruction Long-Term Effect Test

This test was statistically analyzed with the hypothesis composition, similar to the analysis of the card sorting tasks, because I believed that newly learned knowledge could be better remembered when built from a larger number of constituents. For the card sorting tasks, I predicted that the students who interacted with the TAVRs with more modalities would make more meaningful interpretations of their learning experiences because having more constituents, in diverse modalities, results in a more contextualized and meaningfully organized mental model. See Table 4.24 for the data collected and the corresponding analysis methods for each research question.

<table>
<tr><th rowspan="2">Card Sorting Sub-task</th><th rowspan="2">Quantifying Scheme</th><th colspan="3">Statistical analysis</th></tr>
<tr><th>RQ 1</th><th>RQ 2</th><th>RQ 3</th></tr>
<tr><td>Ordering</td><td>Used a coding rubric.</td><td rowspan="3">One-Way between subjects ANOVA with a contrast subcommand (a priori contrast)</td><td rowspan="3">One-Way between subjects ANOVA</td><td rowspan="3">One-Way between subjects ANOVA with a contrast subcommand (a priori contrast)</td></tr>
<tr><td>Grouping</td><td>Counted the number of the groups.</td></tr>
<tr><td>Labeling</td><td>Counted the number of the descriptive words and exclamation marks.</td></tr>
</table>

Table 4.24. The summary of the quantifying schemes for the data that were collected from each sub-task of the card sorting task, and the employed statistical analysis.

In this chapter, I discussed the research questions, overall approach, research design, research context, data collection and methods of analysis. In the next chapter I present the results of this study.

Limits of the Study

As discussed earlier, the study is quasi-experimental because the participating students were not randomly selected. Causal inferences in this study were established under the conditions that were specific to this study. The inferences may not apply to other situations, such as different types of participants, treatments, settings and measures. For example, a study using students with higher mathematical ability or richer background knowledge of the sizes of imperceptible objects may result in the construction of different mental models in students.

CHAPTER 5 RESULTS

In this chapter, I present the results of this study in response to each research question and corresponding hypothesis. The first section addresses research question 1, discussing the effects of different combinations of aural and visual modalities that were employed to augment temporal experiences. The second section addresses research question 2, which compared compressed and extended temporal experiences. Finally in the third section, I discuss the results of research question 3, which aimed to examine the different effects of different temporal manipulations on students' mental model refinements.

As discussed in Chapter 4, for each research question, I assessed the changes in the students' mental models of the range of imperceptible objects that were influenced by their interaction with different types of TAVRs by examining their responses to a survey, card sorting tasks 1 and 2, reflections on a seven-point Likert scale, focus group interviews, and post-instruction long-term effect test. In the following I present the results for each research question in this order, except the focus group interviews, student background information and their preexisting knowledge survey. I use these as supplements for understanding the context of their responses in depth within the discussion of the results of survey, card sorting tasks 1 and 2, and post-instruction long-term effect test (which I call card sorting task 3).

Research Question 1

How do the combinations of the supporting modalities support learners to construct the mental model of the range of imperceptible sizes?

Survey

The students were first asked to answer survey questions that aimed to check their familiarity with the names and sizes of the objects (see Figure 3.6 in Chapter 3). The two questions were: (1) if they had heard of the names of each object, and (2) if they thought the sizes of the objects were smaller than the size of the head of a pin.[6] See Table 5.1 for the summary of each student group's responses to the two survey questions. The number of the students' answers to the second question includes responses only from the students who answered "yes" to the first question. Hence, the total number of the student responses to the second question is equal to the number of the students who answered "yes" to the first question.

As the table shows, almost all students in all four groups had heard of the names of the objects; however, many students thought that the sizes of some objects were bigger than the size of the head of a pin. Specifically, almost 50% of the students (of all groups) thought that the sizes of red blood cells and DNA helixes were bigger than the size of the pinhead. Overall, at least about 20% of the students from all groups thought that the size of the each object was bigger than the pinhead. I conducted a One-Way ANOVA with the student responses to each object in order to see if there existed statistically significant differences between the student groups, specifically for the second survey question, because significantly poorer preexisting knowledge would influence the interpretation of the results from other student tasks. However, the results showed that each group's responses to each object were not significantly different ($p > 0.05$) between the student groups, specifically for the second survey question, because significantly poorer

[6] A real pin (poked into a piece of cork) was provided to each individual student to help him or her better understand the size of the head of a pin.

Question		Student responses (TAVR group) N=27					
		Atom	Molecule	DNA	Virus	Bacterium (Bacteria)	Red blood cell
Heard?	Yes	26	26	27	26	26	27
	No	1	1	0	1	1	0
Smaller than pinhead?	Yes	19	16	14	20	21	14
	No	7 (27%)	10 (38%)	13 (50%)	6 (23%)	5 (19%)	13 (48%)

a. TAVR group.

Question		Student responses (TVR group) N=26					
		Atom	Molecule	DNA	Virus	Bacterium (Bacteria)	Red blood cell
Heard?	Yes	25	25	25	26	25	24
	No	1	1	1	0	1	2
Smaller than pinhead?	Yes	19	14	12	20	20	14
	No	6 (24%)	11 (44%)	13 (52%)	6 (23%)	5 (20%)	10 (42%)

b. TVR group.

Question		Student responses (TAR group) N=26					
		Atom	Molecule	DNA	Virus	Bacterium (Bacteria)	Red blood cell
Heard?	Yes	25	26	26	26	25	26
	No	1	0	0	0	1	0
Smaller than pinhead?	Yes	20	14	14	20	20	14
	No	5 20%	12 46%	12 46%	6 23%	5 20%	12 46%

c. TAR group.

Question		Student responses (TR group) N=25					
		Atom	Molecule	DNA	Virus	Bacterium (Bacteria)	Red blood cell
Heard?	Yes	24	23	25	24	24	23
	No	1	2	0	1	1	2
Smaller than pinhead?	Yes	18	15	12	17	19	12
	No	6 25%	8 35%	13 52%	7 29%	5 21%	11 48%

d. TR group.

Table 5.1. The student groups' responses to two survey questions.

preexisting knowledge would influence the interpretation of the results from other student tasks. However, the results showed that each group's responses to each object were not significantly different (*p* > *0.05*).

Card Sorting Task 1 (Hypothesis Setup Phase)

Card sorting task 1 was given to the students with the purpose of detecting their preexisting knowledge regarding the range of imperceptible sizes, like a pretest. To summarize, most of the students performed poorly on card sorting task 1. No student correctly ordered the objects, the number of the size groups was mostly less than three, and the number of the descriptive words in the labels for the groups tended to be too concise.

Ordering

In the ordering task of card sorting task 1 (which I will call "ordering task 1"), the results showed that none of the students in any group ordered the objects correctly. Although there were some students who knew that the atom was the smallest object, they arranged the order of the other objects incorrectly. See Table 5.2 for the student distribution for each possible score of the ordering task. No difference between the student groups was found from the One-Way ANOVA ($p>0.05$).

Groups	Count of the student responses					
	0	**1**	**2**	**3**	**Mean**	**SD**
TAVR (N=27)	18 67%	6 22%	3 11%	-	1.44	0.70
TVR (N=26)	18 69%	5 19%	3 12%	-	1.42	0.70
TAR (N=26)	16 62%	7 27%	3 12%	-	1.50	0.71
TR (N=25)	17 68%	5 20%	3 12%	-	1.44	0.71
Total (N=104)	69 66%	23 22%	12 12%	-	-	-

Table 5.2. The number of the students for each score (0-3) of the ordering task on card sorting task 1.

Grouping

In the grouping task of card sorting task 1 ("grouping task 1"), most of the students overestimated the sizes of the imperceptible objects and underestimated the range of imperceptible sizes. See Table 5.3 for the minimum, maximum, mean, standard deviation of the results, and the count of the student responses for each possible number of the size groups. Many students (*N*=71, 68%) generated two size groups, and only a few of the students from each student group created three or four size groups. The one-way ANOVA did not show any significant difference between the groups ($p>0.05$).

Student Group (total N.)	**Count of the student responses**							
	1	**2**	**3**	**4**	**5**	**6**	**Mean**	**SD**
TAVR (N=27)	2 7.4%	17 63.0%	7 25.9%	1 3.7%	-	-	2.26	0.66
TVR (N=26)	2 7.7%	17 65.4%	7 26.9%	-	-	-	2.19	0.57
TAR (N=26)	2 7.7%	18 69.2%	6 23.1%	-	-	-	2.15	0.54
TR (N=25)	1 4.0%	19 76.0%	5 20.0%	-	-	-	2.16	0.47
Total (N=104)	7 7%	71 68%	25 24%	1 0.01%	-	-	-	-

Table 5.3. Frequencies of the students for each number of the size groups that the students (in each treatment group) generated for grouping task 1.

Labeling

Considering that the total score of each student's labeling task was generated by the sum of the scores of (1) the student's logic of expressing the differences between size groups (2 points maximum) and (2) the number of the descriptive words or expressions, such as exclamation marks and capital letters (see Table 4.19 in Chapter 4), the results indicated that the students did not much care to use descriptive words to express the size of the smallest group. To name the smallest group, the students tended to use one or two adverbs (e.g., "very," "super," "really," "extremely") to emphasize the adjectives that described the smallness (e.g., "small," "tiny," "mini"). Interestingly, some students used their own terms for the labels, as introduced in Table 4.20 in Chapter 4.

Student Group (total N.)	**Count of the students for each score**					
	1	**2**	**3**	**4**	**Mean**	**SD**
TAVR (N=27)	2 7.4%	12 44.4%	8 29.6%	5 18.5%	2.6	0.9
TVR (N=26)	2 7.7%	7 26.9%	14 53.8%	3 11.5%	2.7	0.8
TAR (N=26)	2 7.7%	9 34.6%	10 38.5%	5 19.2%	2.7	0.9
TR (N=25)	1 4.0%	6 24.0%	14 56.0%	4 16.0%	2.8	0.7
Total (N=70)	7 6.7%	34 32.7%	46 44.2%	17 16.3%	2.7	0.8

Table 5.4. Frequencies of the students for each number of the size groups that the students (in each student group) generated for labeling task 2.

Across all student groups, the number of the adverbs that the students used to name the smallest group tended to be highly related to the number of the object groups. For example, if a student formed three size groups and named the biggest group "small", then he labeled the medium group as "very small," and the smallest group as "very very small." Although most of the students clearly distinguished the different sizes of the groups in the labels, the number of the groups and the descriptive words in the labels tended to be minimal. The one-way ANOVA did not show any significant difference between the groups ($p>0.05$).

Student Group (total N.)	Count of the students for each score					
	1	**2**	**3**	**4**	**Mean**	**SD**
TAVR (N=27)	2 7.4%	12 44.4%	8 29.6%	5 18.5%	2.6	0.9
TVR (N=26)	2 7.7%	7 26.9%	14 53.8%	3 11.5%	2.7	0.8
TAR (N=26)	2 7.7%	9 34.6%	10 38.5%	5 19.2%	2.7	0.9
TR (N=25)	1 4.0%	6 24.0%	14 56.0%	4 16.0%	2.8	0.7
Total (N=70)	7 6.7%	34 32.7%	46 44.2%	17 16.3%	2.7	0.8

Table 5.4. Frequencies of the students for each number of the size groups that the students (in each student group) generated for labeling task 2.

Card Sorting Task 2

In card sorting task 2, the students were asked to revise the cards that they sorted in card sorting task 1, after they interacted with the TAVRs (and TVRs, TARs, and TRs) looking at the total accumulation durations that were shown in the representations. In the following, I present the results from the three sub-tasks of card sorting task 2.

Ordering

In the results, all students in every group correctly ordered the imperceptible objects by size in ordering task 2 (scored 5 points). The combinations of augmenting modalities did not influence the way the students perceived and conceptualized the order of different imperceptible sizes because they could complete this task only by looking at the total accumulation durations once they understood the inverse relationship between the duration and the represented size. The focus group interviews revealed that the repetitive instruction (*"...said it many times,"* student 1-A) and the *"strong emphasis"* (student 1-B) on the inverse relationship between the accumulation durations and the sizes effectively helped the students grasp the concept and correctly order the objects. A student also commented that the quizzes (see Figure 3.11 and 3.12 in Chapter 3) also helped her to *"focus to think about the relationship between the time that takes to put things on the pin… and how small the things are."*

Grouping

In grouping task 2, the students in all groups had increases in the number of size groups that they generated. However, their performances varied; as shown in Table 5.5, while almost 80% of the TAVR group (*N*=21) created four size groups or more, about 58% the TVR group (*N*=15) made four size groups or more. Only 26% of the TAR group (*N*=7) and 20% of the TR group (*N*=5) generated four or more size groups. The results of the statistical analysis were consistent with the hypothesis (TAVR > TVR = TAR > TR) except that the TAR group had a significantly lower result than the TVR group and did not have a significantly different result from the TR group; hence, **TAVR > TVR > TAR = TR.** The result of the a priori test for the posttest data showed that the TAVR group (*M*=4.2, *SD*=0.86) created significantly more size groups than the TVR group (*M*=3.7, *SD*=0.7); *t*=2.01 (*df*, 50.22), *p*=0.045 (one-tailed). The TVR group created significantly more size groups than the TAR group (*M*=332, *SD*=0.55); *t*=2.25 (*df*, 47.89), *p*=0.029 (one- tailed), while the difference between the TAR group and the TR group was not significant (*p*>0.05).

The focus group interviews revealed how such differences resulted. Overall, in the focus group interview, the students in all groups commented that they made the size groups mainly by looking at the biggest units of time in the total accumulation durations and then by trying to interpret the differences between the actual total durations that had

Student Group (total N.)	Count of the student responses							
	1	2	3	4	5	6	Mean	SD
TAVR (N=27)	-	-	6 22.2%	13 48.1%	6 22.2%	2 7.4%	4.2	0.86
TVR (N=26)	-	-	11 42.3%	12 46.2%	3 11.5%	-	3.7	0.68
TAR (N=26)	-	-	19 73.1%	6 23.1%	1 3.8%	-	3.3	0.55
TR (N=25)	-	1 4.0%	19 76.0%	4 16.0%	1 4.0%	-	3.2	0.58
Total (N=104)	-	1 1.0%	55 52.9%	35 33.7%	11 10.6%	2 1.9%	3.6	0.77

Table 5.5. Minimum, maximum, mean, and standard deviation of each group's performance on grouping task 2.

the same biggest units of time. In order to interpret the differences, they recalled their experiences of the waiting, which is the perception of the passage of time while the accumulations were in progress. However, the perceived waiting experiences varied by student group, and in consequence, their interpretations of the units of time differed by group as well. In summary, it seemed that the visual representation independently augmented the perception of the temporal experiences, but the aural representation was able to augment the temporal experience only when the visual representation was present; the visual representation seemed to become even more useful in combination with the aural representation (the TAVR). The visual representation alone (the TVR condition) more effectively augmented the temporal experience than the aural representation alone (the TAR) condition.

The focus group interviews implied that the aural representation amplified the perception of the passage of time in association with the visual representation. The TAVR focus group commented that the audio was *"so annoying," "so boring,"* or *"so long and tiring,"* but it was helpful for them to *"…realize how slow the accumulation* *was going on when I looked at the red line."* The students in the TAR group also commented on how annoying the aural representation was; however, considering that the TAVR group formed significantly more size groups than the TAR group, and the TAR and TR group were not significantly different from each other, it seems that the aural representation effectively augmented the temporal experience only when the visual representation was present.

The reason why the aural representation was useful only when used with the visual representation is that it helps the students better grasp the awareness of the accumulations in progress. Student 1-TAVR-C commented, *"first I thought my computer was down because there was nothing happening on the pinhead, but the sound was still coming out and I could move around the things on the screen. Then I realized that the things [TAVRs] were going on."* It also seemed that the aural representation might have provided a context that the students could use to try to guess the amount of the objects that were placed on the pinhead. Student 1-TAVR-D stated, *"[when grouping the objects by similar sizes] I tried to calculate how many objects were put on the pin trying to remember how fast the beeps [the clicks] were playing."* Moreover, the students seemed to lose concentration in the middle of the learning activity. Many students merely watched the clock and mindlessly pressed the fast forward button to see the clock stop, which is an indicator of the completion of the accumulation. There were students who kept clicking even after the accumulations were completed without noticing it.

The students in the TR group commented on how boring the waiting was as well; however, it seems that their kind of boredom was a different kind of boredom from what the students in the TAVR or TAR group mentioned. The TAVR group students' boredom was rather "the perceivable annoyance over a certain passage of time" that eventually made them interpret the same passage of time differently from the students in other groups, while the TR group students' boredom was the true boredom of perceiving no stimuli from the TRs (except the ticking clocks) on the computer screens. In fact, although I emphasized that the accumulation was happening on the pinhead, many students asked me if their computers were properly working or how they would know when the accumulation was completed. This genuine boredom that was accompanied by uncertainty, not annoyance, in the TR condition seemed to make their attention stray away from the learning activity.

The visual modality, alone (TVR) or in combination with the aural representation (TAVR), seemed to have provided the learners with three waves of "wow, it is smaller than I thought!" experience. With the presence of the visual modality, learners have continuous support for iterative revision of their mental models in three steps. The first experience happens when they have to wait for a certain amount of time, wondering when they will start to see the red line. I observed that when they did not see the red line within the time they expected (which usually is shorter than the actual wait time), students realized that their *"prediction was much shorter"* (student 1-TAVR-A) and thought, *"This is extremely smaller than I thought."* Then they reset their predicted wait time to be longer, which in most cases is still shorter than the real wait time. Hence, through this repetitive waiting-and-re-anticipating process, the students iteratively refined the mental models of the range of imperceptible smallness.

The second wave occurs when they finally see the appearance of the red line. It was observed that most of the students took this event as a pleasant message because they hastily thought that the wait "*would not be too long from this moment*" (student 1-TVR-C) (but it usually remains quite long, especially for a sub-nanoscopic object such as an atom). On this event, the students intentionally looked at the clock and read it. In fact, during the instruction in the classrooms, the students in the TAVR and TVR groups made joyful comments (e.g., *"Oh! Now I can see it!" or "Look! It took 14 hours to see this tiny little red line!"*). These reactions were followed by revisions of the students' previous predictions, which consequently resulted in the modification of their mental models. Additionally, this event brought the students a reinforcing message about the size of the object: *"it took a long time because the object is extremely small"* (student 1-TAVR-D).

The last wave begins from the time when the red line first appears. Most of the students thought that it would only take a couple of minutes more to see the completion once the red line appeared. However, the students actually had to wait for a longer time (about thirty times longer than the wait until the first appearance of the red line). During this time, they either had to press the fast forward button repetitively or had to watch the computer screen for a very long time. At the same time, they watched the accumulation progressing (very slowly), revising their predictions in their minds. Such iterative mental model revisions seemed to become emotional (e.g., *"very surprised," "disappointed," or "too boring"*).

Finally, in addition to the visual and aural representations, many students in all focus groups (about 80%, N=3.5 students on average) actively made use of their experiences with the skip-ahead button when interacting with the temporal representations (TAVR, TVR, TAR, and TR). For example, the students who put only the hydrogen atom in the smallest group in grouping task 2 explained their logic as, *"my* *hand got so tired clicking for this one [pointing at hydrogen atom] ... because it took so long, because it is too small"* (student 1-TAVR-B) or *"I had to press the mouse forever"* (student 1-TAVR-C). Student 1-TVR-A commented, *"I was not sure if I was going to finish this thing [pointing at hydrogen Atom] in time before the bell [that indicates the end of a class] because I worked really hard with this button [pointing at the skip-ahead button] but there was nothing happening on the pin for about five minutes."* Student 1-TAR-A grouped the hydrogen atom and the water molecule together into the smallest group because *"these two [pointing at hydrogen atom and water molecule] hurt my fingers bad."* Even a student in the TR group enthusiastically clicked the skip-ahead buttons although no aural or visual feedback appeared on the computers. A student in the TR group commented, *"I pressed the button very hard, I mean, very fast and quickly for many many times because I had no clue about what was going on"* (student 1-TR-C). These statements from the students imply that the kinesthetic interaction with the temporal representations supported them in generating verbal information about their kinesthetic-temporal experiences as well as kinesthetic mental images of the temporal experiences. Since the effect of the kinesthetic manipulation of the temporal experiences was not investigated in varying combinations with other modalities (aural and visual), in this study it is not possible to determine for which student group it was most useful.

Labeling

In labeling task 2, the number of descriptive words for the smallest group slightly increased in all groups overall. My hypothesis was that (1) the TAVR group would outperform the TVR and the TAR groups and (2) the TVR and the TAR groups would perform similarly to each other but better than the TR group (TAVR > TVR = TAR >

TR) on labeling task 2. However, the result was **TAVR = TVR > TAR = TR**.

See Table 5.6 for the minimum, maximum, mean, and standard deviation of the student scores of each treatment group. The a priori contrast test results showed that the TAVR (M=5.1. SD=0.9) and the TVR (M=4.8, SD=1.3) groups were not significantly different (p=0.328, one-tailed) from each other. The TAVR group performed significantly better than the TAR group (M=4.1, SD=1.0); t=3.75 (df=49.33), p<0.001, one-tailed, and the TR group (M=4.2 SD=1.0); t=3.28 (df=48.01), p=0.002. The TVR group also performed significantly better than the TAR group; t=2.13 (df=47.61), p=0.038, one-tailed. Hence, the results can be summarized as: TAVR = TVR > TAR = TR.

Student group	**Mean of the N. of size groups**	**Labeling task 2 score**			
		Min	**Max**	**Mean**	**SD**
TAVR	4.1	4	8	5.1	0.9
TVR	3.6	3	8	4.8	1.3
TAR	3.3	3	7	4.1	1.0
TR	3.2	3	8	4.2	1.0

Table 5.6. Minimum, maximum, mean, and standard deviation of each student group's performance on labeling task 2.

The students in the TAVR and the TVR groups generated the richest descriptions for the smallest object group they created. Most of the students in both groups tended to use more adjectives, adverbs, and exclamation marks to emphasize the smallness of the smallest group than the students in other groups (TAR and TR). For example, they gave the smallest group labels such as, *"VERY VERY VERY VERY small!", "These guys are the smallest in the world!!!!!!", or "the really extremely teeny-weeny tiny smallest!!!",* while the students in the TAR and the TR groups used fewer adjectives and exclamation marks in the labels, which generated plainer and simpler descriptions (e.g., "the smallest").

In each student group, there were two or three students who did not change the number of size groups but only changed the label for the smallest group. In such cases, their labels became more elaborate. For example, a student in the TVR group labeled the smallest group as "very small" in labeling task 1, then changed it to *"very very very very very small"* in labeling task 2. There also were a few students, regardless of the student groups they belonged to, who made more size groups in grouping task 2 but did not change the labels. In those cases, the labels tended to be very definitive such as, *"need an electron microscope to see this," "number one small,"* or *"smallest of the smalls."*

In the focus group interviews, the students in all groups commented that they were surprised by the sizes of the objects, especially the hydrogen atom and the water molecule, and they tried to reflect such experiences in the label of the smallest group. In order to label the smallest group, they *"thought about how different the times [the accumulation durations] are with each other"* (student 1-TAVR-E) or *"compared the clocks and thought about how small the atom is"* (student 1-TAR-B), or *"tried to make this label look different with others as much as I could because this group is way too smaller than the others"* (student 1-TVR-C). A student in the TR group also stated that he *"tried to make this label [pointing at the smallest group] sound small as much as possible."* As these examples show, the students in all groups seemed to reflect their temporal experiences regarding the smallest sizes in the labels of the smallest size group.

However, considering that the TAVR and the TVR group students used more descriptive words for their labels than the TAR and the TR group students, the visual representation seems to have influenced the students in generating more elaborate labels, although the students' responses in the focus group interviews did not explicitly expose this. I think the differences in labeling task 2 were due to the three waves of mental model revision, which were mainly influenced by the visual representation that I discussed in the previous section for the results of grouping task 2.

Card Sorting Task 1 vs. Card Sorting Task 2

In this section, I explicitly compare the student performances on card sorting tasks 1 and 2 and discuss the difference.

Ordering

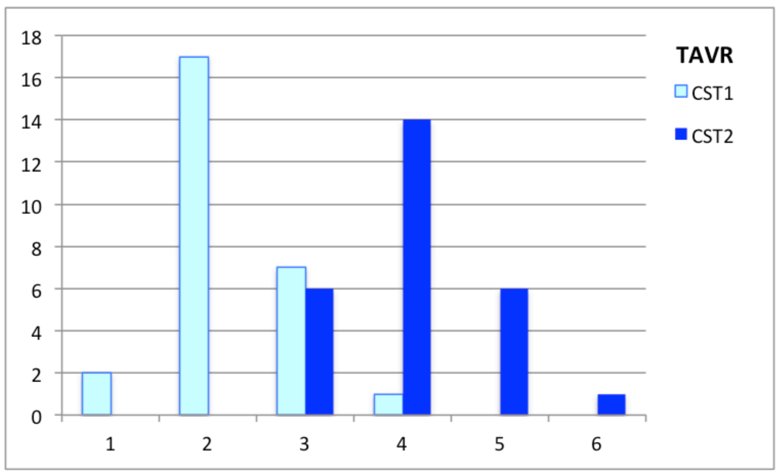
In ordering task 1, all students performed poorly and then properly ordered the objects in ordering task 2, implying that (1) they correctly understood the inverse relationship between the sizes and the accumulation durations, and (2) the different combinations of the augmenting modalities did not differently influence the way the students perceived and interpreted their temporal experiences. During the focus group interviews, the students mentioned that the inverse relationship between the accumulation durations and the sizes was *"pretty clear"* or *"easy to understand"* after learning activity phase 2, where they were introduced to how the TAVR (and TVR, TAR, TR) works.

Grouping

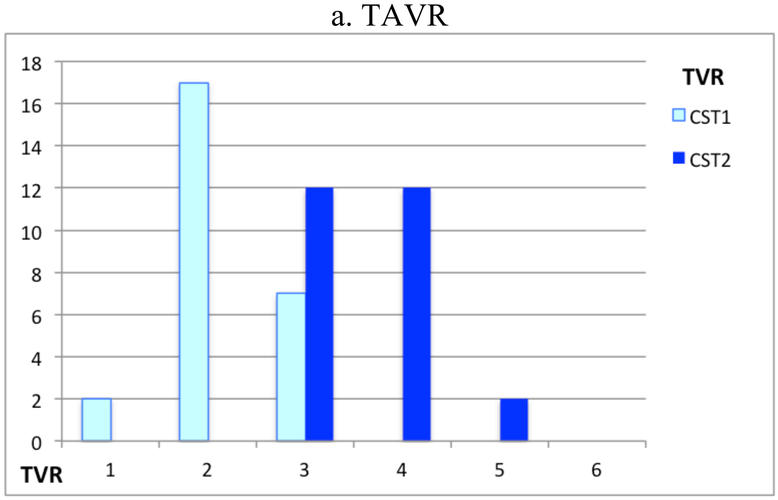
As discussed in the previous section, the student groups did not significantly differ from each other in their results on grouping task 1. However, the result of the a priori contrast analysis of the data from grouping task 2 showed: TAVR > TVR > TAR = TR. These results are represented in the bar graphs in

Figure 5.1. They represent the number of students (Y-axis) for each number of size groups (X-axis) by each treatment group (TAVR, TVR, TAR, and TR) in both grouping task 1 (represented in light blue bars) and grouping task 2 (blue bars). CST1 and

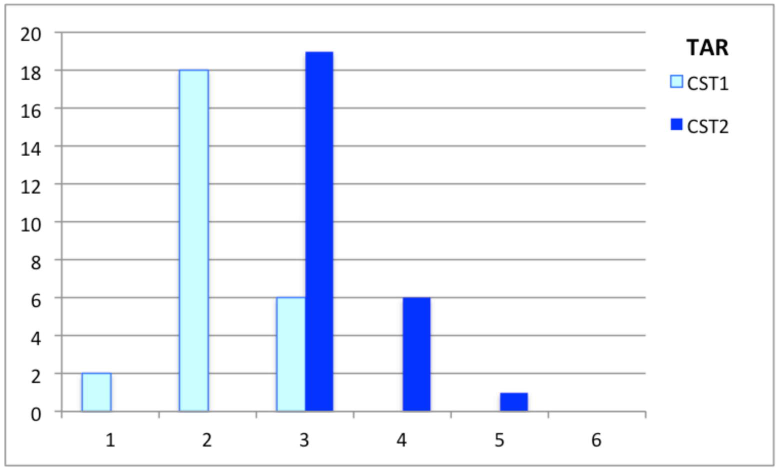
CST2 refer to card sorting task 1 and card sorting task 2. The graphs show that, in general, the students in all four groups exhibited similar patterns of size groups in grouping task 1. Slightly over half of the students in each group created two size groups in grouping task 1, and about 20% of the students in each group made three size groups. In grouping task 2, the students in all groups made more size groups in general. However, the TAVR group had more students who made five or six size groups in grouping task 2 than other groups. About 70% of the students in the TAR and TR groups made three size groups.



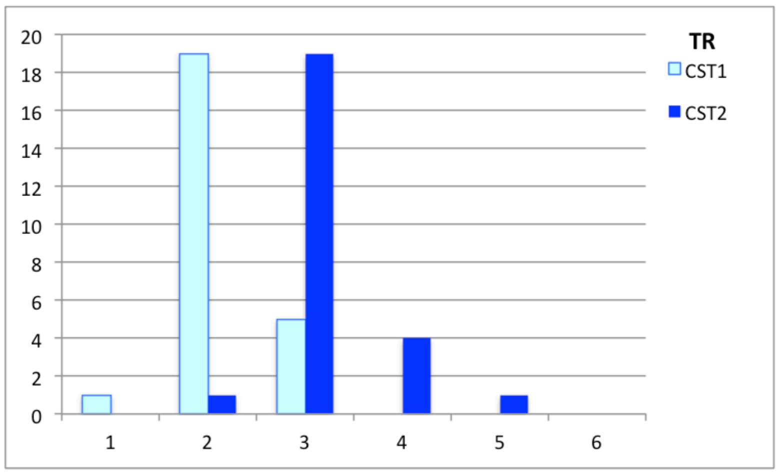
a. TAVR



b. TVR



c. TAR



d. TR

Figure 5.1. The number of the students (Y-axis) for each number of size groups (X-axis) that were classified by similar sizes in the grouping tasks in card sorting tasks 1 (CST1) and 2 (CST2).

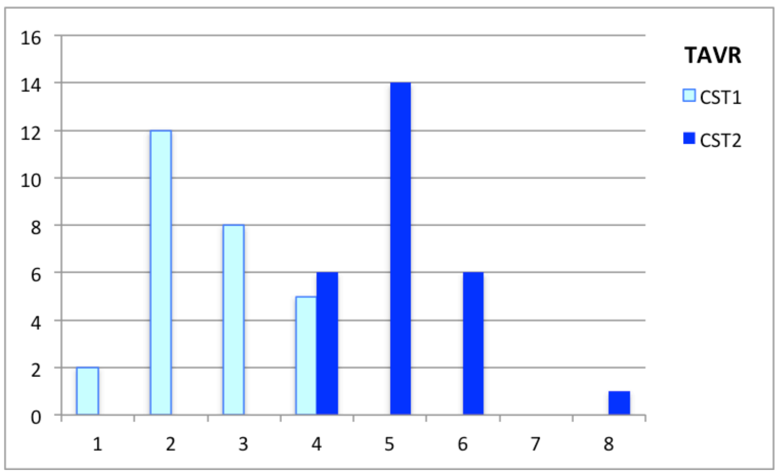
Table 5.7 explains how these differences in grouping task 2 (TAVR > TVR > TAR = TR) occurred. The table presents the count of the students for each difference between the number of the size groups in grouping tasks 1 and 2. The difference, the increase, in other words, for individual students, was calculated by subtracting the number of the size groups they created in grouping task 1 from the number of the size groups in grouping task 2. For example, the table indicates that seventeen students from the TAVR group created two more groups in grouping task 2 than in grouping task 1. About 75% of the students (*N*=20) in the TAVR group increased the number of the size groups by two or more in grouping task 2, while about 45% of the TVR group students had increases of two or more. Furthermore, contrastingly, over 75% of the TAR and TR group students had increases of one or zero in the number of size groups in grouping task 2.

Student Groups	The count of the increase of the number of the student-generated size groups from grouping task 1 to grouping task 2.					
	0	1	2	3	4	Total
TAVR	2 7.4%	5 18.5%	17 63.0%	2 7.4%	1 3.7%	27
TVR	5 19.2%	9 34.6%	8 30.8%	4 15.4%	-	26
TAR	2 7.7%	18 69.2%	6 23.1%	-	-	26
TR	1 4.0%	22 88.0%	2 8.0%	-	-	25
Total	9.6%	51.9%	31.7%	5.8%	1.0%	104

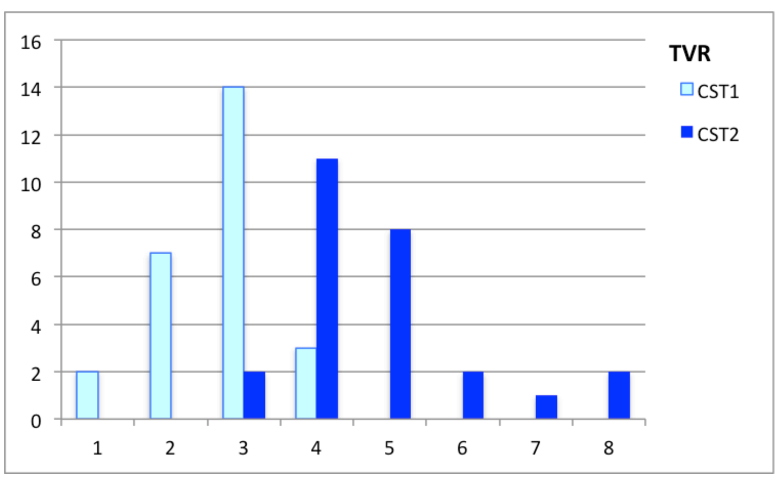
Table 5.7. The count of the increase of the number of the student-generated size groups from grouping task 1 to grouping task 2.

Labeling

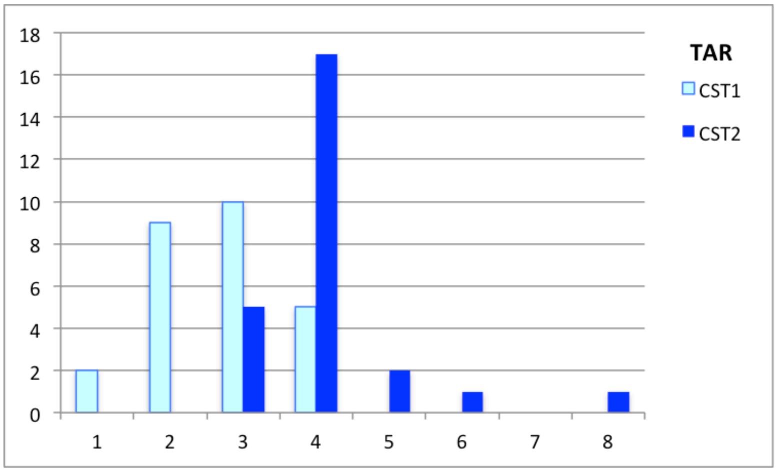
Figure 5.2 shows the number of students (Y-axis) for the scores of the labeling task (X-axis) by each student group (TAVR, TVR, TAR, and TR) in both grouping task 1 (represented in light blue bars; TAVR = TVR = TAR = TR) and grouping task 2 (blue bars; TAVR = TVR > TAR = TR). CST1 and CST2 represent card sorting tasks 1 and 2. The graphs indicate that, in general, the students in all four groups exhibited similar results in labeling task 1. The maximum scores of all groups were 4 (minimum was 1), and about 70% of the students in each group scored 2 or 3 points. In labeling task 2, the scores of all groups were higher than in labeling task 1. Overall, the minimum scores of all groups increased to 3 or 4, and the maximum scores became 8. There were only zero to two students in each group who scored 8 points. However, the distribution of the students differed across the groups; while about 50% (*N*=14) of the students in the TAVR group scored 5 points, only 30% (*N*=8) of the students in the TVR group scored 5 points. Furthermore, 65% (*N*=17) of the TAR group and 50% (*N*=13) of the TR group achieved 4 points.



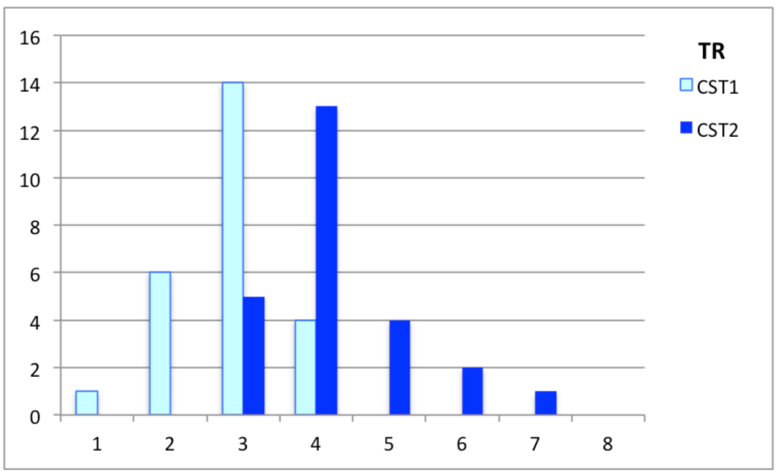
a. TAVR



b. TVR



c. TAR



d. TR

Figure 5.2. The numbers of the students (Y-axis) for the scores on the labeling task (X-axis) that were classified by similar sizes in the grouping tasks in card sorting tasks 1 (CST1) and 2 (CST2).

Table 5.8 presents the count of the students for each possible score difference between labeling task 2 and 1. The increase of the scores of all individual students, in other words, was calculated by subtracting the score of labeling task 1 from that of labeling task 2. About 85% of the students (*N*=23) in the TAVR group had increases in scores of two or more in labeling task 2, while about 55% of the TVR group students (*N*=16) had increases of two or more. Furthermore, contrastingly, only about 50% of the TAR (*N*=11) and TR group students (*N*=11) had increases of two or more in labeling task

2.

Student Groups	The count of the increases in the scores from labeling task 1 to labeling task 2 (= labeling task 2 score - labeling task 1 score)							
	0	1	2	3	4	5	6	Total
TAVR	-	3 11.1%	11 40.7%	11 40.7%	1 3.7%	-	1 3.7%	27
TVR	3 11.5%	5 19.2%	11 42.3%	3 11.5%	2 7.7%	-	2 7.7%	26
TAR	4 15.4%	11 42.3%	7 26.9%	3 11.5%	1 3.8%	-	-	26
TR	6 24.0%	8 32.0%	8 43.0%	1 4.0%	2 8.0%	-	-	25
Total	13 12.5%	27 26.0%	37 35.6%	18 17.3%	6 5.8%	-	3 2.9%	104

Table 5.8. The count of the increase in the scores from labeling task 1 to labeling task 2.

Predictions and Reflections

The students' predictions were similar across the groups. Table 5.9 shows the summary of the spectrums of the student predictions for each imperceptible object. Although significant differences between any student groups have been found from the a priori contrast analysis and the student predictions of all student groups varied drastically, the predictions were inversely relative to the sizes of the objects. In other words, the smaller the object, the longer the predicted accumulation durations.

In the reflections, the student reflections on the seven-point Likert scale did not differ by student group. The result of the a priori contrast analysis using the hypothesis (TAVR > TVR = TAR > TR) that I ran for the student responses regarding each imperceptible object indicated that their responses did not differ by student group ($p>0.05$). As in the ordering task, it seems that the augmenting modalities (visual and aural) did not affect the students in refining their mental models because this task only required them to look and compare the accumulation durations in order to respond on the scale.

Objects	Actual total accumulation time	**Student predictions**	
		min	**max**
Atom	~12 days	10 sec	1 day
Molecule	~4 days	8 sec	1 hour
DNA	~14 hours	2 sec	30 min
Virus	~1 hours	7 sec	10 min
Bacteria	~ 50 sec	5 sec	40 sec
Red blood cell	~ 14 sec	3 sec	30 sec

Table 5.9. The range of the student predictions for each imperceptible object's accumulation time.

Although the students' responses did not differ by student group, in general their reflections tended to be more extreme as the differences between their predictions and the actual accumulation durations became larger. For example, a student who predicted that it would take less than one hour (it actually takes about eleven days and thirteen hours for the hydrogen atom) answered that the size of an atom was "*extremely smaller than I thought*" on the Likert scale. The same student responded that the size of a red blood cell was similar to what he thought because his prediction was 10 seconds (it takes about sixteen seconds for a red blood cell). In this manner, many of the students in all groups marked "extremely smaller than I thought" for the hydrogen atom. Hence, the students' reflections on the Likert scale tended to be more extreme toward the "smaller than I thought" side, as the size of the object became smaller. See Table 5.10 for the number of the students for each of the seven points on the Likert scale, counted for each imperceptible object.

Student group	Count of the student responses **(hydrogen atom)**						
	Extremely smaller than I thought	Much smaller than I thought	Somewhat smaller than I thought	Similar with what I thought	Somewhat bigger than I thought	Much bigger than I thought	Extremely bigger than I thought
TAVR (N=27)	17 63.0%	4 7.7%	6 30.8%	-	-	-	-
TVR (N=26)	19 73.1%	3 11.5%	3 11.5%	1 3.8%	-	-	-
TAR (N=26)	16 61.5%	2 7.7%	8 30.8%	-	-	-	-
TR (N=25)	18 72.0%	2 8.0%	5 20.0%	-	-	-	-
Total (N=104)	70 67.3%	11 10.6%	22 21.2%	1 1.0%	-	-	-

a. Hydrogen atom

Student group	Count of the student responses **(water molecule)**						
	Extremely smaller than I thought	Much smaller than I thought	Somewhat smaller than I thought	Similar with what I thought	Somewhat bigger than I thought	Much bigger than I thought	Extremely bigger than I thought
TAVR (N=27)	16 59.3%	6 22.2%	5 18.5%	-	-	-	-
TVR (N=26)	11 42.3%	10 38.5%	5 19.2%	-	-	-	-
TAR (N=26)	16 61.5%	5 19.2%	5 19.2%	-	-	-	-
TR (N=25)	15 60.0%	5 20.0%	5 20.0%	-	-	-	-
Total (N=104)	58 55.8%	26 25.0%	20 19.2%	-	-	-	-

b. Water molecule

Student group	Count of the student responses **(DNA helix)**						
	Extremely smaller than I thought	Much smaller than I thought	Somewhat smaller than I thought	Similar with what I thought	Somewhat bigger than I thought	Much bigger than I thought	Extremely bigger than I thought
TAVR (N=27)	10 37.0%	13 48.1%	4 14.8%	-	-	-	-
TVR (N=26)	9 34.6%	10 38.5%	7 26.9%	-	-	-	-
TAR (N=26)	8 30.8%	14 53.8%	4 15.4%	-	-	-	-
TR (N=25)	7 28.0%	11 44.0%	7 28.0%	-	-	-	-
Total (N=104)	34 32.7%	48 46.2%	22 21.2%	-	-	-	-

c. DNA helix

Student group	Count of the student responses **(rhinovirus)**						
	Extremely smaller than I thought	Much smaller than I thought	Somewhat smaller than I thought	Similar with what I thought	Somewhat bigger than I thought	Much bigger than I thought	Extremely bigger than I thought
TAVR (N=27)	13 48.1%	13 48.1%	1 3.7!	-	-	-	-
TVR (N=26)	12 46.2%	11 42.3%	3 11.5%	-	-	-	-
TAR (N=26)	12 46.2%	10 38.5%	4 15.4%	-	-	-	-
TR (N=25)	12 48.0%	11 44.0%	2 8.0%	-	-	-	-
Total (N=104)	49 47.1%	45 43.3%	10 9.6%	-	-	-	-

d. Rhinovirus

Student group	Count of the student responses **(e coli bacterium)**						
	Extremely smaller than I thought	Much smaller than I thought	Somewhat smaller than I thought	Similar with what I thought	Somewhat bigger than I thought	Much bigger than I thought	Extremely bigger than I thought
TAVR (N=27)	6 22.2%	9 33.3%	10 37.0%	2 7.4%	-	-	-
TVR (N=26)	2 7.7%	13 50.0%	11 42.3%	-	-	-	-
TAR (N=26)	5 19.2%	8 30.8%	12 46.2%	1 3.8%	-	-	-
TR (N=25)	2 8.0%	9 36.0%	13 52.0%	1 4.0%	-	-	-
Total (N=104)	15 14.4%	39 37.5%	46 44.2%	4 3.8%	-	-	-

e. e coli bacterium

Student group	Count of the student responses **(red blood cell)**						
	Extremely smaller than I thought	Much smaller than I thought	Somewhat smaller than I thought	Similar with what I thought	Somewhat bigger than I thought	Much bigger than I thought	Extremely bigger than I thought
TAVR (N=27)	5 18.5%	1 3.7%	15 55.6%	6 22.2%	-	-	-
TVR (N=26)	9 34.6%	1 3.8%	13 50.0%	3 11.5%	-	-	-
TAR (N=26)	8 30.8%	2 7.7%	13 50.0%	3 11.5%	-	-	-
TR (N=25)	4 16.0%	1 4.0%	16 64.0%	4 16.0%	-	-	-
Total (N=104)	26 25.0%	5 4.8%	57 54.8%	16 15.4%	-	-	-

f. Red blood cell

Table 5.10. The number of the students for each reflection point on the Likert scale, counted for each imperceptible object.

To compare the common pattern of the student responses across the student groups in depth, I converted the student responses to numeric data. The response option “similar with what I thought” was converted to number 1, “somewhat smaller than I thought” to 2, “much smaller than I thought” to 3, and “extremely smaller than I thought” to 4. See Table 5.11 for the means of the responses from each student group for each imperceptible object. The means are also visually compared in Figure 5.3, as stacked line graphs. As both Figure 5.3 and Table 5.11 display, the reflections of the students, in general, regardless of their groups, tended to be greater for the sub-nano or nanoscopic objects (e.g., hydrogen atom and water molecule) than for the microscopic objects (i.e., e coli bacterium and red blood cell).

Student group	Means of the student responses, coded into numbers, for each imperceptible object					
	Hydrogen atom	Water molecule	DNA helix	Rhinovirus	e Coli bacterium	Red blood cell
TAVR	3.6	3.4	3.2	3.5	2.4	2.7
TVR	3.7	3.2	3.1	3.3	2.6	2.6
TAR	3.6	3.4	3.1	3.3	2.6	2.6
TR	3.7	3.3	3	3.3	2.2	2.5

Table 5.11. Means of the student responses, coded into numbers, for each imperceptible object.
The codes are: “similar with what I thought” = 1; “somewhat smaller than I thought” = 2; “much smaller than I thought” = 3; “extremely smaller than I thought” = 4.

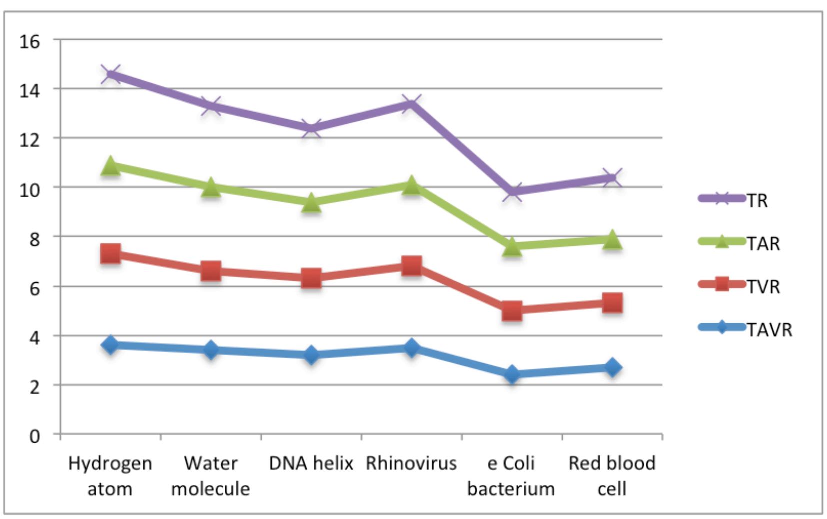


Figure 5.3. Stacked line graph of the average student responses of each student group. Please note that the lines are stacked.

Considering that the student groups that experienced the temporal representations with the visual component (TAVR and TVR) outperformed other two groups (TAR and TR) in the grouping and labeling tasks, I believe this result was due to the limits in the testing method that provided just seven points on the Likert scale, which only allowed three degrees of "smaller than I thought" – "extremely," "much," and "somewhat." In fact, many students in the extended condition group explicitly asked for more options that would allow them to mark on a smaller scale than "extremely smaller than I thought."

Card sorting task 3

In card sorting task 3, which was conducted about six weeks after card sorting task 2, the numbers of the participants from each group were reduced by two or three students, either because they were absent or because they had to participate in extracurricular activities (e.g., marching band practice). I present the total number of the participants of each student group in the tables of the results in the following.

Ordering

About 90% of the students across all groups remembered that the atom was the smallest object and took about twelve days to accumulate; however, most of them could not get the order of the other five imperceptible objects right. See Table 5.12 for the count of the students for each possible ordering task score. I think this result is natural because there was no follow-up instruction that reinforced their memorization of the order of the objects.

However, although the total ordering task scores of the student groups decreased in the long-term effect test, significant group differences were found (**TAVR = TVR >**

TAR = TR). This result is the same as the finding from labeling task 2. The a priori contrast analysis revealed that the TAVR group (*M*=2.6 *SD*=0.7) performed similarly to the TVR group (*M*=2.4, *SD*=0.8); $p>0.05$, one-tailed, and the TAR group (*M*=2.0, *SD*=0.7) and the TR group (*M*=2.0, *SD*=0.7) achieved similar scores ($p>0.05$), while the scores of the TAR group were significantly lower than the TVR group; $t=2.22$ ($df=44.97$), $p=0.032$, one-tailed. The correlation between labeling task 2 and ordering task 2 revealed that these two variables were strongly correlated, $r(92)=0.83$, $p<0.001$. Hence, I did not proceed to conduct a multivariate ANOVA (MANOVA) for one cannot use MANOVA when the dependent variables are highly correlated; it will produce the risk of multicollinearity.

Student group	1	2	3	4	Mean	SD
TAVR (N=24)	-	13 54.2%	8 33.3%	3 12.5%	2.6	0.7
TVR (N=24)	1 4.2%	14 58.3%	7 29.2%	2 8.3%	2.4	0.8
TAR (N=23)	5 21.7%	15 65.2%	2 8.7%	1 4.3%	2.0	0.7
TR (N=23)	5 21.7%	13 56.5%	4 17.4%	1 4.3%	2.0	0.7
Total (N=94)	11 11.7%	55 58.5%	21 22.3%	7 7.4%	2.3	0.7

Table 5.12. The count of the students for each possible ordering task score.

Grouping

Most of the students in all groups remembered how many size groups they had created six weeks ago. Specifically, the students who created five or six groups in grouping task 2 accurately remembered the number of the size groups they had created previously. Hence, the statistical analysis of the he a priori contrast analysis of grouping task 3 showed a similar relationship between the student groups as the findings from grouping task 2: **TAVR > TVR > TAR = TR**. See Table 5.13 for the count of the students for each possible number of size groups. The TAVR group (*M*=4.3, SD=0.86) created significantly more size groups than the TVR group (*M*=3.6, *SD*=0.71); *t*=2.93 (*df*, 44.45), *p*=0.005, one-tailed, and a significant difference was found between the TVR group and the TAR group (*M*=3.2, *SD*=0.52); *t*=2.25 (*df*, 42.06), *p*=0.03, one-tailed, as well. The results of the TAR group and the TR group (*M*=3.4, *SD*=0.78) did not differ (*p*>0.05). The changes in the numbers of the size groups tended to be between -1 and 1. I re-coded the number of the changes in positive numbers (-1 to 1, 0 to 2, 1 to 3) and checked for a correlation with the results of grouping task 2, and the results indicated a weak negative linear relationship; *r*(92)= -0.214, *p*=0.039.

Student Group (total N.)	**Count of the student responses**							
	1	**2**	**3**	**4**	**5**	**6**	**Mean**	**SD**
TAVR (N=24)	-	-	4 16.7%	11 45.8%	7 29.2%	2 8.3%	4.3	0.86
TVR (N=24)	-	-	12 50.0%	9 37.5%	3 12.5%	-	3.6	0.71
TAR (N=23)	-	1 4.3%	16 69.6%	6 26.1%	-	-	3.2	0.52
TR (N=23)	-	1 4.3%	16 69.6%	3 13.0%	3 13.0%	-	3.4	0.78
Total (N=94)	-	2 2.1%	48 51.1%	29 30.9%	13 13.8%	2 2.1%	3.6	0.82

Table 5.13. Minimum, maximum, mean, and standard deviation of each group's performance on grouping task 3.

Labeling

In labeling task 3, the students showed the same pattern of the mean scores as they did in labeling task 2: **TAVR = TVR > TAR = TR.** The a priori contrast analysis revealed that the TAVR group (*M*=6.0 *SD*=1.00) performed similarly with the TVR group (M=5.9, SD=1.60); p>0.05, one-tailed, and the TAR group (M=5.0, SD=1.49) and the TR group (M=5.0, SD=1.24) achieved similar scores (p>0.05), while the scores of the TAR group were significantly lower than the TVR group; t=2.06(df=45.00), p=0.045, one-tailed.

Student Group (total N.)	**Count of the student responses**									
	3	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Mean**	**SD**
TAVR (N=24)	-	-	8 33.3%	9 37.5%	5 20.8%	2 8.3%	-	-	6.0	1.0
TVR (N=24)	1 4.2%	2 8.3%	7 29.2%	10 41.7%	-	2 8.3%	1 4.2%	1 4.2%	5.9	1.6
TAR (N=23)	3 13.0%	7 30.4%	7 30.4%	2 8.7%	3 13.0%	-	1 4.3%	-	5.0	1.5
TR (N=23)	3 13.0%	4 17.4%	9 39.1%	5 21.7%	1 4.3%	1 4.3%	-	-	5.0	1.2
Total (N=94)	7 7.4%	13 13.8%	31 33.0%	26 27.7%	9 9.6%	5 5.3%	2 2.1%	1 1.1%	5.5	1.4

Table 5.14. Minimum, maximum, mean, and standard deviation of each group's performance on labeling task 3.

Interestingly, the mean scores of all groups were higher than those for labeling task 2. See Table 5.15 for the comparison of the mean scores of each student group in labeling tasks 2 and 3.

Student Group (total N.)	Labeling task 2		Labeling task 3	
	Mean	SD	Mean	SD
TAVR (N=24)	5.2	0.9	6.0	1.0
TVR (N=24)	4.9	1.2	5.9	1.6
TAR (N=23)	4.2	1.1	5.0	1.5
TR (N=23)	4.3	1.1	5.0	1.2
Total (N=94)	4.7	1.1	5.5	1.5

Table 5.15. The mean and standard deviation of each student group in labeling tasks 2 and 3.

Summary

From the results of research question 1, the following findings were resulted:

1. *The TAVR was most effective:* The combination of the visual representation and the aural representation was most effective for augmenting the temporal experiences.
2. *The visual representation was effective with or without the aural representation:* The visual representation, even when used alone, was effective for augmenting the temporal experiences by facilitating a better awareness of the accumulation progresssions. The visual cue of accumulation seemed to reinforce the learners' experience of the sequential accumulation of an imperceptible object at three different stages: (1) while waiting from the beginning of the accumulation until the first appearance of the visual representation, (2) at the moment of the first appearance of the visual representation, (3) while waiting for the completion of the accumulation after the first appearance of the visual representation.
3. *The aural representation was useful only when used in combination with the visual representation:* The aural component did not play a critical role in supporting learners to better perceive and conceptualize the temporal experiences of imperceptible sizes when the visual component was absent. I believe this is because the students in the TAR group could not observe the moment when the accumulation became macroscopic and the red line finally started to appear.
4. The students who interacted with the TAVRs tended to amplify their learning experiences in their memory. In particular, the students remembered that the hydrogen atom was the smallest object.

Research Question 2

How do different intervals of temporal experiences influence learners' conceptualization of the range of imperceptible sizes?

Before discussing the results, I would like to remind the readers of the fact that the participants for this research question were different students from the participants for research question 1. They were from a different middle school, and hence, a different science teacher.

Survey

The two questions that were asked of the students in the survey were (1) if they have heard of the names of each object, and (2) if they thought the sizes of the objects were smaller than the size of the head of a pin (which was shown to the students). See Table 5.16 and Table 5.17 for the summary of each student group's responses to the two survey questions. The number of student answers to the second question includes responses only from the students who answered, "yes," to the first question. Hence, the total numbers of answers to the second question are equal to the number of the students who answered "yes" to the first question.

Question		**Student responses (extended group)** N=29					
		Atom	Molecule	DNA	Virus	Bacterium (Bacteria)	Red blood cell
Heard?	Yes	28	28	28	28	29	28
	No	1	1	1	1	-	1
Smaller than pinhead?	Yes	26	27	25	24	28	22
	No	2 7%	1 4%	3 11%	4 14%	1 3%	6 21%

Table 5.16. The student responses to two survey questions (the extended group).

Question		Student responses (compressed group) N=28					
		Atom	Molecule	DNA	Virus	Bacterium (Bacteria)	Red blood cell
Heard?	Yes	28	28	28	28	26	28
	No	-	-	-	-	2	-
Smaller than pinhead?	Yes	28	27	20	26	24	21
	No	-	1 4%	7 26%	2 7%	1 4%	7 25%

Table 5.17. The student responses to two survey questions (the compressed group).

As the Tables show, almost all students in both groups had heard of the names of the objects; however, many students thought that the sizes of some objects were bigger than the size of the head of a pin. Specifically, over 20% of the students in both groups (extended group: 21%, *N*= 6; compressed group: 25%, *N*=7) thought that the diameter of a DNA helix was bigger than the diameter of a pinhead. Moreover, 26% of the compressed group students also thought the diameter of a DNA helix is bigger than the diameter of a pinhead. Except for DNA and red blood cells, only one or two students in the compressed group thought that the other objects were bigger than the diameter of pinhead. However, in the extended group, three students (11%) thought that the diameter of a DNA helix was bigger than the pinhead, and four students (14%) answered that the virus was bigger than the pinhead.

I conducted a Chi-Square test for each object to see if there existed statistically significant differences between the student groups, specifically for the second survey question, because significantly poorer preexisting knowledge would influence the interpretation of the results from other student tasks. The results showed that each group's responses to each object were not significantly different from those of the other groups ($p > 0.05$).

Card Sorting Task 1

In card sorting task 1, the students in both groups performed poorly and did not show any significant group difference (*p>0.05*). All students overestimated the sizes of the imperceptible objects and underestimated the range of imperceptible sizes (i.e., number of size groups less than 2, minimal use of descriptive words in labels).

Ordering

The results showed that none of the students in any group had the order of the objects correct. See Table 5.18 for the summary of the count of the students for each ordering score and the descriptions. No difference between the student groups was found using One-Way ANOVA (*p>0.05*). Most of the students in both groups achieved scores of two in ordering task 1; about 72.4% of the extended group (*N*=21) and 68.4% of the compressed group (*N*=18) scored two points.

Groups	**Count of the student scores**					
	0	**1**	**2**	**3**	**Mean**	**SD**
Extended (N=29)	-	3 10.3%	21 72.4%	5 17.2%	2.1	0.53
Compressed (N=28)	-	5 17.9%	18 64.3%	5 17.9%	2.0	0.60
Total (N=104)	-	8 14.0%	39 68.4%	10 17.5%	2.04	0.57

Table 5.18. The number of the students for each score (0-3) of the ordering task 1.

Grouping

In grouping task 1, most of the students overestimated the sizes of the imperceptible objects and underestimated the range of imperceptible sizes. Hence, about 95% of the students of each group generated two or three size groups, and only one student in each group made four size groups. See Table 5.19 for the minimum, maximum, mean, and standard deviation of the results, and the count of the student responses for each possible number of size groups that are summarized according to student group (extended vs. compressed). The one-way ANOVA did not show any significant difference between the groups ($p>0.05$).

Student Group (total N.)	Count of the student responses							
	1	**2**	**3**	**4**	**5**	**6**	**Mean**	**SD**
Extended (N=29)	-	15 51.7%	13 44.8%	1 3.4%	-	-	2.5	0.57
Compressed (N=28)	-	14 50.0%	13 46.4%	1 3.6%	-	-	2.5	0.58
Total (N=57)	-	29 50.9%	26 45.6%	2 3.5%	-	-	-	-

Table 5.19. Frequencies of the students for each number of the size groups that the students (in each treatment group) generated for grouping task 1.

Labeling

Table 5.20 shows the distribution of students (for each group) across the scores of labeling task 1. Considering that the total scores of each student's labeling task were generated by the sum of the scores of (1) the student's logic of expressing the difference between size groups (two points maximum) and (2) the number of descriptive expressions the student used (see Table 4.19 in Chapter 4), the results imply that the students tended not to use many descriptive words in the labels. Although most of the students clearly distinguished the different sizes of the groups in the labels (about 93% of the students of each student group scored two or higher), the descriptive words in the labels for the smallest size group tended to be minimal; the highest score in both groups is only four points. The One-Way ANOVA did not show any significant difference between the groups ($p>0.05$).

Student Group (total N.)	Distribution of the students across the scores					
	1	2	3	4	Mean	SD
Extended (N=29)	2 6.9%	10 34.5%	11 37.9%	6 20.7%	2.7	0.88
Compressed (N=28)	2 7.1%	12 42.9%	9 32.1%	5 17.9%	2.6	0.87
Total (N=57)	4 7.0%	22 38.6%	20 35.1%	11 19.3%	2.7	0.87

Table 5.20. Distribution of the students across the scores of labeling task 1.

Card Sorting Task 2

Ordering

In ordering task 2, all students in both groups correctly ordered (scored 5 points) the objects by size. The accumulation rapidity did not make a difference in the ways the students perceived and conceptualized the order of different imperceptible sizes because, as in research question 1, they could complete this task by looking only at the total accumulation durations once they understood the inverse relationship between the accumulation duration and the represented size. The focus group students in both groups commented that it was *"easy"* (student 2-Extended-A) or *"very clear"* (student 2-Compressed-B) to order the objects from the smallest to the biggest.

Grouping

In grouping task 2, as I hypothesized, the students exhibited different responses according to the TAVR types (extended vs. compressed) they interacted with; the extended group created more size groups than the compressed group. As Table 5.21 shows, the extended group created significantly more size groups ($M=4.5$, $SD=0.82$);

($F(1, 55) = 6.2$, p=0.016) than the compressed group (*M=3.9, SD=0.85*). This implies that the difference between the accumulation durations of the imperceptible objects seemed relatively bigger to the students of the extended group than to those of the compressed group.

I noticed that about 86% of the students in the extended group (*N*=25) classified the hydrogen atom as the sole member of the smallest group, while only 43% of the students in the compressed group (*N*=12) did so. About 65% of the students in both groups (the extended: *N*=19, the compressed: *N*=18) grouped the e coli bacterium and red blood cell together. Most of the variance in the numbers of the size groups was due to the way the students grouped the rest of the imperceptible objects: water molecule, DNA helix, and rhinovirus. Eleven students in the extended group (38%) created five size groups by classifying these objects as the single members of four individual groups, plus the biggest group that was made up of e coli bacterium and red blood cell. In contrast, only three students in the compressed group (11%) generated five size groups in this way.

Student Group (total N.)	**Count of the student responses**							
	1	**2**	**3**	**4**	**5**	**6**	**Mean**	**SD**
Extended (N=29)	-	-	3 10.3%	12 41.4%	11 37.9%	3 10.3%	4.5	0.82
Compressed (N=28)	-	-	9 32.1%	14 50.0%	3 10.7%	2 7.1%	3.9	0.85
Total (N=57)	-	-	12 21.1%	26 45.6%	14 24.6%	5 8.8%	4.2	0.88

Table 5.21. Minimum, maximum, mean, and standard deviation of each group's performances on grouping task 2.

The focus group interviews revealed that the students in both groups actively read the total accumulation durations and attempted to interpret them by reflecting on their daily temporal experiences. It seemed that the students in the extended group tended to interpret the accumulation durations within the context of their daily lives. In particular, it seemed that the accumulation duration of the hydrogen atom was quite surprising to the students. For example, when I asked about the logic behind the way they classified the size groups, a student in the extended group (student 2-Extended-A), who put only the hydrogen atom in the smallest group, commented, *"think about how many things you can do for one day. There are just so many. ... and for 115 days, oh, man!"* Student 2-Extended-B, who classified only the hydrogen atom into the smallest group as well, said, *"one hour is long enough to me, and 12 hours is too long. 5 days is also too long. But 115 days? This is unacceptable."* Student 2-Extended-C converted 115 days into *"almost 4 months,"* which was a long enough time for him to *"grow 2 inches,"* and placed the hydrogen atom alone in the smallest group.

In contrast, the students in the compressed group did not comment much on what kinds of meanings the eleven days (the accumulation duration of the hydrogen atom in the compressed condition) had to them in their explanations of the logic behind the way they formed the size groups. The focus group students commented that they focused on the biggest units of time in the accumulation durations. For example, student 2-Compressed-A, who generated three size groups commented, *"these [pointing at the red blood cell and e coli bacterium] are seconds, these [pointing at the Rhinovirus and DNA helix] are some hours, and these [pointing at the hydrogen atom and the water molecule] are many days."* Student 2-Compressed-C, who created four size groups said, *"I grouped e coli bacterium and red blood cell together because they are shorter than one minute, and I put DNA and rhinovirus together because they are shorter than one day."*

These student responses imply that the extended interval of the sequential placement, which results in relatively longer accumulation durations than the compressed interval, made the relative difference between the accumulation durations look more drastic to learners. Consequently, the extended group created more size groups than the compressed group. Although the numbers of the accumulated objects were the same, it was the total accumulation durations of the sequential placement that affected the modification of the students' mental models of the range of imperceptible sizes. This result may become different (e.g., no significant difference between groups) if the total number of the accumulated object were provided in the TAVRs. If the TAVRs showed the number of the accumulated objects to the students, the two student groups might have generated similar numbers of the size groups.

Labeling

In general, many students in both groups used more adjectives, adverbs, and exclamation marks to emphasize the smallness of the smallest group than they did in labeling task 2. However, as I hypothesized, the students in the extended group generated richer descriptions for the smallest size group than the students in the compressed group (see Table 5.22) did for their smallest size group; the students in the extended group used more adjectives and adverbs to name the smallest size group ($M=6.3, SD=1.4$); ($F(1, 55) = 14.0$, $p < 0.001$) than the compressed group ($M=4.9, SD=1.3$) did. A moderate positive correlation between labeling task 2 and grouping task 2 was found; $r(55)=0.32$, $p=0.016$.

As was also observed in the results of research question 1, some students used a number of exclamation marks or capital letters to emphasize the smallness of the size of the smallest group. For example, they gave the smallest group labels such as, *"This is the* *smallest of the smallest x 1000000000!!!!!!"* or *"the really surprisingly teeny-weeny smallest!!!"* Although some students in the compressed group used such decorations as well, they used a smaller number of those in the labels than the extended group.

Student Group (total N.)	Count of the students for each score										
	3	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Mean**	**SD**
Extended (N=29)	-	-	11 37.9%	7 24.1%	7 24.1%	2 6.9%	1 3.4%	-	1 3.4%	6.3	1.4
Compressed (N=28)	2 7.1%	10 35.7%	9 32.1%	4 14.3%	1 3.6%	2 7.1%	-	-	-	4.9	1.3
Total (N=57)	2 3.5%	10 17.5%	21 35.1%	11 19.3%	8 14.0%	4 7.0%	1 1.8%	-	1 1.8%	5.6	1.5

Table 5.22. Frequencies of the students for each number of the size groups that the students (in each treatment group) generated in labeling task 2.

In the focus group interviews, the students in both groups commented that they were surprised by the sizes of the objects, especially by the hydrogen atom, and they tried to reflect such experiences in the label for the smallest size group, which included the hydrogen atom. The explanations of the students regarding the way they named the smallest group were similar for both student groups, although the scores of the extended group were significantly higher than the scores of the compressed group. The focus group students from both student groups similarly commented that the size of a hydrogen atom was *"surprisingly small"* (student 2-Extended-C) or *"very very very very small"* (student 2-Compressed-B) but the students in the extended group scored higher. This implies that both student groups thought that the size of the hydrogen atom was extremely small in their answers to my interview questions, but the perceived and interpreted "smallness" of a hydrogen atom by the extended group students was much smaller than that of the compressed group students.

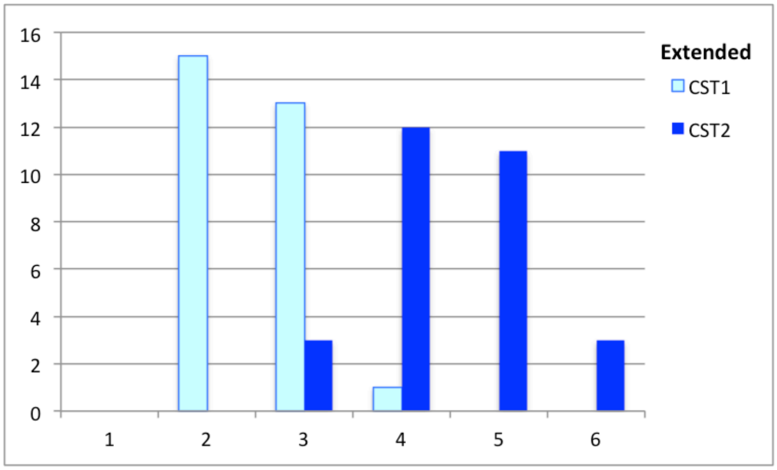
Card Sorting Task 1 vs. Card Sorting Task 2

Ordering

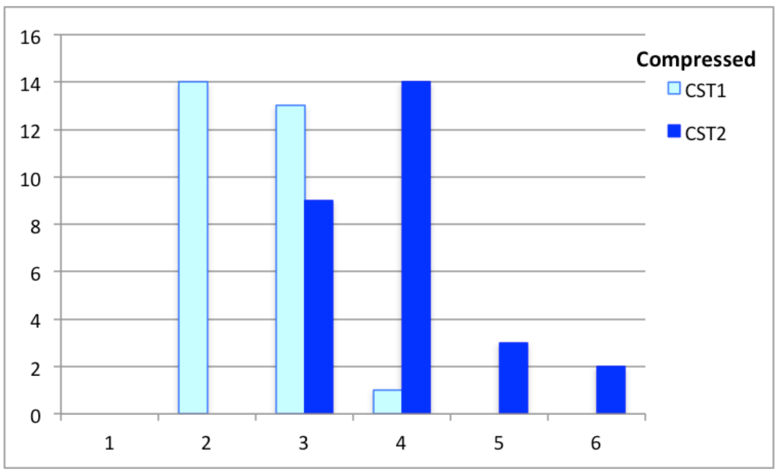
In ordering task 1, all students in both groups performed poorly and then properly ordered the objects in ordering task 2, implying that they correctly understood the inverse relationship between the sizes and the accumulation durations. The One-Way ANOVA did not return significant differences between the two groups, meaning that the length of the interval unit of the temporal experience did have a different influence on the way the students perceived and interpreted their temporal experiences.

Grouping

The bar graph in Figure 5.4 presents the number of the students (Y-axis) for the number of the student-generated size groups (X-axis) by each student group (extended vs. compressed) in both grouping task 1 (represented in light blue bars) and grouping task 2 (blue bars). CST1 and CST2 represent card sorting tasks 1 and 2. The graphs show that the students in both groups exhibited similar results in grouping task 1; the minimum and maximum number of the student-generated size groups are the same between the two groups (min=2; max=4), and the distributions of the students for each number of the size groups are also alike. However, in grouping task 2, the extended group had more students who made five or six size groups in grouping task 2 than the compressed group. About 48% of the students in the extended group (N=14) made five or six size groups, while only 18% of the students in the compressed group (N=5) made five or more size groups.



a. Extended group



b. Compressed group

Figure 5.4. The number of the students (Y-axis) for each number of size groups (X-axis) that were classified by similar sizes in the grouping tasks in card sorting tasks 1 (CST1) and 2 (CST2).

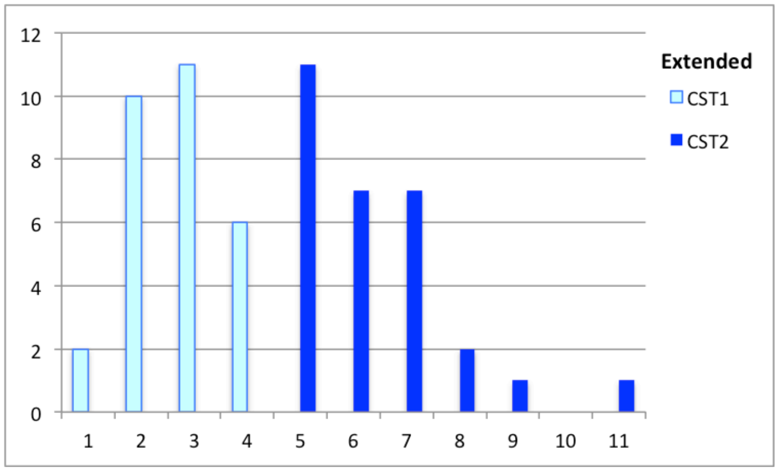
Table 5.23 provides an explanation for how these increases in grouping task 2 resulted (the extended group > the compressed group). The table shows the count of the students for each difference between the number of the size groups in grouping tasks 2 and 1. For example, the table indicates that twelve students of the extended group made two more size groups in grouping task 2. As Table 5.23 shows, almost 70% of the students (N=20) of the extended group had increases in the number of the size groups of two or more, while only about 43% of the TVR group students (N=12) had increases of two or more.

Student Groups	The count of the increase of the number of the student-generated size groups from the grouping task 1 to 2.				
	0	1	2	3	4
Extended (N=29)	-	9 31.0%	12 41.4%	8 27.6%	-
Compressed (N=28)	5 17.9%	11 39.3%	8 28.6%	4 14.3%	-
Total (N=57)	5 8.8%	20 35.1%	20 35.1%	12 21.1%	-

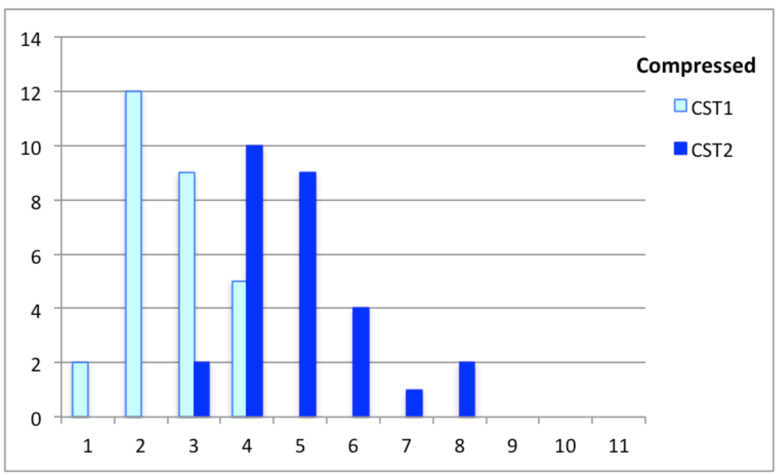
Table 5.23. The count of the increase of the number of the student-generated size groups from grouping task 1 to 2.

Labeling

The graphs in Figure 5.5 show the number of the students (Y-axis) for the scores of the labeling task (X-axis) by each student group (extended and compressed) in both grouping task 1 (represented in light blue bars) and grouping task 2 (dark blue bars). CST1 and CST2 represent card sorting tasks 1 and 2. The graphs indicate that the students in both groups exhibited similar results in labeling task 1. The maximum scores of all groups were 4 (minimum was 1), and over 70% of the students in each group (the extended group: 72%, N=21; the compressed group: 75%, N=21) scored 2 or 3 points. In labeling task 2, the scores of both groups were higher than in labeling task 1. Overall the minimum and the maximum scores of both groups increased; however, the distribution of the students differed across the groups; the minimum score of the extended group changed from 1 point to 5 points, and that of the compressed group changed from 1 point to 3 points. The maximum score of the extended group changed from 4 points to 11 points, and that of the compressed group changed from 4 points to 8 points. While about 62% (N=18) of the students in the extended group scored 6 points or higher, only 25% (N=7) of the students in the compressed group scored 6 points or higher.



a. Extended



b. Compressed

Figure 5.5. The number of the students (Y-axis) for each possible labeling task score in labeling tasks 1 (CST1) and 2 (CST2).

Table 5.24 presents the count of the students for each of the score differences between labeling tasks 1 and 2 (calculated by subtracting the score of labeling task 1 from that of labeling task 2). While about 38% of the students (N=11) in the extended group had increases in the scores of 4 points or more in labeling task 2, only about 14% of the compressed group (*N*=4) had increases of 4 points or more.

Student Groups	The difference between the scores of labeling tasks 1 and 2 (= labeling task 2 score - labeling task 1 score)							
	0	1	2	3	4	5	6	7
Extended (N=29)	-	2 6.9%	3 10.3%	13 44.8%	4 13.8%	3 10.3%	3 10.3%	1 3.4%
Compressed (N=28)	2 7.1%	5 17.9%	11 39.3%	6 21.4%	2 7.1%	1 3.6%	-	1 3.6%
Total (N=57)	2 3.5%	7 12.3%	14 24.6%	19 33.3%	6 10.5%	4 7.0%	3 5.3%	2 3.5%

Table 5.24. The count of the increase of the number of the student-generated size groups from labeling task 1 to labeling task 2.

Predictions and Reflections

The students' predictions varied by student group, naturally due to the difference in the accumulation intervals of the two groups (one object / sec vs. ten objects / sec). b. The compressed group

Table 5.25 shows the summary of the spectrums of each student group's predictions for each imperceptible object. The predictions seemed to be inversely relative to the sizes of the objects: the smaller the object, the longer the predicted accumulation duration. Although I could not run One-Way ANOVA, because (1) the student predictions were very sparsely dispersed over a wide range of values, (2) there existed too many extreme outliers, and (3) the distribution was normally curved, I noticed a difference between the student responses. As Table 5.25 shows, the students in the extended group tended to make more extreme predictions (e.g., 900 years for the hydrogen atom or 200 years for the water molecule), which were tremendously longer than the actual accumulation durations in the TAVRs, when compared to the compressed group. This could be related to the longer accumulation interval (one object / sec) of the

TAVRs that the extended group students interacted with.

The extended group			
Objects	Actual total accumulation time	**Student predictions**	
		min	**max**
Atom	~ 115 days	1 minute	900 years
Molecule	~ 23 days	15 sec	200 years
DNA	~5 d	15 sec	30 min
Virus	~12 hours	20 sec	0.7 days
Bacteria	8 min 20 sec	10 sec	5 min
Red blood cell	2 min 40 sec	3 sec	25 seconds

b. The compressed group

The compressed group			
Objects	Actual total accumulation time	**Student predictions**	
		min	**max**
Atom	~12 days	8 seconds	2 days
Molecule	~4 days	5 sec	5 hours
DNA	~14 hours	2 sec	30 min
Virus	~1 hours	7 sec	15 min
Bacteria	~ 50 sec	5 sec	75 sec.
Red blood cell	~ 14 sec	2 sec	4 min

b. The compressed group

Table 5.25. The range of the student predictions for each imperceptible object's accumulation time.

In the reflections, the student reflections on the seven-point Likert scale did not differ by student group. The result of the One-Way ANOVA indicated that the two groups' responses did not differ (p>0.05). I believe this is due to the limit of the assessment tool – the seven-point Likert scale, as discussed with respect to the previous research question. However, although the students' responses did not differ by student group, their reflections tended to be more extreme in general as the differences between their predictions and the actual accumulation duration became larger. However, there also existed students who offered contradictory responses on the reflection task. For example, the student (in the extended group) who predicted that the hydrogen atom would take 900 years did not select "extremely bigger than I thought" on the Likert scale; instead, he marked "extremely smaller than I thought," implying that he just attempted to come up with longest time that he could think of when predicting the accumulation duration because he *"knew that the hydrogen atom was the smallest one among all."*

Card Sorting Task 3

In card sorting task 3, which was conducted about three weeks after card sorting task 2, one student from the extended group could not participate because she was absent that day.

Ordering

About 90% of the students in both groups remembered that the atom was the smallest object; however, most of them could not get the order of the other five imperceptible objects right. See Table 5.26 for the count of the students for each possible ordering task score. As discussed in research question 1, I think this result is natural, because there was no follow-up instruction that reinforced their memorization of the order of the objects. The total ordering task scores of the student groups decreased in ordering task 3 from ordering task 2 (where they all scored 4). Although the mean of the scores of the extended group (M=2.6, SD=0.69) was higher than that of the compressed group (M=2.3, SD=0.67), no significant difference between them was found from the One-Way ANOVA (p>0.05).

Student group	Count of the students for each possible score					
	1	2	3	4	Mean	SD
Extended (N=28)	-	16 55.2%	10 34.5%	3 10.3%	2.6	0.69
Compressed (N=28)	2 7.4%	16 59.3%	8 29.6%	1 3.7%	2.3	0.67
Total (N=56)	2 3.6%	32 57.1%	18 32.1%	4 7.1%	2.4	0.68

Table 5.26. The count of the students for each possible ordering task score for ordering task 3.

Grouping

Most of the students in both groups remembered how many size groups they had created. In particular, the students who created five or six groups in grouping task 2 still remembered the number of the size groups they had created previously. Hence, the statistical analysis of the a priori contrast analysis of grouping task 3 showed the same relationships between the student groups as the findings from grouping task 2: **the extended > the compressed**. See Table 5.27 for the count of the students for each possible number of the size groups. The extended group (M=4.5, SD=0.87) created significantly more size groups than the TVR group (M=3.9, SD=0.81); $F(1, 54)=7.78$, p=0.007.

I checked for differences between individual student answers in grouping tasks 2 and 3. The changes in the number of the size groups tended to be between -1 and +1 (-1: one size group fewer, 0: same number of size groups, +1:one size group more). Only two students in each student group created one size group fewer, and only two students in the extended group and one student in the compressed group made one more size group in grouping task 3. It seems that the student responses remained in their memory in a relatively fresh state because it had been only three weeks since they had taken card sorting task 2.

Student Group (total N.)	Count of the students for each possible number of the size groups							
	1	2	3	4	5	6	Mean	SD
Extended (N=28)	-	-	4 13.8%	10 34.5%	12 41.4%	3 10.3%	4.5	0.87
Compressed (N=28)	-	-	10 37.0%	12 44.4%	4 14.8%	1 3.7%	3.9	0.81
Total (N=56)	-	-	14 25.0%	22 39.3%	16 28.6%	4 7.1%	4.2	0.90

Table 5.27. Minimum, maximum, mean, and standard deviation of each group's performance on grouping task 3.

Labeling

In labeling task 3, the students showed the same pattern of mean scores as they did in labeling task 2: **the extended group > the compressed group.** See Table 5.28 for the summary of the count of the students for each possible labeling task score. The result of a One-Way ANOVA indicated that the scores of the extended group (*M*=7.0 *SD*=1.7) were significantly higher than those of compressed group (*M*=5.6, *SD*=1.5); *F*(1, 54)=10.26, *p*=0.002. No significant correlation was found between the amount of the increase in the scores of labeling task 3 and the scores of labeling task 2 (*p*>0.05).

Student Group (total N.)	Count of the student responses											
	4	5	6	7	8	9	10	11	12	13	Mean	SD
Extended (N=28)	-	5 17.2%	7 24.1%	8 27.6%	5 17.2%	2 6.9%	1 3.4%	-	-	1 3.4%	7.0	1.7
Compressed (N=28)	5 18.5%	11 40.7%	7 25.9%	-	2 7.4%	1 3.7%	1 3.7%	-	-	-	5.6	1.5
Total (N=56)	5 8.9%	16 28.6%	14 25.0%	8 14.3%	7 12.5%	3 5.4%	2 2.6%	-	-	1 1.8%	6.4	1.8

Table 5.28. Minimum, maximum, mean, and standard deviation of each group's performance on the grouping task in card sorting task 3.

Interestingly, the means of the scores of both groups were higher than those for labeling task 2. The mean of the extended group for labeling task 2 was 6.3 (*SD*=1.4) and the mean of the compressed group was 4.9 (*SD*=1.3). These increases in the scores of both groups in labeling task 3 imply that their interpretation of the accumulation duration of the smallest group (which involves the hydrogen atom) became "exaggerated" over the passage of time. When I asked one student in the extended group why he used three more words in the label for the smallest group, he said, *"I thought this is exactly what I wrote in the answer last time."* Another student in the extended group responded, *"I've changed my mind [and used 2 more words] because now I feel like 200 days [this student was incorrectly remembering the accumulation duration of hydrogen atom] is a very long time."*

<u>Summary</u>

The findings from the data for research question 2 can be summarized as follows:

1. *The TAVRs with the extended accumulation interval were more effective:* The students who interacted with the TAVRs with the extended accumulation interval (one object in one second) generated more size groups and used more descriptive words for the groups than the students who interacted with the TAVRs with the compressed accumulation interval (ten objects in one second).
2. Regardless of the interval difference, the range of imperceptible sizes in the students' mental models became expanded over the passage of time after the learning activity. In card sorting task 3, the labeling task scores of the students in both groups increased (maintaining the significant difference), while the students tended to believe that their labels were the same as the ones they used in card sorting task 2. Considering that

most of the students correctly remembered the number of the size groups, it is thought that the students either intended to emphasize the size of the smallest group or developed mental models of an exaggerated range of imperceptible sizes over time after the learning activity.

Research Question 3

How do active temporal manipulations vs. passive observation by students influence their perceptions of the durations of the temporal experiences with TAVRs and the conceptualization of the range of imperceptible sizes?

Survey

The two questions that were asked of the students in the survey were (1) if they had heard of the names of each object, and (2) if they thought the sizes of the objects were smaller than the size of the head of a pin (which was shown to the students). See Table 5.29 for the summary of each student group's responses to the two survey questions. The number of the student answers to the second question includes responses only from the students who answered, "yes" to the first question. Hence, the total number of responses to the second question is equal to the number of the students who answered "yes" to the first question.

As the tables show, almost all students in all three groups had heard of the names of all the objects except mitochondria; only one or two students in each group had heard of an object called mitochondria. Although many students said that they were familiar with the names of the flu virus, e coli bacterium, red blood cell, and skin cell, about 20% – 30% of the students in each group answered that they thought the sizes of the objects were bigger than the diameter of the pinhead. The ratio of such students was bigger for the No-Manipulation group (NM group) than for the Automatic-Manipulation group (AM group) and the Interactive-Manipulation group (IM group). About 35% of the NM group responded that they thought the sizes of the flu virus and e coli bacterium were bigger than the diameter of the pinhead, while less than 25% of the students in the other groups responded so. However, no significant group difference was found using One-Way ANOVA ($p > 0.05$).

Question		**Student responses (No-Manipulation group, N=23)**				
		Flu virus	Mitochondria	e Coli bacterium	Red blood cell	Skin cell
Heard?	Yes	23	1	22	20	23
	No	-	22	1	3	-
Smaller than pinhead?	Yes	15	-	14	16	16
	No	8 34.8%	-	8 36.3%	4 20.0%	7 30.4%

a. No Manipulation group (NM)

Question		**Student responses (Automatic Manipulation group, N=23)**				
		Flu virus	Mitochondria	e Coli bacterium	Red blood cell	Skin cell
Heard?	Yes	23	2	23	20	23
	No	-	21	-	3	-
Smaller than pinhead?	Yes	17	-	18	17	18
	No	6 26.1%	-	5 21.7%	3 15%	5 21.7%

b. Automatic Manipulation group (AM)

Question		Student responses (Interactive Manipulation group, N=24)				
		Flu virus	Mitochondria	e Coli bacterium	Red blood cell	Skin cell
Heard?	Yes	24	2	23	21	24
	No	-	22	1	3	-
Smaller than pinhead?	Yes	20	-	17	18	19
	No	4 16.7%	-	6 26.0%	3 14.3%	5 20.8%

c. Interactive Manipulation group (IM)

Table 5.29. The student responses to the survey questions.

Card Sorting Task 1

Ordering

The result showed that none of the students in any group had the order of the objects correct. See Table 5.30 for a summary of the count of the students for each ordering score and the descriptions. No difference between the student groups was found using One-Way ANOVA ($p>0.05$). Most of the students in both groups scored 2 points on ordering task 1; almost 70% of the students in each group scored 2 points. There were only one or two students who achieved 3 points, and no student scored 4 points, which is the perfect score.

Student group	Count of the student scores						
	0	1	2	3	4	Mean	SD
NM (N=23)	-	7 30.4%	15 65.2%	1 4.3%	-	1.7	0.5
AM (N=23)	-	6 26.1%	16 69.6%	1 4.3%	-	1.8	0.5
IM (N=24)	-	5 20.8%	17 70.8%	2 8.3%	-	1.9	0.5
Total (N=70)	-	18 25.7%	48 68.6%	4 5.7%	-	1.8	0.5

Table 5.30. The number of the students for each score (0-3) for ordering task 1.

Grouping

In grouping task 1, most of the students overestimated the sizes of the imperceptible objects and underestimated the range of imperceptible sizes. See Table 5.31 for the minimum, maximum, mean, and standard deviation of the results, and the count of the students for each possible number of the size groups. About 70% of the students across all groups (*N*=15-17) generated two size groups, and only five or six students in each group created three size groups. As the table shows, over 60% of the students in all groups created two or three size groups. The One-Way ANOVA did not show any significant difference between the groups ($p>0.05$).

Student Group (total N.)	**Count of the student responses**						
	1	**2**	**3**	**4**	**5**	**Mean**	**SD**
NM (N=23)	2 8.7%	16 69.6%	5 21.7%	-	-	2.13	0.55
AM (N=23)	2 8.7%	15 65.2%	6 26.1%	-	-	2.17	0.58
IM (N=24)	2 8.3%	17 70.8%	5 20.8%	-	-	2.13	0.54
Total (N=70)	6 8.6%	48 68.6%	16 22.9%	-	-	2.14	0.55

Table 5.31. Count of the students for each possible number of the size groups for grouping task 1.

Labeling

The results of labeling task 1 indicate that the students did not much care to use descriptive words to express the sizes of either the smallest group or the biggest group. To name the smallest group, the students tended to use a minimal number of adverbs (e.g., “very,” “super,” “really,” “extremely”) to emphasize the adjectives (e.g., “small,” “tiny,” “mini”) that describe the size of the smallest object. Table 5.32 shows the distribution of students (of each group) across the scores of labeling task 1. Although most of the students clearly distinguished the different sizes of the groups in the labels, the use of the descriptive words in the labels tended to be minimal. The One-Way ANOVA did not show any significant difference between the groups (p>0.05). No significant correlation between the number of the size groups and the labeling scores was found.

Student Group (total N.)	**Count of the students for each score**					
	1	**2**	**3**	**4**	**Mean**	**SD**
NM (N=23)	2 8.7%	9 39.1%	8 34.8%	4 17.4%	2.6	0.7
AM (N=23)	1 4.3%	7 30.4%	13 56.5%	2 8.7%	2.7	0.9
IM (N=24)	2 8.3%	9 37.5%	8 33.3%	5 20.8%	2.7	0.9
Total (N=70)	5 7.1%	25 35.7%	29 41.4%	11 15.7%	2.7	0.8

Table 5.32. Count of the students for each possible score for labeling task 1.

Card Sorting Task 2

Ordering

In ordering task 2, as in the previous research questions, all students in all groups correctly ordered (scored 5 points) the objects by size. The temporal manipulation style did not influence the ways the students perceived and conceptualized the order of different imperceptible sizes because, as in the previous research questions, they could complete this task by looking only at the total accumulation durations once they understood the inverse relationship between the duration and the represented size.

Grouping

In grouping task 2, the students in all groups had increases in the number of the size groups that they generated. However, their performances varied; while about 70% of the students in the NM group (*N*=16) made four size groups, only 13% of the AM group (*N*=3) and 38% of the IM group (*N*=9) generated four size groups. Hence, unlike my hypothesis, it seemed that the NM group, who did not manipulate the temporal experience, perceived the accumulation duration to be longer than both the IM and AM groups. See Table 5.33 for the summary of the result. Out of a possible number of five size groups, no student generated one or five size groups. Only three students in the AM group made two size groups.

The results of the statistical analysis did NOT support the hypothesis (IM > NM > AM), except that the IM and NM groups performed better than the AM group. The results of the a priori contrast analysis showed that the NM group created significantly more size groups (M=3.7, SD=0.5) than the IM group (M=3.4, SD=0.5); t=2.28 (df=45.0), p=0.028 (one-tailed), and the IM group made significantly more size groups than the AM group (M=3.0, SD=0.5); t=4.76 (df=43.53), p<0.001, (one-tailed). Hence, the result indicated, **NM > IM > AM**.

Student Group (total N.)	**Count of the student responses**						
	1	**2**	**3**	**4**	**5**	**Mean**	**SD**
NM (N=23)	-	-	7 30.4%	16 69.6%	-	3.7	0.5
AM (N=23)	-	3 13.0%	17 72.9%	3 13.0%	-	3.0	0.5
IM (N=24)	-	-	15 62.5%	9 37.5%	-	3.4	0.5
Total (N=70)	-	3 4.3%	39 55.7%	28 40.0%	-	3.4	0.5

Table 5.33. Frequencies of the students for each number of the size groups that the students (in each treatment group) generated for grouping task 2.

In the focus group interviews, the students in all groups commented that they made the size groups mainly by looking at the biggest units of time in the total accumulation durations and then trying to interpret the differences between the actual total durations that had the same biggest units of time. However, because the numbers of the size groups differ by the student groups, it implies that each student group differently perceived the accumulation durations of the objects and differently classified the size groups.

As appeared in Table 5.33, the difference between the student groups occurred due to the difference in the number of students who made three or four size groups. The sixteen students in the NM group who made four size groups shared similar object memberships of the groups. For instance, among sixteen students, thirteen students classified the skin cell as the single member of the biggest group, the red blood cell and the e coli bacterium as the second biggest group, the mitochondria as the second smallest group, and finally, the flu virus as the smallest group (skin cell / red blood cell, e coli bacterium / mitochondria / flu virus). The other three students grouped the objects in the same manner, except that they put the skin cell and the red blood cell together into the biggest group and the e coli bacterium alone into the second biggest group (skin cell, red blood cell / e coli bacterium / mitochondria / flu virus). These results imply that, for the students, the difference between 3 seconds and 13 seconds seemed to be bigger than the difference between 13 seconds and 33 seconds.

One out of the three students in the AM group and the nine IM group students, who made four size groups, classified the objects in the same way as the thirteen NM group students (skin cell / red blood cell, e coli bacterium / mitochondria / flu virus). The other two students, who made four size groups, put the skin cell and the red blood cell into the biggest size group, and the rest of the objects on their own single groups (skin cell, red blood cell/ e coli bacterium / mitochondria / flu virus). These results imply that the difference between 3 seconds and 13 seconds looked larger to the students who actively interacted with the TAVRs or were forced to endure the accumulation process than to the students who just needed to watch the accumulation being automatically accelerated.

All thirty-nine students (across all student groups) who made three size groups put the skin cell, the red blood cell, and the e coli bacterium together in the biggest group because they *"all end in several seconds,"* and put the mitochondria and flu virus in two separate groups on their own (skin cell, red blood cell, e coli bacterium / mitochondria / flu virus). Considering that only 30% of the NM group (N=7) made three size groups while 73% of the AM group (N=17) and 63% of the IM group (N=15) made three groups, it seems that the perceived differences between 3 seconds, 13 seconds, and 33 seconds were larger for the students in the AM and IM groups.

It seems that such differences in the student responses are related to the kinds of experiences that the students had when they were waiting for the accumulations to be completed. I observed that while the AM group students did not have a problem of losing focus on the learning activity, the students in the NM group, in particular, kept wondering how long they would have to wait and then began complaining about boredom, which in consequence drove the class almost out of control. The IM group students appeared to be having fun pressing the "skip ahead" buttons, sometimes competing with a peer student in the next seat, although it took about a minute until they realized that they had better press the button arduously. I observed that the arduous button-pressing action made their fingers, hands, and wrists get tired. However, this arduous button-pressing action was merely the boring repetition of simple kinesthetic action with minimal feedback from the simulation (i.e., the accumulation progresses not as quickly as the students expected). Therefore, I believe this kinesthetic action generated negative emotional experiences, which were stored as verbal information of their experiences with the imperceptible sizes. In fact some students commented, *"frustrated but surprising," "boring," "unexpected."*

To put it simply, the NM group had a more irritating experience waiting until the accumulations were completed than the IM group did, and the AM group had the least irritating experience. Such emotional experiences seem to have influenced the way the students perceived and interpreted different durations.

Labeling

In labeling task 2, the number of the descriptive words for the smallest group increased in all groups. Most of the students in all groups tended to use more adjectives, adverbs, and exclamation marks to emphasize the size of the smallest group than they did in labeling task 1. However, the result was inconsistent with the hypothesis, which predicted that the scores of the student groups would be highest in this order: IM > NM > AM. The NM group's score was significantly higher than the IM group, and the score of the IM group was similar to that of the AM group (hence, NM > AM = IM). See Table 5.34 for the minimum, maximum, mean, and standard deviation of the student scores of each treatment group. The a priori contrast test results showed that the NM (M=5.0. SD=0.8) group had a significantly higher score than the AM group (M=4.3, SD=0.7); t=4.08 (df=44.23), p<0.001, (one-tailed). The AM group performed similarly to the IM group (M=4.2, SD=1.0); p<0.001, one-tailed. Hence, the results can be summarized as: **NM > IM = AM**.

Student Group (total N.)	Count of the students for each score						
	3	**4**	**5**	**6**	**7**	**Mean**	**SD**
NM (N=23)	-	5 21.7%	13 56.5%	4 17.4%	1 4.3%	5.0	0.8
AM (N=23)	3 13.0%	11 47.8%	9 39.1%	-	-	4.3	0.7
IM (N=24)	2 8.3%	18 75.0%	2 8.3%	2 8.3%	-	4.2	0.7
Total (N=70)	5 7.1%	34 48.6%	24 34.3%	6 8.6%	1 1.4%	4.5	0.8

Table 5.34. Frequencies of the students for each number of the size groups that the students (in each treatment group) generated for labeling task 2.

While 75% of the IM group (N=18) scored 4 points, only 22% of the NM group (N=5) and less than a half of the AM group (N=11) achieved 4 points. The scores of about 80% of the students in the NM group (N=18) were 5 points or higher, but only 40% of the AM group (N=9) and 16% of the IM group (N=4) scored 5 points or higher.

Card Sorting Task 1 vs. Card Sorting Task 2

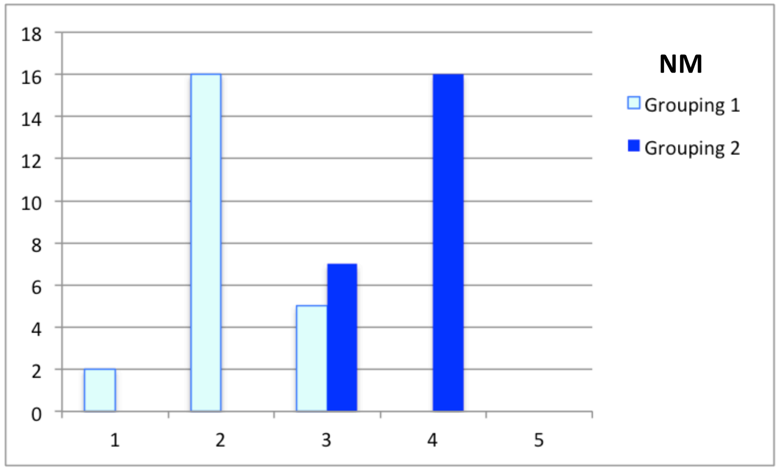
Ordering

In ordering task 1, all students performed poorly, and then all students correctly ordered the objects in ordering task 2, implying that the different combinations of the augmenting modalities did not have different influences on the way the students perceived and interpreted their temporal experiences.

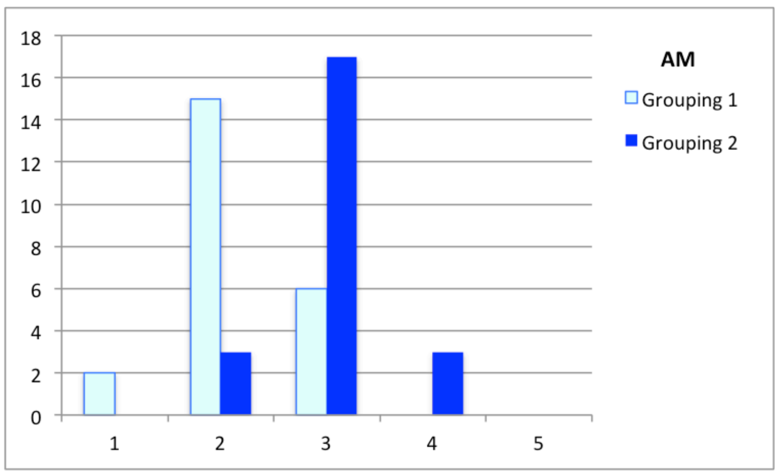
Grouping

As discussed in the previous section, the student groups did not significantly differ from each other in their results for grouping task 1. However, as discussed earlier, the results of the a priori contrast analysis of the results from grouping task 2 showed: NM > IM > AM. These results are represented in the bar graphs in Figure 5.6. They show the number of the students (Y-axis) for the possible numbers of the student-generated size groups (X-axis), by each student group (NM, AM, and IM), in both grouping task 1 (represented in light blue bars) and grouping task 2 (blue bars). The graphs indicate that, in general, the students in all three groups exhibited similar patterns of size groups in grouping task 1. Most of the students in each group created two size groups in grouping task 1. However, in grouping task 2, the NM group had more students who made four size groups than other student groups. About 70% of the students in the NM group (N=16) made three size groups, while only 13% of the AM group (N=3) and 38% of the IM group (N=9) created four size groups. 74% of the AM group (N=17) and 63% of the IM group (N=15) made three size groups.

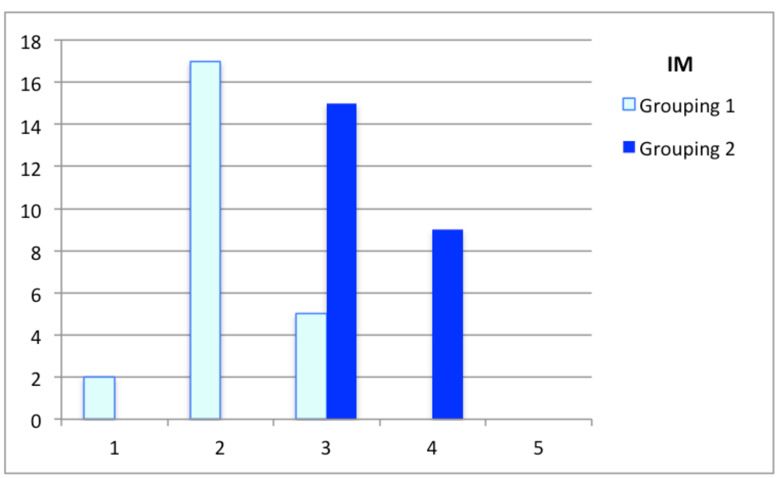
Table 5.35 presents the count of the students for each difference between the number of size groups in grouping task 2 and grouping task 1. For example, the table indicates that twelve students of the NM group made two more size groups in grouping task 2 than they did in grouping task 1. Almost 57% of the students (N=13) of the NM group increased the number of the size groups by two or more, while only about 8.7% of the AM group (N=2) and 25% of the IM group (N=6) had increases of two or more.



a. NM



b. AM



c. IM

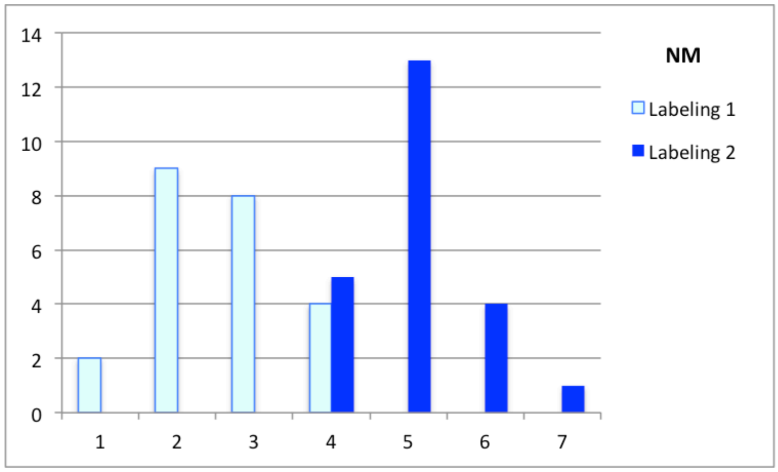
Figure 5.6.The bar graphs of each group's performances on grouping tasks 1 and 2.

Student Groups	The count of the increase of the number of the student-generated size groups from grouping task 1 to grouping task 2.				
	0	1	2	3	4
NM (N=23)	1 4.3%	9 39.1%	12 52.2%	1 4.3%	-
AM (N=23)	4 17.4%	17 73.9%	2 8.7%	-	-
IM (N=24)	1 4.3%	17 70.8%	5 20.8%	1 4.2%	-
Total (N=70)	6 8.6%	43 61.4%	19 27.1%	2 2.9%	-

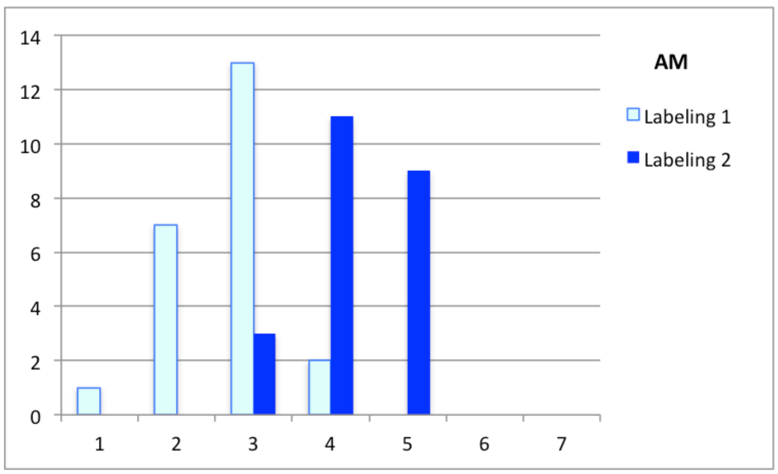
Table 5.35. The count of the increase of the number of the student-generated size groups from grouping task 1 to grouping task 2.

Labeling

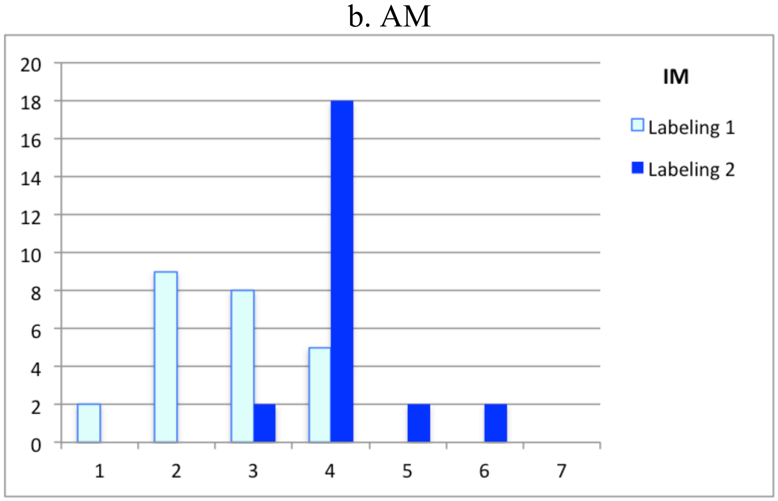
The graphs in Figure 5.7 show the number of the students (Y-axis) for the scores of the labeling task (X-axis) by each student group in both grouping task 1 (represented in light blue bars) and grouping task 2 (dark blue bars). The graphs indicate that the students in all groups exhibited similar results in labeling task 1. The maximum scores of all groups were 4 (minimum was 1), and over 70% of the students in each group (the NM group: 75%, N=17; the AM group: 87%, N=20; the IM group: 71%, N=17) scored 2 or 3 points. In labeling task 2, the mean scores of all groups were higher than those in labeling task 1. Overall, the minimum and the maximum scores of all groups increased. However, the distribution of the scores differed across the groups; the minimum score of the NM group changed from 1 point to 4 points, and that of the AM and the IM group changed from 1 point to 3 points. The maximum score of the NM group changed from 4 points to 7 points, and that of the AM group changed from 4 points to 5 points, and the IM group from 4 points to 6 points.



a. NM



b. AM



c. IM

Figure 5.7. The bar graphs of each student group's scores for labeling tasks 1 and 2.

Table 5.36 presents the count of the individual increases in the scores from labeling task 1 to labeling task 2 (calculated by subtracting the score of labeling task 1 from that of labeling task 2). While about 40% of the students (N=9) in the NM group had increases in their scores of 3 points or more in labeling task 2, only about 17% of the AM group and IM group (N=4) had increases of 3 points or more.

Student Groups	The count of the increase in the numbers of the scores from labeling task 1 to labeling task 2.					
	0	1	2	3	4	5
NM (N=23)	-	3 13.0%	11 47.8%	6 26.1%	2 8.7%	1 4.3%
AM (N=23)	4 17.4%	6 26.1%	9 39.1%	4 17.4%	-	-
IM (N=24)	3 12.5%	10 41.7%	7 29.2%	4 16.7%	-	-
Total (N=70)	7 10.0%	19 27.1%	27 38.6%	14 20.0%	2 2.9%	1 1.4%

Table 5.36. The count of the increase in the number of the student-generated size groups from labeling task 1 to labeling task 2.

Predictions and Reflections

The students' predictions were similar across the groups. Table 5.37 shows the summary of the spectrums of the student predictions for each imperceptible object. Although no significant differences between any student groups were found from the a priori contrast analysis or from the post-hoc analysis of One-Way ANOVA, the student predictions were inversely related to the sizes of the objects. In other words, the smaller the object, the longer the predicted accumulation duration.

The student reflections on the seven-point Likert scale did not differ by student group. The result of the a priori contrast analysis using the hypothesis (IM > NM > AM) indicated that their responses did not differ by their groups ($p>0.05$); hence,

Objects	Actual total accumulation time	Student predictions	
		min	max
Skin cell	3 seconds	1 sec	10 sec
Red blood cell	13 seconds	1 sec	10 sec
e-coli bacterium	33 seconds	1 sec	44 sec
Mitochondria	1 minute 40 seconds	1 sec	30 sec
flu virus	12 minutes 49 seconds	1 sec	50 sec

Table 5.37. The range of the student predictions for each imperceptible object's accumulation time.

IM=NM=AM. As in the ordering task, it seemed that the types of the temporal manipulations did not affect the students in refining their mental models because this task only required them to look and compare the accumulation durations in order to respond on the scale. Although the students' responses did not differ by student group, their reflections tended to be more extreme as the differences between their predictions and the actual accumulation durations became larger.

<u>Card sorting task 3</u>

In card sorting task 3, which was conducted about four weeks after card sorting task 2, the numbers of the participants from each group had to be reduced by one or two either because they were absent or because they had to participate extracurricular activities. I present the total number of the participants in each student group in the tables of the results in the following.

Ordering

All of the students in all groups remembered that the flu virus was the smallest object and took about twelve minutes to accumulate and the skin cell was the biggest and needed only three seconds to accumulate. However, many of them could not get the order of the other three imperceptible objects right. See Table 5.38 for the count of the students for each possible ordering task score. I think this result is natural because there was no follow-up instruction that reinforced their memorization of the order of the objects.

See Table 5.38 for the count of the students for each possible score of labeling task 3. Although the total ordering task scores of all student groups decreased in ordering task 3, the mean scores of the NM group and the IM group were similar to each other and higher than that of the AM group. However, no significant difference between the groups was found from the a priori contrast analysis (p>0.05).

Student group	1	2	3	4	Mean	SD
NM (N=21)	1 4.8%	12 57.1%	6 28.6%	2 9.5%	2.4	0.7
AM (N=22)	2 9.1%	12 54.5%	7 31.8%	1 4.5%	2.3	0.7
IM (N=23)	1 4.3%	16 69.6%	5 21.7%	1 4.3%	2.3	0.6
Total (N=66)	4 6.1%	40 60.6%	18 27.3%	4 6.1%	2.3	0.7

Table 5.38. The count of the students for each possible ordering task score.

Grouping

The statistical analysis of the a priori contrast analysis of grouping task 3 showed different relationships between the student groups from those found in grouping task 2. In grouping task 2, the NM group made significantly more size groups than the IM group, and the IM group made significantly more size groups than the AM group (NM > IM > AM). However, in grouping task 3, the mean scores of the IM and the AM groups remained similar to the results of grouping task 2, but the mean of the NM group decreased. See Table 5.39 for the count of the students for each possible number of size group, the minimum, the maximum, the mean, and the standard deviation. The results for the NM group (*M*=3.3, SD=0.46) were similar to those of the IM group (M=3.2, SD=0.42), p>0.05, but the NM group created significantly more size groups than the AM group (*M*=3.0, *SD*=0.49); *t*=2.29 (*df*, 41.0), *p*=0.027, one-tailed; hence, **NM = IM > AM**.

Student Group (total N.)	**Count of the student responses**						
	1	**2**	**3**	**4**	**5**	**Mean**	**SD**
NM (N=21)	-	-	15 71.4%	6 28.6%	-	3.3	0.46
AM (N=22)	-	3 13.6%	17 77.3%	2 9.1%	-	3.0	0.49
IM (N=23)	-	-	18 78.3%	5 21.7%	-	3.2	0.42
Total (N=66)	-	3 4.5%	50 75.8%	13 19.7%	-	3.2	0.47

Table 5.39. Minimum, maximum, mean, and standard deviation of each group's performance on the grouping task in card sorting task 3.

This result seems to be related to the way the students in the NM group classified the skin cell and the red blood cell. Many of the students who classified these objects separately in two different groups (making the skin cell the sole member of the biggest group), and generated four size groups got confused by the common word in the names of the objects: "cell." In fact, the students whom I briefly interviewed after card sorting task 3 stated that they grouped the skin cell and the red blood cell together, because they are cells, so the students thought that their sizes would be similar.

Labeling

In labeling task 3, the students showed the same pattern of the mean scores as they did in labeling task 2: **NM > IM = AM.** The a priori contrast analysis revealed that the IM group (*M*=4.1 *SD*=0.82) performed similarly to the AM group (*M*=4.4, *SD*=1.0);

$p>0.05$, one-tailed, but the NM group (M=5.1, SD=0.96) achieved significantly higher scores than the AM group; t=2.60 (df=41.0), p=0.013, one-tailed. The mean scores of all groups were higher than those from labeling task 2. See Table 5.41 for the comparison of the mean scores of each student group in labeling tasks 2 and 3.

Student Group (total N.)	**Count of the student responses**						
	3	**4**	**5**	**6**	**7**	**Mean**	**SD**
NM (N=21)	-	7 33.3%	5 23.8%	8 38.1%	1 4.8%	5.1	0.96
AM (N=22)	4 18.2%	10 45.5%	4 18.2%	4 18.2%	-	4.4	1.00
IM (N=23)	5 21.7%	11 47.8%	6 26.1%	1 4.3%	-	4.1	0.82
Total (N=66)	9 13.6%	28 42.4%	15 22.7%	13 19.7%	1 1.5%	4.5	1.00

Table 5.40. Minimum, maximum, mean, and standard deviation of each group's performance on the grouping task in card sorting task 3.

Student Group (total N.)	Labeling task 2		Labeling task 3	
	Mean	SD	Mean	SD
NM (N=21)	5.0	0.67	5.1	0.96
AM (N=22)	4.3	0.70	4.4	1.00
IM (N=23)	4.2	0.72	4.1	0.82
Total (N=66)	4.5	0.77	4.5	1.01

Table 5.41. The mean and standard deviation of each student group in labeling tasks 2 and 3.

Summary

The findings from the data for research question 3 can be summarized as follows:

1. *The naturalistic temporal experience was most effective*: Although the students in all groups were provided with the same results of the accumulation durations in the

TAVRs they interacted with, their interpretations of the durations of time became different according to the types of temporal manipulation. The naturalistic temporal experience (no manipulation features – the NM group) was most effective for producing more sizes groups and the use of more descriptive words in the labels than the kinesthetic (the IM group) or automatic manipulation (the AM group) of the temporal experiences.

2. *The interactive manipulation was more effective than the automatic manipulation:* The perceptually abbreviated passage of time, even though it was accomplished by tiring kinesthetic manipulation, resulted in mental models of a narrower range of imperceptible sizes than the naturalistic passage of time. Although not as effective as the NM condition, the IM condition was still significantly more effective than the AM condition for generating more differentiated perceptions of the passage of time. The interactive manipulation of the accumulation process influenced the way the students interpreted the difference between the accumulation durations of different imperceptible objects. The students considered their kinesthetic experience as an active input that they had to carry out in order to see the completion of the accumulations. Although the actual perceived accumulation durations were much shorter than what was shown on the clocks in the TAVRs, the tiring sensation that the students felt facilitated them to conclude: the more tiring, the smaller the object.
3. *The boredom of natural passage of time resulted in the most augmented temporal experience:* The focus group interview revealed that the NM group students, who generated more size groups and used more descriptive words in labels, had a more irritating experience waiting until the accumulations were completed than the IM

group did, while the AM group had the least irritating experience. Such emotional experiences seem to have different influences on the way the students perceived and interpreted different durations of object accumulations.

CHAPTER 6 DISCUSSION

Development of the Student Mental Models of the Range of Imperceptible Sizes

The Effects of Augmenting/Manipulating Features

In card sorting task 1, which the students had to take as a pretest, the students showed alternative conceptions of the range of imperceptible sizes that were discussed in previous research that studied students' conceptions of the range of different sizes (Tretter, Jones, Andre, et al., 2006; Tretter, Jones, & Minogue, 2006). The students, across different treatment groups, overestimated the sizes of each imperceptible object and underestimated the differences between the sizes of the objects; they incorrectly ordered the objects by size, generated two size groups on average, and labeled the size groups with only two or three descriptive words.

The increased scores of the student performances in card sorting task 2, compared to those from card sorting task 1, implied that the students' mental models of the range of imperceptible sizes were reconstructed into more refined ones, regardless of the students' treatment groups. This indicated that the temporal representations of the imperceptible sizes were useful for supporting learners in constructing a more refined mental model of the range of imperceptible sizes, no matter what kinds of augmenting/manipulating features they interacted with. The students' responses showed that the temporal experiences were particularly useful for refining their mental models of objects with extremely small sizes, such as an atom or a molecule. Considering that students tend to develop the most naïve conceptions of the sizes of such extremely small objects (Tretter, Jones, Andre, et al., 2006), these temporal representations can be useful for supporting students in refining their mental models of the range of imperceptible sizes.

However, in-depth examinations of the student tasks and focus group interviews revealed that the final products of the students' mental models differed by the augmenting/manipulating features of the temporal representations that they interacted with. Overall, the students who interacted with the temporal representations that provided (1) the visual representation (TAVR or TVR), (2) the extended accumulation interval, or (3) no feature for temporal manipulation (natural experiences of the passage of time) generated more refined mental models of the range of imperceptible sizes. A common aspect of these features is that they all provide students with additional information that helps them realistically perceive the passage of time with different durations and accumulation progressions. In Table 6.1, I present summaries of the roles of each feature that I investigated in three different research questions, followed by explanations of each feature.

In the first research question, the visual representation was useful for facilitating better awareness of the accumulation process and the elapsed time in TAVRs, regardless of the presence of the aural representation. It provided additional information for displaying the progress of the accumulation and, hence, allowed the students to guess the remaining accumulation time. In this process the visual representation triggered the students to keep checking the clock at three phases: (1) from the beginning of the accumulation and until the first appearance of the visual representation, (2) when the visual representation appeared, and (3) from the first appearance of the visual representation until the completion of the accumulation. The aural representation was

<table>
<tr><th>RQ</th><th>Feature</th><th>Function</th></tr>
<tr><td rowspan="4">1</td><td>Aural and Visual</td><td>▪ The aural and visual representations together provided the best augmentation for the perception and cognition of the accumulation processes.</td></tr>
<tr><td>Visual</td><td>▪ The visual representation was consistently useful regardless of the presence of the aural representation.
▪ The visual representation facilitated better awareness of the accumulation process across three different phases of the accumulation. The students looked at the clock at these three phases:
1) From the beginning of the accumulation to the first appearance of the visual representation,
2) The moment when they notice the visual representation appears,
3) From the first appearance of the visual representation until the completion of the accumulation.</td></tr>
<tr><td>Aural</td><td>▪ Notifies that the accumulation is in progress. The aural representation was not as useful as the visual representation because the clock also worked as an indicator of the progress of accumulation.</td></tr>
<tr><td>None (only temporal)</td><td>▪ Only the clock indicates that the accumulation is in progress.</td></tr>
<tr><td rowspan="2">2</td><td>Extended interval</td><td>▪ Dramatized difference between the accumulation durations.</td></tr>
<tr><td>Compressed interval</td><td>▪ Less dramatized difference between the accumulation durations.</td></tr>
<tr><td rowspan="3">3</td><td>No manipulation</td><td>▪ Realistic perception of the passage of time was more effective than the abbreviated perception of the passage of time.
▪ Although the students developed more refined mental models of the range of imperceptible sizes, they claimed to have had negative emotional experiences (e.g., boredom). However, it seems that the negative emotions influenced the development of a mental model of a wider range of the sizes.</td></tr>
<tr><td>Interactive manipulation</td><td>▪ Although the acceleration of the accumulation was accomplished by kinesthetic manipulation, its effect was not as strong as the no-manipulation condition because the temporal experience was perceptually abbreviated.</td></tr>
<tr><td>Automatic manipulation</td><td>▪ Automatic acceleration of the accumulation was the least effective.</td></tr>
</table>

Table 6.1. Summary of the findings for each augmenting/manipulation feature.

useful for indicating that the accumulation was in progress, especially when the length of the accumulation had not become macroscopic. However, this role was also carried out by the clock; hence, the TAR and the TR groups performed similarly in card sorting task 2 and showed no significant difference between them.

However, although the aural representation did not seem to have played a critical role in augmenting the students' temporal experiences, literature on sonification still sheds light on the potential roles of aural representation. The aural modality can be used as a sonification feature that conveys the amount of the accumulated objects. Sonification is a form of auditory representation, which uses non-speech audio to convey or perceptualize information (McGookin & Brewster, 2004), often used as an alternative or complement to visualization techniques. Sonification is considered to have a high potential in making use of presentations of information in order to understand it in new ways that may accelerate, simplify, and support the information perception process. An observation that sonifications have been successful in supporting visually impaired learners to perceive and conceptualize scientific concepts (e.g., Cohen, Meacham, Skaff, 2006; Lunney & Morrison, 1990; Levy & Lahav, 2011) encourages further exploration of its use for teaching and learning imperceptible phenomena. In such cases, sonifications have been used to translate abstract scientific data (e.g., reactions between chemical particles) into amplitude values of the waveform. The research commonly argues that a sonification feature can be applicable and useful if the data have a temporal sequence, and techniques such as repeating beeps with varying pitch or output power are used to represent the different values in the data. In the TARs, the aural representation did not provide any additional stimuli other than the repeating clicks. Changing the pitch or tone of the aural representation as the accumulation reaches the end might create improved results.

Second, the extended accumulation interval produced more distinctive differences between the final accumulation durations of the imperceptible objects in the TAVRs compared to the compressed intervals, and as a result, the students generated more size groups and more elaborate labels for the smallest group. It did not seem that it was the single interval between accumulations that actually influenced the students' perceptions of the durations; rather, it was the total accumulation durations. The category learning theory proposed by Rosch (1988) suggests a possible reason why such differences between the student groups resulted. The theory proposed that two basic cognitive principles are involved in the process of category learning. The first is to achieve maximum differentiation between categories, and the second is to avoid cognitive overload, which would result from over differentiating and a consequent loss in flexibility in grouping exemplars that share important characteristics. Reflecting on these principles, it seems that the TAVRs with the extended accumulation interval successfully maximized the differentiation between categories (i.e., many students generated five or six size groups in card sorting task 2), but they overly differentiated the differences; hence, many students made more size groups than the three groups that expert scientists would create.

In the third research question, the naturalistic temporal experience (no manipulation features – the NM group) was most effective for supporting the students in distinguishing different durations of time than the interactive (the IM group) or automatic manipulation (the AM group) of the accumulation. It seemed that the perceptually abbreviated passage of time, even though it was accomplished through tiring kinesthetic manipulation, resulted in mental models of a narrower range of imperceptible sizes. Although the students in all groups were provided with the same results of the accumulation durations in the TAVRs they interacted with, the interpretation of the durations of time became different according to the types of the temporal manipulation (including non-manipulation).

Although not as effective as the NM condition, the IM condition was still significantly more effective than the AM condition for generating more differentiated perceptions of the passage of time. It was clear that the interactive manipulation made to the accumulation process influenced the way the students interpreted the differences between the accumulation durations. The kinesthetic sensations that the students felt in their fingers and hands while interactively manipulating the accumulation for about five minutes (to complete the accumulation of flu viruses) were significantly more severe than the sensation they felt for about thirty seconds while manipulating the accumulation of mitochondria. Although the actual perceived accumulation durations were much shorter than what were shown on the clocks in the TAVRs, the students connected their kinesthetic experiences as an active input that they had to carry out to see the completion of the simulations, and the strength of the tiring sensation they felt was inversely related to the sizes of the imperceptible objects: the more tiring, the smaller the object.

<u>Verbal and Non-verbal Information of the Mental Model of Imperceptible Sizes</u>

In Chapter 3, it was discussed that the students' mental models of imperceptible sizes may be shaped by information interwoven from verbal and non-verbal (mental images), based on the literature on mental models. In the results, it was indeed observed that the students actively utilized both verbal and non-verbal information from their learning experiences with the temporal representations when explaining their logic regarding the card sorting tasks. Although this study did not explicitly attempt to detect the existence of the mental images of different modalities (e.g., visual, aural, and kinesthetic), the student responses in the focus group interviews provided insights into how the students interpreted their perceptual experiences with the temporal representations into verbal information and constructed semantic connections between the mental images of the perceptual experiences and the verbal information.

Further, the results from card sorting task 3 showed that the students tended to remember their learning experiences in an exaggerated manner. The students who interacted with the more influential augmenting/manipulating features used more descriptive words in their labels for the smallest size groups. This observation implies that they remembered their learning according to the meanings of the learning experiences, not according to the precise mental images that made up the perceptual experiences. This phenomenon seems to be due to the nature of memory. Cognitive scientists argue that memories are structured by their meanings. For example, people who listened to a story later confidently "recognized" sentences that never appeared in the story as long as the new sentences were consistent with the story's meaning (Bransford & Franks, 1971). In another example, people quickly lost the memory of precise images that made up a picture story (i.e., whether a character faced left or right), but they retained the meaning of point of the story (Gernsbacher, 1985). Cognitive scientists suggest that this is not because people do not store the mental images in their memories; rather, it is because they try to interpret the learning experiences to generate meanings out of them.

Implications for Designing Representations for Learning

Using a Learner-Centered Design Framework for Educational Representations

Constructivist learning theories suggest that effective learning happens when learners can construct new knowledge upon their preexisting knowledge, because learners actively refer to their preexisting knowledge (Jonassen, 1994; Piaget et al., 1977; Wood et al., 1976). The way learners perceive and conceptualize a representation is heavily shaped by what they already know and have experienced. When a learner attempts to comprehend an external representation, he or she retrieves the elements from his or her mental model and compares them with the representation to decode which components of the representation stand for which aspects of the referent (Buckley, 2000; Rohr & Reimann, 1998). This implies that the poorer the preexisting knowledge, the more likely one is to inappropriately interpret an external representation. Hence, the design of a representation for learning requires extra attention to what learners already know and how much they can comprehend.

However, some representations that are used in science education are too abstract and difficult for middle school students because they often do not respect the students' preexisting knowledge and their cognitive capacities. The representations are mostly first created by scientists who possess ample background knowledge that covers not only the necessary domain-specific knowledge, but also the representational literacy. Those representations are introduced to students mostly because they have been conventionally and commonly used in science textbooks and classrooms for a long time. Prior research shows that many students encounter difficulty when trying to understand a representation that was designed from an expert's perspective in three different aspects: (1) figuring out each component of the representation, (2) comprehending the relationship between the components of the representation, and (3) understanding the relationship between two different representations that share the same referent (Kozma et al., 2000). Such challenges occur mainly because learners lack background knowledge and, hence, tend to focus on the surface features of a representation rather than the underlying information that is represented by the features. The representations that have been used for teaching and learning about imperceptible sizes (e.g., powers of ten, measurement units such as nanometer, micrometer, or macroscopic visual representations) are developed and used by scientists as well, and learners face challenges when trying to understand imperceptible sizes using these representations. It has been discussed that learners have many difficulties in understanding imperceptible phenomena, and their poor understanding of various representations has been discussed as one of the main reasons (diSessa, 2004; Kozma et al., 2000; Lemke, 2004). Since learners must depend on representations to understand phenomena at imperceptible scales, the role of representations becomes even more critical in teaching and learning imperceptible phenomena. It is the representations, which were created by domain experts and were inherited from them into middle school classrooms, that cause the problem, not the learners, who are just being learners. One solution to this problem is to provide learners with a representation that respects their preexisting knowledge and directly represents target concepts in its surface features rather than forcing them to use the complex and abstract representations that are used by scientists.

In order to design a temporal representation, literature on how people perceive and conceptualize different sizes was reviewed, then the problems associated with existing representations were examined. Then finally an alternative representation that addresses both the target knowledge and the cognitive capability of learners was designed. In order to identify the interfaces that provide not only effective learning but also a cognitively optimized learning experience to learners, this study assessed different augmenting or manipulating features of the representation. Through this process, a representation that was effective and easy for learners to understand was created.

The representation design approach taken in this study coincides with the principles of learner-centered design (LCD). Learner-centered design (LCD) is a framework that argues that a learning technology tool must be designed around the specific needs of learners to foster learning (Quintana, Krajcik, & Soloway, 2003; Quintana, Soloway, & Norris, 2001; Soloway, Guzdial, & Hay, 1994). According to this principle, because learners do not possess the same domain-specific (and even domain-general) expertise as experts do, a designer of a learning technology must consider three things: (1) the tasks that learners must undertake (to learn), (2) the tools that they can use to deal with those tasks, and (3) the interface for those tools (Soloway et al., 1994). Employing the principles of LCD into designs of representations and using representations that were engineered for learners for science teaching and learning may solve some challenges that have long existed in science education.

Implications for Using Multimodal Representations in Science Learning

Effect of Multiple Representations

In this study, it was found that different modalities in a representation served different roles in relation to providing different reasoning processes for conceptualizing represented sizes. As discussed by several scholars (Ainsworth, 1999; Oviatt & Cohen, 2000; Rapp & Kurby, 2008; Sweller, 2005b; Zhang & Norman, 1994), multimodality allowed the students to benefit from varying computational processes. Different mental images and their products - verbal information – came together and provided meaningful perceptual and cognitive experiences by complementing the weaknesses of one modality (e.g., the aural) with the strengths of another (e.g., the visual). These modalities were combined to achieve a greater level of expressiveness.

This finding is consistent with what researchers say about multiple representations (Ainsworth, 1999; Kozma & Russell, 2005; Richard E. Mayer & Sims, 1994). According to Ainsworth (1999), learners can benefit from using multiple representations in two ways. First, multiple representations may complement each other with regard to their content and computational efficiency and with regard to learner characteristics and preferences. This is primarily because both recall and memory are improved when information is presented in multiple ways. Second, multiple representations may constrain the interpretation of another representation. A combination of representations enables a learner to deal with the material from different perspectives and with different strategies. The manipulation of two different representational formats induces two different paths of insight into the same learning content, and may have synergetic effects on the construction of coherent knowledge structures. In the TAVRs, the individual visual and aural modalities in the TAVRs served as separate representations with different perceptual channels but with the same referent, and together they provided semantically rich and complimentary information to the students.

Number of Modalities and the Effect of a Multimodal Representation

Although learners can benefit from a representation with multiple modalities, the results from research question 3 imply that incorporating more modalities in a representation does not guarantee better learning; rather, what matters is how close the representation is to the referent. If the students' kinesthetic manipulation had been to place the imperceptible objects one by one on the head of the pin, those students would have generated the largest number of size groups and most elaborate labels for the smallest size group. This issue of how to provide realistic perceptual experiences with a multimodal representation could be valid for representations of imperceptible phenomena because learners cannot have direct perceivable experiences of phenomena at imperceptible scales, and such phenomena are mediated only by representations.

Implications for Emotional Experiences and Learning Impact

It is known that positive emotional experiences during multimedia learning are highly correlated with positive learning impacts (Chauncey & Azevedo, 2010; Kaiser & Oertel, 2006; Um, 2008). Accordingly, in this study, the students who interacted with the most effective types of temporal representations (e.g., the TAVRs in research question 1 or the extended interval TAVRs in research question 2) responded that the learning activities were fun, exciting, and worthwhile. They developed positive emotional experiences during the learning activities and performed better in card sorting task 2 than other student groups. The results of card sorting task 3 indicated that they even tended to develop exaggerated memories of the learning experience over a certain period of time.

However, interestingly, it was also noticed that a negative emotional experience also resulted effective learning, in terms of the students' performances. It seems that the verbal and non-verbal information from the irritating experiences that occurred during the waiting process resulted in the formation of a wider conceptual range of imperceptible sizes (as shown in the more elaborate labels for the smallest group in labeling task 2) and more scale groups (more size groups in grouping task 2). The NM group students (for research question 3) who had to watch the accumulation progressing without making any manipulations complained about the boredom of the dull waiting process. However, this made them interpret their temporal experiences to be longer than the students in other groups did; they created more size groups and used more descriptive words in the labels.

This finding is similar with what was observed from the student group that interacted with the TAVRs with the interactive temporal manipulation feature - the "Skip ahead" button (the IM group). Although many students in this group complained about the unpleasant sensation of the muscles in their fingers, hands, and wrists that emerged after a certain period of arduous button clicking, they quickly related such kinesthetic-temporal experiences to the sizes of the imperceptible objects and developed a kind of inverse relationship between the kinesthetic sensation and the size; *"the more painful my hands get, the smaller the object."* Although these students exhibited relatively lower performances on the card sorting tasks than the NM group students, their scores increased in card sorting task 2 and they performed significantly better than the student group that only watched the accumulation being automatically accelerated (the AM group). It was clear that the temporal experiences that were true to the temporal scale or were kinesthetically accelerated were more effective than the temporal experience that was not true to the scale and involved no manipulation, although the former type of temporal experience generated negative emotions, while the latter did not.

Possible Learning Activities Using Temporal Representations

Measurement Units

The goal of designing and researching the TAVRs is not to propose temporal representation as the single learning tool that helps students understand the range of imperceptible sizes; rather, it was to design a learning tool that middle school students can use as a first stepping stone for understanding and using more complex types of representations. In fact, students will have to learn to understand and to use the commonly used size notions in the field of science, such as the powers of ten or absolute sizes, which are still difficult concepts to understand for middle school students. Measurement units such as nanometer or micrometer are meaningless to middle school students unless they know that one nanometer is one billion times smaller than one meter. The TAVRs can come in useful to help them more easily comprehend such scientific notions that involve proportional reasoning with large numbers. This study showed that the concept of the passage of time, when written in the units of time, was a very useful resource for representing abstract information. When contextualized within an advanced curriculum that addresses measurement units or powers of ten, the TAVRs can be used to help students better conceptualize how small one nanometer or one micrometer is.

Imperceptibly Large Sizes

As I discussed previously, it was found that labor-intensive kinesthetic experience over certain duration of time could successfully augment a learning experience with a representation. Based on this finding, exploring the use of the temporal representation to represent sizes that are too big to see (e.g., the size of Earth, solar system, and galaxy) using kinesthetic-temporal representation seems to be a potential research topic. In addition to sub-macroscopic sizes and scales, planetary sizes and scales are other concepts that many middle school students have difficulty in conceptualizing as well

(Tretter, Jones, Andre, et al., 2006). Research in cognitive psychology and linguistics indicates that the concept of space and time (the duration of an event) are tightly interwoven (Boroditsky, 1997; Casasanto & Boroditsky, 2008; Moore, 2006) in a relative relationship. For example, many people conceptualize distances that cannot be viewed holistically by inferring from physical travel time, such as, "it takes only ten minutes by car to get to A, while B is about one hour away from here." On this basis and the findings from this research that exhibited the effects of the temporal modality for perceiving abstract sizes, the possibility of using the kinesthetic-temporal interface to teach learners about sizes that are too big to be holistically experienced seems promising.

Future Research

Combinations of Successful Features

In this study, the TAVR features that were effective for augmenting or manipulating the perception of the temporal representation were identified. The next step could involve exploring combinations of such features (i.e., a TAVR in which students have to place imperceptible objects across the top of a pinhead). The temporal experience might be more dramatically experienced when the extended temporal experience and the kinesthetic experience are combined. However, as it was found in this study that one cannot guarantee that students will use all modalities that are included in a representation and that the modalities will generate the most influential cognitive activity, the combinations of successful features need to be carefully explored as well.

Tactile Modality

In the design and analysis of this study, four different modalities that were involved in the TAVR interaction – temporal, aural, visual, and kinesthetic – were included. However, there exists a modality that was overlooked - the tactile. The tactile modality is a sensory modality affecting to the sense of contact via the skin with external objects. The arduous button clicking action that the students made in order to accelerate the object accumulation in TAVR not only produces kinesthetic sensation in the muscles of their fingers and hands, but also generates tactile sensations on their skin. This aspect was not addressed during the student interviews mainly because this tactile dimension of button clicking had been disregarded and also because the students tended to focus on describing the sensation they had felt in their fingers and hands, which probably was a more dominant sensation.

However, the literature implies that it is possible that the tactile sensation, in addition to the visual, the aural, and the kinesthetic, might have influenced the students' temporal experiences as well. Prior research suggests that tactile interfaces can become particularly effective in a number of application domains in which other communication channels, such as vision and audition, are heavily overloaded or weakened (L. A. Jones & Sarter, 2008; Sarter, 2006). Considering that the visual and the aural modalities were not solely potent in both conveying the object accumulations and augmenting the students' perceptions of time, a need emerges to investigate whether and how the tactile modality, in association with the kinesthetic modality (the button clicking) influences learners' interpretations of their temporal perceptions.

Other Options for Representing a Temporal Event

It was observed that the students' perceptions and interpretations of their temporal experiences were influenced both by the types of the manipulations they applied to the object accumulation and the types of representations that conveyed the accumulation processes. This observation leads to the discussion of what other potential options for both the manipulation and the representation of a temporal event can be incorporated into TAVR for augmenting learners' temporal experiences. In this section, possible options for representing a temporal event are proposed first, and then in the following section the candidates for the temporal manipulation are presented.

In TAVR, although the aural modality was not as effective as the visual modality in providing better awareness of the accumulation process to learners, it was evident that it did provide additional information to the students regarding how objects accumulate. There exists a possibility that it was the way an aural representation was used in TAVR that made it less useful than the visual representation. Recalling that the visual representation was useful, because it facilitated the students in developing better awareness of the accumulation progressions, adding awareness-supporting features to the aural representation in TAVR may improve its effect.

Following this discussion, an aural representation that provides not only the clicks, but also shifting pitch can be suggested. Aural representations with a pitch-shifting feature are occasionally used in devices that are used in situations where visual modalities are unavailable or inappropriate. For example, a visually impaired chemist waits to hear the rising pitch of whistling sound that is played by a thermometer placed in a beaker, in order to become aware that the water in the beaker is boiling (Nees & Walker, 2009). Applied to the design of the TAVR, raising the pitch of the clicking sound in TAVR when the accumulation becomes macroscopic, or raising the pitch even higher as the accumulation reaches to its completion, may help learners become better aware of the critical phases in the accumulation process.

Other Options for Manipulating a Temporal Event

This study found that the kinesthetic manipulations that the students conducted in order to accelerate the object accumulation in TAVR was not as effective as the natural perception of the passage of time, which was accompanied by boredom, in directing the students to perceive duration of time as "long." To the students, the irritating tactile-kinesthetic action was not as powerful as the annoying boredom. However, it cannot be concluded that natural perception of the passage of time that creates boredom is the most effective way of augmenting learners' temporal experiences, because the influence of the tactile-kinesthetic input itself may have been not as equivalently strong as the boredom. If the temporal manipulation method was a type that generated perceptual experiences of either analogous or greater intensity than the boredom, the result of the students' card sorting tasks might have been different. Thus, it remains unknown as to whether it was specifically due to the type of the manipulation interface or due to the magnitude of the strengths of the manipulation interface that affected the differences in the students' performances.

This discussion triggers the development of future research topics that aim to explore the effects of other possible types of tactile modalities for augmenting learners' temporal experiences. Kinesthetic input interface that induce more active and large-scaled movements and, in consequence, generate more intense kinesthetic mental images may result in promoting learners to interpret their temporal experiences to be longer than the real length of time. There exist potential manipulation interfaces of which scales are larger than those of finger-oriented button click actions. For example, a dialing gesture that is performed on an image of a clock or a sliding gesture on a dial requires one to employ his or her arm, and might provide greater augmentation of learners' perception of their temporal experiences. A kinesthetic movement that involves movements of the limbs of the human body, such as running or jumping, might also produce different influence on learners' interpretations of temporal experiences. Interactive motion-sensing input devices like Wii™ or KINECT™ can facilitate such activities in a classroom.

Tactile modality, in a way that was not explored in the previous design of TAVR interactions, can also be studied. Unlike vision and hearing, touch is capable of simultaneously sensing and acting on the environment (Sarter, 2006). Applying this principle, a computer mouse or a unique input device that additionally informs the object accumulation progression in a TAVR by changes in the conditions of its surface, such as changes in temperature or texture, can be designed. In this design, the accumulation progression is conveyed not only by the visual and aural feedback, but also by the tactile sensation (either temperature or texture) that a learner may feel whenever he or she touches the computer mouse or the unique input device.

Another input interface that may enable learners to manipulate a temporal event is speech (or sound) in which a user makes voice commands to issue instructions to the system. For example, the acceleration of object accumulation in TAVR can be initiated whenever a student simply tells the computer, "accelerate," "skip," or more simply, just the sound of clapping hands, although this is not exactly a speech input interface that requires natural language processing. Alternatively, if the speech input processor is advanced, a learner can also attempt to instruct the temporal event simulator in complex sentences using various words. With its increasing expressive power, thanks to the evolution of natural language processing technology, a speech input interface allows one to envision many possible designs of interactive systems. It will be interesting to observe the speech syntax of learners to instruct the program to accelerate the temporal event and how they interpret their temporal experiences afterwards.

Concluding Remarks

This dissertation introduced a temporal representation called, TAVR, which was designed in order to support learners in perceiving and conceptualizing the range of imperceptibly small sizes, and explored the effects of its features that were included to augment learners' temporal experiences of imperceptibly small sizes. One observation that was consistent throughout the pilot study and the present study was that learners possess well-developed concept of time. Although individual learners interpret certain duration of time differently (i.e., one may think one day is short while the other considers it long), learners do discern different lengths of time and attach meanings to the lengths of time in their own ways. Since there exist wide range of units of time (second, minute, hour, day, week, month, year, decade, and etc.) that learners are familiar with, temporal representations can be useful for teaching and learning imperceptible scales or abstract information that are too vast to be directly perceived or comprehended.

This study does not insist that TAVR is the best design of a temporal representation that aims to represent imperceptibly small sizes. It must be emphasized that this dissertation does not limit the use of temporal representations for conveying these concepts. There can exist other concepts that can benefit from using temporal representations. With advancement of digital technologies, what learners can do with technologies in classrooms is becoming more multimodal. Paying more attention to the potential uses of temporal representations in innovative learning technologies for various concepts in different subject domains will expand what learners can achieve.

REFERENCES

Acredolo, C. (1989). Assessing children's understanding of time, speed and distance interrelations. In I. Levin & D. Zakay (Eds.), *Time and human cognition: A life-span perspective.* (pp. 219-257). Oxford England: North-Holland.

Ainsworth, S. (1999). The functions of multiple representations. [Article]. *Computers & Education, 33*(2-3), 131-152.

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. [Article]. *Learning and Instruction, 16*(3), 183-198. doi: 10.1016/j.learninstruc.2006.03.001

Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the Effects of Different Multiple Representational Systems in Learning Primary Mathematics. *Journal of the Learning Sciences, 11*(1), 25 - 61.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Anastopoulou, S., Sharples, M., & Baber, C. (2011). An evaluation of multimodal interactions with technology while learning science concepts. [Article]. *British Journal of Educational Technology, 42*(2), 266-290. doi: 10.1111/j.1467-8535.2009.01017.x

Baddeley, A. (1986). *Working memory*. New York, NY US: Clarendon Press/Oxford University Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423. doi: 10.1016/s1364-6613(00)01538-2

Barak, M., Ashkar, T., & Dori, Y. J. (2011). Learning science via animated movies: Its effect on students' thinking and motivation. [Article]. *Computers & Education, 56*(3), 839-846. doi: 10.1016/j.compedu.2010.10.025

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577-660. doi: 10.1017/s0140525x99002149

Belardinelli, M. O., & Di Matteo, R. (2002). Is mental imagery prominently visual? [Editorial Material]. *Behavioral and Brain Sciences, 25*(2), 204-+.

Boroditsky, L. (1997, Aug 07-10). *Evidence for metaphoric representation: Perspective in space and time.* Paper presented at the 19th Annual Conference of the Cognitive-Science-Society, Stanford, CA.

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. [Article]. *Psychological Science, 13*(2), 185-189.

Borst, G., & Kosslyn, S. M. (2008). Visual mental imagery and visual perception: Structural equivalence revealed by scanning processes. [Article]. *Memory & Cognition, 36*(4), 849-862. doi: 10.3758/mc.36.4.849

Bransford, J. D., & Franks, J. J. (1971). The Abstraction of Linguistic Ideas. *Cognitive Psychology, 2*, 331-350.

Brown, A. L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *The Journal of the Learning Sciences, 2*(2), 141-178.

Buckley, B. C. (2000). Interactive Multimedia and Model-based Learning in Biology. *International Journal of Science Education, 22*(9), 895-935.

Casasanto, D., & Boroditsky, L. (Writers). (2008). Time in the mind: Using space to think about time [doi: DOI: 10.1016/j.cognition.2007.03.004].
Chang, H. Y., Quintana, C., & Krajcik, J. S. (2010). The Impact of Designing and Evaluating Molecular Animations on How Well Middle School Students Understand the Particulate Nature of Matter. [Article]. *Science Education, 94*(1), 73-94. doi: 10.1002/sce.20352
Chauncey, A., & Azevedo, R. (2010). *Emotions and motivation on performance during multimedia learning: how do i feel and why do i care?* Paper presented at the Proceedings of the 10th international conference on Intelligent Tutoring Systems - Volume Part I, Pittsburgh, PA.
Chi, M. T. H. (1997). Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. *Journal of the Learning Sciences, 6*(3), 271 - 315.
Chi, M. T. H. (2005). Commonsense Conceptions of Emergent Processes: Why Some Misconceptions Are Robust. *Journal of the Learning Sciences, 14*(2), 161-199. doi: 10.1207/s15327809jls1402_1
Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121-152.
Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.
Chi, M. T. H., & Roscoe, R. D. (2002). The Processes and Challenges of Conceptual Change In M. L. a. L. Mason (Ed.), *Reconsidering Conceptual Change: Issues in Theory and Practice* (pp. 3-27). Dordrecht, Netherlands: Springer Netherlands.
Cohen, R. F., Meacham, A., & Skaff, J. (2006). *Teaching graphs to visually impaired students using an active auditory interface*. Paper presented at the Proceedings of the 37th SIGCSE technical symposium on Computer science education, Houston, Texas, USA.
de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. [Article]. *Instructional Science, 38*(2), 105-134. doi: 10.1007/s11251-009-9110-0
diSessa, A. A. (2004). Metarepresentation: Native competence and targets for instruction. [Article]. *Cognition and Instruction, 22*(3), 293-331.
diSessa, A. A., & Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *Journal of the Learning Sciences, 13*(1), 77-103. doi: 10.1207/s15327809jls1301_4
Domagk, S., Schwartz, R. N., & Plass, J. L. (2010). Interactivity in multimedia learning: An integrated model. [Article]. *Computers in Human Behavior, 26*(5), 1024-1033. doi: 10.1016/j.chb.2010.03.003
Downs, R. M., & Stea, D. (1973). Cognitive Maps and Spatial Behavior: Process and Products. In R. M. Downs & D. Stea (Eds.), *Image & environment: Cognitive mapping and spatial behavior.* (pp. 8-26). New Brunswick, NJ US: AldineTransaction.
Droit-Volet, S. (2001). Temporal Bisection in Children. *Journal of experimental child psychology, 80*(2), 142-159.
Eames Office. (1977). Powers of Ten.
Encyclopedia Britannica. (2012). Chicago: Encyclopedia Britannica, inc.
Evans, V. (2003). *The structure of time language, meaning, and temporal cognition*.

Amsterdam ; Philadelphia: J. Benjamins.
Friedman, K. C. (1944). Time Concepts of Junior and Senior High School Pupils and of Adults. *The School Review, 52*(4), 233-238.
Friedman, W. (1979). *The Development of Relational Understandings of Temporal and Spatial Terms*.
Frost, C. (2001). The role of mental models in a multi-modal image search. In E. Aversa & C. Manley (Eds.), *Asist 2001: Proceedings of the 64th Asist Annual Meeting, Vol 38, 2001* (Vol. 38, pp. 52-57). Medford: Information Today Inc.
Gentner, D., Imai, M., & Boroditsky, L. (2002). As Time Goes By: Evidence for Two Systems in Processing Space--Time Metaphors. *Language and Cognitive Processes, 17*(5), 537-565.
Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology, 17*, 324-363.
Gillet, A., Sanner, M., Stoffler, D., Goodsell, D., & Olson, A. (2004). *Augmented Reality with Tangible Auto-Fabricated Models for Molecular Biology Applications*. Paper presented at the Proceedings of the conference on Visualization '04.
Goldstone, S., Boardman, W. K., & Lhamon, W. T. (1958). Kinesthetic cues in the development of time concepts. *Journal of Genetic Psychology, 93*, 185-190.
Han, I., & Black, J. B. (2011). Incorporating haptic feedback in simulation for learning physics. *Computers & Education, 57*(4), 2281-2290. doi: 10.1016/j.compedu.2011.06.012
Hasler, B. S., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology, 21*(6), 713-729. doi: 10.1002/acp.1345
Hutchins, E. (1995). *Cognition in the wild*. Cambridge, Mass.: MIT Press.
Hynd, C. R., & Guzzetti, B. (1998). When knowledge contradicts intuition: Conceptual change. In C. R. Hynd, S. A. Stahl, M. Carr & S. M. Glynn (Eds.), *Learning from text across conceptual domains.* (pp. 139-163). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.
Jarman, R. F. (1979). Matching of auditory–visual and temporal–spatial information by seven- and nine-year-old children. *Child Development, 50*(2), 575-577. doi: 10.2307/1129438
Jenkins, M. H. (2010). *The effects of using mental imagery as a comprehension strategy for middle school students reading science expository texts.* (70), ProQuest Information & Learning, US. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2010-99051-321&site=ehost-live Available from EBSCOhost psyh database.
Johnson, H. M., & Seifert, C. M. (1999). Modifying mental representations: Comprehending corrections. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading.* (pp. 303-318). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.
Jonassen, D. H. (1994). Thinking Technology: Toward a Constructivist Design Model. *Educational Technology, 34*(4), 34-37.
Jones, L. A., & Sarter, N. B. (2008). Tactile Displays: Guidance for Their Design and Application. *Human Factors, 50*(1).
Jones, M. G., Minogue, J., Tretter, T. R., Negishi, A., & Taylor, R. (2006). Haptic

augmentation of science instruction: Does touch matter? [Article]. *Science Education, 90*(1), 111-123. doi: 10.1002/sce.20086
Kaiser, R., & Oertel, K. (2006). *Emotions in HCI: an affective e-learning system*. Paper presented at the Proceedings of the HCSNet workshop on Use of vision in human-computer interaction - Volume 56, Canberra, Australia.
Kosslyn, S. M. (1980). *Image and mind*. Cambridge, Mass.: Harvard University Press.
Kosslyn, S. M. (1994). *Image and brain: the resolution of the imagery debate*. Cambridge, Mass.: MIT Press.
Kosslyn, S. M. (2005). Reflective thinking and mental imagery: A perspective on the development of posttraumatic stress disorder. *Development and Psychopathology, 17*(3), 851-863. doi: 10.1017/s0954579405050406
Kosslyn, S. M., Behrmann, M., & Jeannerod, M. (1995). The cognitive neuroscience of mental imagery. *Neuropsychologia, 33*(11), 1335-1344. doi: 10.1016/0028-3932(95)00067-d
Kosslyn, S. M., Murphy, G. L., Bemesderfer, M. E., & Feinstein, K. J. (1977). Category and Continuum in Mental Comparisons.
Kozma, R. B., Chin, E., Russell, J., & Marx, N. (2000). The Roles of Representations and Tools in the Chemistry Laboratory and Their Implications for Chemistry Learning. [Article]. *Journal of the Learning Sciences, 9*(2), 105-143.
Kozma, R. B., & Russell, J. (2005). Multimedia Learning of Chemistry. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning.* (pp. 409-428). New York, NY US: Cambridge University Press.
Kuchera-Morin, J. (2010). Performing in Quantum Space: A Creative Approach to N-Dimensional Computing. *Journal of Arts, Sciences and Technology*.
Lakoff, G. (1987). *Women, fire, and dangerous things : what categories reveal about the mind*. Chicago: University of Chicago Press.
Lakoff, G., & Johnson, M. (1980a). THE METAPHORICAL STRUCTURE OF THE HUMAN CONCEPTUAL SYSTEM. [Article]. *Cognitive Science, 4*(2), 195-208. doi: 10.1207/s15516709cog0402_4
Lakoff, G., & Johnson, M. (1980b). *Metaphors we live by*. Chicago: University of Chicago Press.
Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: the embodied mind and its challenge to Western thought*. New York: Basic Books.
Lamon, S. J. (1994). Ratio and proportion: Cognitive foundations in unitizing and norming. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics.* (pp. 89-120). Albany, NY US: State University of New York Press.
Lee, P., Tversky, B. (1999). *Pictorial and verbal tools for conveying routes.* Paper presented at the COSIT.
Lemke, J. L. (2004). The Literacies in Science. In E. W. Saul (Ed.), *Crossing borders in literacy and science instruction: Perspectives in theory and practice* (pp. 13-32). Newark, DE: International Reading Association/National Science Teachers Association.
Lunney, D., & Morrison, R. (1990). High technology laboratory aids for visually handicapped chemistry students. *Journal of Chemistry Education, 58*(228).
Lyle, J. (2003). Stimulated Recall: A report on its use in naturalistic research. *British*

Educational Research Journal, 29(6), 861-878. doi: 10.1080/0141192032000137349
Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of educational psychology, 83*(4), 484-490. doi: 10.1037/0022-0663.83.4.484
Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. [Article; Proceedings Paper]. *Educational Psychologist, 38*(1), 43-52. doi: 10.1207/s15326985ep3801_6
Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of educational psychology, 86*(3), 389-401. doi: 10.1037/0022-0663.86.3.389
McCormack, T., & Hoerl, C. (2008). Temporal Decentering and the Development of Temporal Concepts. *Language Learning, 58*(s1), 89-113.
McGookin, D. K., & Brewster, S. A. (2004). Understanding concurrent earcons: Applying auditory scene analysis principles to concurrent earcon recognition. *ACM Trans. Appl. Percept., 1*(2), 130-155. doi: 10.1145/1024083.1024087
Miller, L. M., Chang, C. I., Wang, S., Beier, M. E., & Klisch, Y. (2011). Learning and motivational impacts of a multimedia science game. [Article]. *Computers & Education, 57*(1), 1425-1433. doi: 10.1016/j.compedu.2011.01.016
Moore, K. E. (2006). Space-to-time mappings and temporal concepts. *Cognitive Linguistics, 17*(2), 199-244. doi: 10.1515/cog.2006.005
Moreno, R., & Mayer, R. (1999). Multimedia-Supported Metaphors for Meaning Making in Mathematics. *Cognition and Instruction, 17*(3), 215-248.
Moreno, R., & Mayer, R. (2007). Interactive Multimodal Learning Environments. *Educational psychology review, 19*(3), 309-326.
Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of educational psychology, 87*(2), 319-334. doi: 10.1037/0022-0663.87.2.319
Nanoscale Informal Science Education. (2010). Scale Ladder, from http://www.nisenet.org/viz_lab/size-scale
National Academy & Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academy Press.
National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
Nees, M. A., & Walker, B. N. (2009). Auditory Interfaces and Sonification. In C. Stephanidis (Ed.), *The Universal Access Handbook* (pp. 507-522). New York: CRC Press.
Norman, D. A. (1993). *Things that make us smart : defending human attributes in the age of the machine*. Reading, Mass.: Addison-Wesley Pub. Co.
Nugent, G. C. (1982). Pictures, audio, and print: Symbolic representation and effect on learning. *Educational Communication & Technology Journal, 30*(3), 163-174.
Ornstein, R. E. (1997). *On the experience of time*. Harmondsworth: Penguin.
Oviatt, S. (1999). Ten Myths of Multimodal Interaction. *Communications of the ACM, 42*(11), 74-81.
Oviatt, S., & Cohen, P. (2000). Multimodal Interfaces that Process: What Comes Naturally. . *Communications of the ACM, 43* (3), 45-53.

Oviatt, S., Cohen, P., Wu, L., Duncan, L., Suhm, B., Bers, J., . . . Ferro, D. (2000). Designing the user interface for multimodal speech and pen-based gesture applications: State-of-the-art systems and future research directions. *Human-Computer Interaction, 15*(4), 263-322. doi: 10.1207/s15327051hci1504_1

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4. doi: 10.1207/s15326985ep3801_1

Paivio, A. (1986). Mental representations : a dual coding approach. *Oxford psychology series no. 9*

Palmiero, M., Belardinelli, M. O., Nardo, D., Sestieri, C., Di Matteo, R., D'Ausilio, A., & Romani, G. L. (2009). Mental imagery generation in different modalities activates sensory-motor areas. *Cognitive Processing, 10*(Suppl2), S268-S271. doi: 10.1007/s10339-009-0324-5

Pass, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science, 32*(1-2), 1-8. doi: 10.1023/B:TRUC.0000021806.17516.d0

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions* (pp. 47-87). New York: Cambridge University Press.

Pellegrino, J., Krajcik, J., Stevens, S., & al., E. (2008). *Using Construct-Centered Design to align curriculum, instruction, and assessment development in emerging science.* Paper presented at the Proceedings from ICLS '08: International perspectives in the Learning Sciences: Creating a learning world, Utrecht, the Netherlands.

Pfeiffer, D. G. (2008). Listen and learn: An investigation of sonification as an instructional variable to improve understanding of complex environments. [Article]. *Computers in Human Behavior, 24*(2), 475-485. doi: 10.1016/j.chb.2007.02.006

Piaget, J., Gruber, H. E., & Vonèche, J. (1977). *The essential Piaget*. New York: Basic Books.

Pineda, L., & Garza, G. (2000). A model for multimodal reference resolution. [Article]. *Computational Linguistics, 26*(2), 139-193. doi: 10.1162/089120100561665

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences, 25*(2), 157-238. doi: 10.1017/s0140525x02000043

Quintana, C., Krajcik, J., & Soloway, E. (2003). Issues and approaches for developing learner-centered technology *Advances in Computers, Vol 57* (Vol. 57, pp. 271-321). San Diego: Academic Press Inc.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., . . . Soloway, E. (2004). A Scaffolding Design Framework for Software to Support Science Inquiry. *Journal of the Learning Sciences, 13*(3), 337 - 386.

Quintana, C., Soloway, E., & Norris, C. (2001, Aug 06-08). *Learner-centered design: Developing software that scaffolds learning.* Paper presented at the IEEE International Conference on Advanced Learning Technologies, Madison, Wi.

Rapp, D. N. (2005). Mental Models: Theoretical Issues for Visualizations in Science Education. In J. K. Gilbert (Ed.), *Visualization in Science Education* (Vol. 1). Dordrecht: Springer.

Rapp, D. N., & Kurby, C. A. (2008). The 'Ins' and 'Outs' of Learning: Internal Representations and External Visualizations In M. R. a. M. N. John K. Gilbert (Ed.), *Visualization: Theory and Practice in Science Education* (Vol. 3, pp. 29-52). Dordrecht: Springer Netherlands.

Rohr, M., & Reimann, P. (1998). Reasoning with multiple representations when acquiring the particulate model of matter. . In M. W. v. Someren, P. Reimann, H. P. A. Boshuizen & T. d. Jong (Eds.), *Learning with multiple representations* (pp. 41-66). Oxford: Elsevier Science Ltd.

Roland, P., & Gulyas, B. (1994). Visual imagery and visual representation. *Trends in Neurosciences, 17*(7), 281-287.

Rosch, E. (1988). Principles of categorization. In A. M. Collins & E. E. Smith (Eds.), *Readings in cognitive science: A perspective from psychology and artificial intelligence.* (pp. 312-322). San Mateo, CA US: Morgan Kaufmann.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573-605. doi: 10.1016/0010-0285(75)90024-9

Rowe, A. L., & Cooke, N. J. (1995). Measuring mental models: Choosing the right tools for the job. *Human Resource Development Quarterly, 6*(3), 243-255. doi: 10.1002/hrdq.3920060303

Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in Cognition: Extending Human Intelligence with Intelligent Technologies. *Educational Researcher, 20*(3), 2-9. doi: 10.3102/0013189x020003002

Sarter, N. B. (2006). Multimodal information presentation: Design guidance and research challenges. *International Journal of Industrial Ergonomics, 36*(5), 439-445. doi: http://dx.doi.org/10.1016/j.ergon.2006.01.007

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 US: Cambridge University Press.

Schonborn, K. J., Bivall, P., & Tibell, L. A. E. (2011). Exploring relationships between students' interaction and learning with a haptic virtual biomolecular model. [Article]. *Computers & Education, 57*(3), 2095-2105. doi: 10.1016/j.compedu.2011.05.013

Sharona T. Levy, & Lahav, O. (2011). *Listening to Complexity: A Blind Person's learning of Chemistry with a Sonified Model.* . Paper presented at the Chais conference on instructional technologies research: Learning in the technological era, Israel.

Sims, V. K., & Hegarty, M. (1997). Mental animation in the visuospatial sketchpad: Evidence from dual-task studies. *Memory & Cognition, 25*(3), 321-332.

Soloway, E., Guzdial, M., & Hay, K. E. (1994). Learner-centered design: the challenge for HCI in the 21st century. *interactions, 1*(2), 36-48. doi: 10.1145/174809.174813

Song, M., & Quintana, C. (2009). *WIIS: multimodal simulation for exploring the world beyond visual sense*. Paper presented at the Proceedings of the 27th international conference extended abstracts on Human factors in computing systems, Boston, MA, USA.

Song, M., & Quintana, C. (2010). *TAVR: temporal-aural-visual representation to convey*

imperceptible spatial information. Paper presented at the Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems, Atlanta, Georgia, USA.
Stevens, S. Y., Krajcik, J. S., Shin, N., Pellegrino, J. W., Geier, S., Swarat, S., . . . Hagedorn, E. A. (2008). *Using construct-centered design to align curriculum, instruction, and assessment development in emerging science*. Paper presented at the Proceedings of the 8th international conference on International conference for the learning sciences - Volume 3, Utrecht, The Netherlands.
Stieff, M., Bateman, R., & Uttal, D. (2005). Teaching and Learning with Three-dimensional Representations
Visualization in Science Education. In J. Gilbert (Ed.), (Vol. 1, pp. 93-120): Springer Netherlands.
Stix, A. (1996). Creating Rubrics through Negotiable Contracting and Assessment.
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science: A Multidisciplinary Journal, 12*(2), 257-285.
Sweller, J. (2005a). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning.* (pp. 19-30). New York, NY US: Cambridge University Press.
Sweller, J. (2005b). The Redundancy Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning.* (pp. 159-167). New York, NY US: Cambridge University Press.
Taylor, H. A., Renshaw, C. E., & Jensen, M. D. (1997). Effects of computer-based role-playing on decision making skills. *Journal of Educational Computing Research, 17*(2), 147-164. doi: 10.2190/gble-cevt-ywtl-qebg
The Design-Based Research Collective. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher, 32*(1), 5-8.
Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*(4), 257-287. doi: 10.1037/1076-898x.3.4.257
Tourniaire, F., & Pulos, S. (1985). Proportional reasoning: A review of the literature. *Educational Studies in Mathematics, 16*(2), 181-204. doi: 10.1007/bf02400937
Tretter, T. R., Jones, M. G., Andre, T., Negishi, A., & Minogue, J. (2006). Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena. [Article]. *Journal of Research in Science Teaching, 43*(3), 282-319. doi: 10.1002/tea.20123
Tretter, T. R., Jones, M. G., & Minogue, J. (2006). Accuracy of Scale Conceptions in Science: Mental Maneuverings across Many Orders of Spatial Magnitude. *Journal of Research in Science Teaching, 43*(10), 1061-1085.
Tversky, B., Zacks, J. M., & Lee, P. (2004). Events by hands and feet. *Spatial Cognition and Computation, 4*(1), 5-14. doi: 10.1207/s15427633scc0401_2
Um, E. R. (2008). *The effect of positive emotions on cognitive processes in multimedia-based learning.* New York University.
Vallesi, A., Binns, M. A., & Shallice, T. (2008). An effect of spatial-temporal association of response codes: Understanding the cognitive representations of time. [doi: DOI: 10.1016/j.cognition.2007.10.011]. *Cognition, 107*(2), 501-527.
van Oostendorp, H., & Bonebakker, C. (1999). Difficulties in updating mental

representations during reading news reports. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading.* (pp. 319-339). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24*(4), 535-585. doi: 10.1016/0010-0285(92)90018-w

Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science: A Multidisciplinary Journal, 18*(1), 123-183.

Vygotsky, L. S. (1978). *Mind in Society : the development of higher psychological processes*. Cambridge: Harvard University Press.

Waldrip, B., Prain, V., & Carolan, J. (2010). Using Multi-Modal Representations to Improve Learning in Junior Secondary Science. [Article]. *Research in Science Education, 40*(1), 65-80. doi: 10.1007/s11165-009-9157-6

Waldron, A. M., Sheppard, K., Spencer, D., & Batt, C. A. (2005). Too small to see: Educating the next generation in nanoscale science and engineering. In C. S. S. R. Kumar, J. Hormes, C. Leuschner & l. Wiley online (Eds.), *Nanofabrication towards biomedical applications: techniques, tools, applications, and impact* (pp. xxii, 420 p.). Weinheim: Wiley-VCH.

Wiley, J., & Ash, I. K. (2005). Multimedia Learning of History. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning.* New York, NY US: Cambridge University Press.

Wolpert, J. (1964). The Decision-process in Spatial Context. [Article]. *Annals of the Association of American Geographers, 54*(4), 537-558.

Wood, D., Bruner, J. S., & Ross, G. (1976). Role of Tutoring in Problem-solving. [Article]. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 17*(2), 89-100.

Zhang, J., & Norman, D. (1994). The Representation of Relational Information. *Proceedings of the Sixtheenth Annual Conference of the Cognitive Science Society*, 952-957.

Zolna, J. S. (2008). *Two stage process model of learning from multimedia: Guidelines for design.* (69), ProQuest Information & Learning, US. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2008-99200-196&site=ehost-live Available from EBSCOhost psyh database.