Isolated Text as Design Method for Signaling

Learners in a Multimedia Learning Task

by

Joshua Chin

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2017 by the
Graduate Supervisory Committee:

Scotty Craig, Chair
Russell Branaghan
Rod Roscoe

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

Web-based learning resources have been criticized as being developed with minimal consideration as to the effectiveness of the design principles or guidelines used to create them. Extraneous material is oftentimes present and necessary for learners to engage in effective learning with multimedia learning material. Signaling learners towards important information between images and corresponding text has been shown to be an effective method for providing learners a way to quickly find information between the two parts of the learning material. However, not all signaling methods are equally effective in all applications. This study investigates a novel signaling method, using spatial isolation of text, as a way to signal learners in a web-based format compared to a traditional highlighting method and a non-signaled control group. Improved learning performance was observed for knowledge retention using text isolation as the signaling method, but no other significant effects were observed between the other conditions. Additionally, transfer of knowledge across all conditions showed no significant differences either. While minimal support for the effectiveness of isolated text signaling was demonstrated, the statistical means trend across all post-test knowledge assessments suggest that further evaluation of the novel signaling method is justified.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Universities and open access education services around the world have responded to an increased demand for online education by people who otherwise would not be able to attend the more traditional routes to an in-person education. This global push to implement effective online learning has increased the need to provide quality learning materials to meet this demand (EACEA & European Commission, 2014). In order to develop learning materials quickly and efficiently, developers tend to author widely used learning resources, which look and perform as they should but lack proper applications of learning principles and usability design methods (Murray, 2004). While there are suggestions for designing empirically grounded multimedia learning environments, there is little evidence to support the use of specific design methods and guidelines used by user interface (UI) developers (Jin, 2013).

The disconnect between empirically tested pedagogic principles and "best practice" UI design methods necessitate further research to implement effective online learning programs on a large scale. For example, it is common for instructional websites similar to "Instructibles.com" or "eHow.com" to incorporate slideshows/layouts with text-based explanations. Because UI design methods lack empirical grounding for multimedia learning environments, an important step to meeting the global demand for these learning materials is by establishing a relationship between specific design methods and learning principles. Due to the many different design aspects that are used to create a user interface for learning, this current study will investigate the impact that empty space has on learning. Specifically, this study will examine whether the purposeful use of empty space can signal the location of temporally important information to a learner in a web-based learning resource format.

Design Methods

While layout design methods for printed text have been heavily researched, the style guidelines that are often used are, unfortunately, also used for the electronic delivery of text with little research to show that it is effective in multimedia delivery (Jin, 2013; Park & Hannafin, 1993). Furthermore, developers are often instructed to make decisions based on common sense and best practice. For example, Hartley (2004) concludes that when choosing a typeface, it is best to compare individual typefaces instead of deciding whether or not they have a serif; this guideline contradicts the commonly held suggestion to use sans-serif typefaces on low-resolution displays (Lidwell, Holden, Butler, & Elam, 2010). With the development of increasingly higher resolution displays, it is unclear if these guidelines are still applicable. These kinds of guidelines suggest that, depending on the background of the developer, one guideline will be more influential than another.

Beyond textual layout and in a more general sense, best practices of design are, at best, tested in the pre-deployment stage of development through usability testing, at worst, suboptimal design decisions are corrected after users have identified usability issues (Nuland, Eagleson, & Rogers, 2016). Because no single design guideline in any domain can properly inform a developer in the pre-deployment stage of development, continued usability testing of a product should always be necessary following the initial release. However, testing of these design “best practices”, can help to inform developers on the limitations of design guidelines that they are encouraged to follow in the pre-deployment stage. A systematic evaluation to develop a robust design guideline should be a major focus for instructional design in multimedia learning (de Koning, Tabbers, Rikers, & Paas, 2009). This large undertaking would be comprised of numerous experiments designed to evaluate one variable at a time. The issue with this approach is that testing all the combinations between learner individual differences, technologies, design options and etc. would be time-consuming and presumably unable to keep pace with technological advances.

Lidwell, Holden, Butler, & Elam (2010) identify 125 principles of design that developers/designers can apply in numerous design domains. To address each of them would be outside the scope of this work, but by narrowing the design methods that would presumably align with empirically grounded learning principles, one can begin to assess the efficacy of UI design methods which developers use in "best practice". In this current work, *principles of design* will be referred to as design methods to avoid confusion with learning principles.

One specific design method that is commonly used in numerous modern designs, is the purposeful use of empty space. Historically, filling empty space has been a design choice born out of affluence (Elias, 1994; Mortelmans, 2005). Victorian era design styles promote detailed, visually full presentations in most mediums like interior design, pictorial art, clothing, etc. The desire to fill empty space is often referred to as "Horror Vacui", which is a Latin phrase meaning the "fear of emptiness". Mortelmans (2005) describes that the ability to refrain from filling empty space signifies that an individual has self-control over their innate desire to fill empty space in various ways (e.g. art, interior design, resources, etc.). The self-control that he describes has been a result of civilization's behavioral changes over time. Specifically, people were better able to refrain from instincts and lusts (ex. eating with utensils, sleeping in clothes, personal hygiene, etc.), which gave way to materialism and its relationship to human behavior (Coser & Elias, 1978; Mortelmans, 2005).

The modern use of space, however, has evolved into a design choice in which a strategy is formed on how best to present the value of what is being displayed. Furthermore, brand perceptions and the spaces that items are sold (i.e. the interior design of a store), all have to be congruent with the presentation strategy. This type implicit messaging to consumers is common in advertising and is also observed from storefronts. In a study conducted on over 60 clothing stores, Mortelmans (2005) found that retail shops which displayed large amounts of inventory and signage in the window tended to have low-cost items for sale, while luxury stores tended display very few items with little to no signage viewable from the street. Because the evidence in consumer psychology suggests, that conveying the value of objects by ensuring there is empty space around the item/s, it stands to reason that almost anything which is spatially isolated can signal to the observer that the displayed object has a higher value than what surrounds the empty space.

The luxury-item presentation approach seems to have value in introducing the effect to people when a designer chooses to elicit value judgments from customers for any specific purpose. From a cognitive attentional perspective, this effect is well understood and can be described by many of the various models of perception. However, this present study focuses on the use of empty space as a way to improve learning through the design of web-based learning material which uses the same luxury-item presentation approach, spatial isolation. Two influential models of perception describe how people interpret space between objects and noticeability – Gestalt Psychology and Feature Integration Theory.

Models of Perception

Models of visual attention and processing provide comprehensive accounts of how individuals interpret their visual experiences. These models describe our perception at a complex and deep level that is outside the scope of this study. The applications of these models in instructional design are as numerous as the number of possible combinations of signaling methods developed in different technologies. However, interaction with learning material is still mostly limited to one of two senses vision and hearing. There are two major foci which are particularly important in evaluating text isolation as a signaling method. First, how individuals perceive and interpret empty space between objects and second, how and when individuals attend to the objects that are presented to them.

Gestalt psychology. The modern use of empty space (i.e. anti-horror vacui) as a design method largely comes from the progression away from the Victorian era design style, which traditionally fills space with details. Purposeful use of empty space is described in detail within Gestalt theory, which generally refers to a perceptual whole as being more representative than the sum if the perceptual parts (Wolfe et al., 2012). For example, in a visually presented scene, elements that create the scene can be evaluated on an individual basis, but we perceive the scene as a whole regardless of the individual elements. Specifically, the effects of proximity in perceptual grouping depict the tendency of similar objects in space either being more or less grouped together depending on the amount of space between them (Wagemans et al., 2012; Max Wertheimer, 1938). Wertheimer's seminal publication (Max Wertheimer, 1923, 1959) demonstrated this by presenting equally spaced dots on a line, which to the viewer, appear to present as a uniform line. However, when reducing the space between pairs of dots, the proximal pairs appear strongly grouped together.

Two criticisms that are common in regards to the early Gestalt psychologists are that their principles relied heavily on phenomenology and lack quantitative data (Kubovy, Holcombe, & Wagemans, 1998). To address this, more recent researchers have used dot matrices to assess the limitations of perceived grouping. Oyama (1961) demonstrated a perceptual bias by using fixed orientation 4 X 4 matrices and varying the distance horizontally or vertically in 2mm increments. The results suggest that people have a vertical grouping bias, at least for the Japanese population. However, the implicit suggestion is that inter-item distances affect grouping biases, but are affected by culture.

Zucker and Davis (1988) continued to provide evidence of proximity grouping in one experiment that used spaced-dot lines radiating out from a central point with empty space at the center – called the sun illusion. In the sun illusion, the central circle appears brighter than the surrounding empty space when the density of the dotted lines increases. Using three students that were familiar to the task (experiment 1), Zucker and Davis asked the participants to judge whether the sun illusion was present in multiple variations of line density. Their results suggest that there is a noticeable difference below the ratio of 1:4 (dot:space) and the illusion is not present at a ratio of 1:6 or more. The implications of these results suggest that there is a threshold to a person perceiving items as separate or isolated. In the context of UI design, appropriate use of these effects can either convey specific elements as a part of a group or separate items that are intended as distinctly different regardless of their function (Lidwell et al., 2010). Furthermore, this suggests that signaling effects could be induced by varying the amount of space between a featured portion of text within a larger body of text.

Feature-integration theory. Counter to the top-down Gestalt theory which is concerned with the perception as a whole (Gardner, 1987), Feature Integration Theory (FIT) is concerned primarily with elemental stimuli that create perception, a bottom-up approach. In FIT, there is a distinction made between parallel searches and serial searches. For example, suppose a participant's task was to indicate whether they could find a green "T" amongst several brown" T"s, or find a green "T" amongst several green "L"s. Parallel searches are the strategy that people use when there are distinctive features of a specific stimulus amongst distractor stimuli like a green "T" amongst brown "T"s. To contrast parallel searches, serial searches, are when a person has to focus on the small differences in each of the elements in the stimuli like searching for a green "T" amongst green "L"s. The reaction time that a participant would indicate that the target was present would be relatively short compared to a serial search task where the features of the target are similar to the distractor stimuli. The search strategies are different because targets and distractor stimuli have elemental differences that can be perceived in a pre-attentive state, whereas a serial search necessitates focused attention to perceive the elemental qualities of each stimulus in an array, thus increasing the response reaction-time (Treisman & Gelade, 1980). Elemental features of target stimuli in parallel searches have been said to have "pop-out" effects because target stimulus recognition is automatic upon foveal fixation (Quinlan, 2003).

Prior to the development of FIT, Kahneman & Henik (1977) proposed that "group processing" begins hierarchically in the pre-attentive stage of visual processing. They emphasize the role of attentional resources being distributed to features that enable processing of presented items, while Treisman (1982) maintains that attention is the process of selecting features to process. Treisman uses the analogy of a spotlight that varies in intensity to refer to "group processing", while the spotlight in FIT can vary in size and direction. Regardless of each other's views of how grouping information is processed, performance for recall on various grouping experiments has been shown to be an effective means to improve recall of stimuli when presented in rapid serial visual presentations. Furthermore, FIT and "group processing" are both congruent in their depiction of attention being focused on groups as a whole and then to individual elemental features. In other words, first, a person sees groups in parallel searches and then switches to a serial processing strategy to analyze the elemental features of the items that compose a single group.

While the experiments in Kahneman & Henik (1977) and Treisman (1982) focused primarily on a person's ability to briefly see groups of numbers and letters for recall, they describe the perceptual role that attention plays in interpreting space and grouped stimuli, much like the gestalt psychologists of the early 1900s. It is plausible that by designing text layouts to be visually searchable in the pre-attentive stage, signaling effects described by the Cognitive Theory of Multimedia Learning can be achieved. As a result, learning performance outcomes can be improved over traditional text layout presentations in multimedia learning.

The Cognitive Theory of Multimedia Learning

One particularly influential theory that uses interface design as the basis for effective learning materials is the Cognitive Theory of Multimedia Learning (CTML) (R. E. Mayer, 2014b). Researchers have reported both small and large effect sizes throughout the evaluation of CTML, but the trends for different learning outcomes tend to show large effect sizes for knowledge transfer and more varied effect sizes for retention, procedural-motor, and problem-solving (Butcher, 2014). While there is a large amount of effect size variability, overall, the theory has been effective in explaining the limitations of learning principles in the specific contexts used by multimedia instructional designers. Furthermore, CTML provides a thorough account of how individuals process multimedia learning material from the moment of perception to storage into long-term memory.

Three assumptions of CTML. CTML makes three assumptions about learners and their ability to process information. The dual-channel assumption asserts that humans process visual and auditory information in separate channels. For example, if a learner is presented a narrated illustration, the visual information is processed simultaneously and separate from the narration. The dual-channel assumption is similar to widely accepted models of information processing like Paivio's (Clark & Paivio, 1991) Dual-Coding Theory and Baddeley's model of working memory (Baddeley, 2003). The second assumption is that people have limited processing ability per channel. The visual channel enables only a few images to be held and processed in working memory, while the auditory channel allows for a few keywords or phrases representing the auditorily presented information (i.e. not the exact words that were presented) (R. E. Mayer, 2014a). The third assumption asserts that people must actively process information to form a coherent interpretation of their experiences. To learn, people must select relevant information to process, then organize that information in a coherent form, and then integrate that information with prior knowledge structures.

Three memory stores of CTML. The CTML model largely reflects the human information processing system in that it represents three different memory stores (R. E. Mayer, 2014a). First, sensory memory is activated when information is presented to the learner via pictures (visual channel) or words (auditory channel). Sensory memory is thought to have an unlimited capacity, but a very brief duration for holding information. Because words can be presented both visually and auditorily, they are processed in the corresponding channel – i.e. auditorily presented words are processed via the ears and visually presented words are presented via the eyes. The second memory store, working memory (WM), is where the majority of learning takes place and is generally understood to have a limited capacity and short duration for holding information. The CTML model makes the distinction between two phases of processing within WM. First, information entering WM is entering via the visual channel or auditory channel, but the mental representations can change modalities. For example, the auditorily presented word "cat" can be translated into a pictorial form of a cat and processed in its visual form and vice versa. Second, the visual or auditory representation held in WM is then further processed with existing knowledge from long-term memory. The third memory store, long-term memory (LTM), represents the learner's accumulated knowledge. LTM is thought to have an unlimited capacity and able to store information permanently. Any information that is stored in LTM must be actively retrieved and processed in WM to integrate with new information.

Five processes of CTML. CTML asserts that there are five processes that occur when learning from a multimedia environment. The learner must select relevant words (1) or images (2), organize those words (3) or images (4), and then integrate (5) that information with prior knowledge.

When a learner selects relevant words, the learner will receive information through their ears or eyes and must choose which words to hold in WM. Because WM has a limited capacity, the representation that is within WM is not an exact copy of the presentation, rather, the representation is a reflection of what the learner has determined to be important. This strategy is useful in that, according to the dual-channel assumption, there is a limited capacity for processing visual or auditory information, because of this, learners can allocate information to a specific channel so as not to overburden a single processing channel. As mentioned earlier, the modality in which the words are presented dictates the path in which they processed. Depending on the metacognitive strategy the learner uses, words can enter WM visually and be transformed into an auditory representation or vice versa. Similar to selecting words, learners will take the entirety of the visual stimuli (i.e. an illustration or animation) into their sensory memory through their eyes and determine the important information to be processed, because of the limited processing capacity within WM. In determining the important information, the learner might compress multiple images into a single representation as a strategy to reduce the WM load.

After relevant words and images are selected they must be organized into a coherent structure and processed. For words, the learner must organize the selected words into a coherent and efficient or simple representation that Mayer calls the “Verbal

Model”. For images, the process is similar to organizing the selected words with the difference being the processes would most likely take place in the visual channel and the output would produce a coherent representation that Mayer terms the “Pictorial Model”.

The final process of multimedia learning, integration of words and images, is where learning occurs. Whereas the other four processes focus on processing visual and auditory stimuli and representations in their respective channels, integration joins the verbal and pictorial models together into a cohesive representation of the presented material. Furthermore, the integrated representation is processed along with existing knowledge structures in LTM. The success of processing integrated words and pictures is largely dependent upon the efficiency and accuracy of the representations of the organized words and images. These five processes are most likely to occur multiple times in a multimedia presentation – that is, learners create their integrated mental representations segment by segment throughout the multimedia presentation

Three kinds of cognitive processing. In multimedia learning and learning in general, managing appropriate cognitive processes is critical to effective learning. Because learners have a limited WM capacity, it is important to not overburden WM with information that is not conducive to learning. Mayer proposes that there are three types of cognitive processing learners experience in multimedia learning environments – extraneous processing, essential processing, and generative processing. Extraneous processing is described as any cognitive processing that does not support the learning goals and is a result of poor instructional design. Essential processing is defined as the necessary cognitive processes that are inherent within the complexity of the material. That is, the subject matter that makes up the mental representation itself requires the learner to expend effort on creating those mental representations in the selection process (i.e. selecting words and images) and hold them in WM. Generative processing is described as the cognitive processing that a learner experiences when organizing and integrating the learning material and is a result of the learner's motivation to learn the task.

All three of these types of processing must be addressed when designing a multimedia learning environment. The designer must incorporate features to mitigate extraneous processing, manage essential processing, and encourage generative processing if the learning task is to be successful.

Techniques to minimize extraneous processing. There are several principles that researchers have proposed and empirically tested to manage the levels of processing in different multimedia learning scenarios. Since best practice UI design methods tend to focus on aesthetics and usability, rather than empirically tested pedagogical principles of learning, this present study will focus on visual design CTML principles that minimize extraneous processing and have a UI design focused approach.

Coherence principle. The coherence principle refers to the tendency for people to perform better in a learning task when there are no seductive details (i.e. extraneous and irrelevant, but related elements) associated with the learning material. In two experiments, Moreno and Mayer (2000), used learning tasks to test the effects of sound and music that were non-essential to learning material that detailed the processes of lightning formation (experiment 1) and hydraulic brake system operation (experiment 2). In each experiment, participants received illustrations and either narration only, narration and sound effects, narration and looping background music, or narration, sound effects and music as their learning task. Participants that received narration only generally outperformed the other groups on learning material retention and knowledge transfer.

Auditorily presented seductive details are not unique to this effect. In a similar study, where participants were visually presented learning material that used a base text on lightning formation as a control, Harp and Mayer, (1997) conducted an experiment (experiment 1) that added interesting textual details seamlessly into the base text in one group, interesting captioned illustrations as a supplement to the base text in another group, and the final group received both the seductive text and illustrations added to the base text. Their results demonstrated that the learners which received the base only text on retention and transfer.

The results from these studies indicate that attempts to present more engaging learning material by adding factors of entertainment can result in overloading a learner's cognitive capacity and subsequently interfering with the selection, organizing, and integration processes that learners engage in when actively processing learning material (R. E. Mayer, 1999). These results indicate a relationship that can be extrapolated to the formation of the best practice design method in which needlessly filling empty space (i.e. horror vacui) fosters a poor UI design.

Spatial contiguity principle. The spatial contiguity principle refers to the situation in which people perform better in a learning task when words and pictures are spatially close as opposed to spatially distant on the same page or entirely separate page. In an experiment conducted by Moreno and Mayer (1999), participants were subjected to a learning task which required them to learn about lightning formation in three conditions (experiment 1). Participants were provided narrated illustrations in one group, illustrations with integrated text, or illustrations that were spatially distant from the featured portion of the illustration. Their results indicated that learners who received narrated illustrations performed better than the other groups, which is consistent with the CTML. Furthermore, participants that were presented illustrations with spatially close text outperformed participants on retention and transfer tests than the spatially distant group. These results suggest that participants which must hold information in WM for an increased amount of time, have a reduced ability to integrate information and will subsequently perform worse on retention and transfer knowledge tests.

The signaling principle of multimedia learning. The signaling principle of multimedia learning describes the effect that, when learning, signaling important material is more effective when the cues guide the learner's attention. Multiple approaches can be taken to provide a signal or cue that visually instructs the learner to attend to specific information such as highlighting, arrows, or a simple in-text statement that instructs the learner to look at an illustration (van Gog, 2014). Other approaches to signaling important information to a learner have been effective when integrated with narration and animated agents by designing deictic and gestures into the learning material (Craig, Gholson, & Driscoll, 2002). The efficacy and intended level of overt or covert design in the cue highly depend on the appropriate use of a specific signaling method in a particular learning environment. In other words, the type of signaling method used in presenting to-be-learned material must be appropriate for the type of information that is being conveyed.

For example, Florax and Ploetzner (2010) investigated the role of text segmentation, picture labeling, and spatial proximity of learner outcomes when participants are given multi-modal information (text and a diagram) to study. One-hundred and sixty-five students from two different universities were randomly assigned to one of five groups that varied the layout of the learning material. The material to be learned was a diagram with accompanying text that detailed the electrochemical processes that occur in neuronal synaptic transmission. The first group was given a continuous block of text positioned above the diagram that described the synaptic transmission in the context of the diagram. The second group was shown the same text and diagram, except the text, was segmented by sentence, numbered, and placed into columns. The third group was given the continuous block of text and diagram, but the diagram parts were labeled. The fourth group was given segmented text and a corresponding labeled diagram, and the last group was given a diagram with integrated text. The integrated text group saw the text segments distributed throughout the diagram in boxes that had an extended line pointing to the corresponding process in the diagram. Participants were given working-memory capacity tasks, spatial ability tasks, pre- and post-tests. Their results indicated that there were no significant differences in working memory or spatial ability between groups. However, there were significant differences between the group post-test scores. Learners that were given text segmentation and/or labeling performed better than the control group – the group with continuous text and an unlabeled diagram. Learners that received segmented text with a labeled diagram and learners with the integrated text diagram performed significantly better than all other groups. There was not a significant difference between the two best-performing groups, but they partially attribute this effect to the unorganized nature of the integrated text that is scattered throughout the diagram.

Florax & Ploetzner's (2010) use of integrated text within a diagram can be considered as an inappropriate use of the design method because the text detailing the nature of synaptic transmission is serial and not parallel, which is what the design method would imply. However, the implications of these results suggest that reducing visual search by separating important information and making it easily identifiable will have a positive effect on learner outcomes.

While Florax and Ploetzner (2010) provide evidence to support the use segmented text in conjunction with a labeled diagram, their data does not address any degree of efficiency that may be a determining factor in forming associations between multi-modal learning material and learning. Ozcelik, Karakus, Kursun, and Cagiltay (2009) implemented an eye-tracking study to determine why color-coding is effective at improving learner outcomes when text is paired with a diagram versus a monochrome (or conventional) learning environment. Fifty-two normal-vision undergraduates at a Turkish university were randomly assigned to either a conventional group or a color-coding group. Each participant was instructed to study a diagram and accompanying text positioned below the diagram, together, they detailed the electrochemical processes in a neuronal synapse. In the color-coding group, the diagram had colored titles for the individual parts of the synaptic processes that corresponded to identically colored portions within the block of text.

Their results indicated that learners in the color-coded experimental group outperformed the learners in the conventional group in knowledge retention and transfer tests, which suggests a deep level of knowledge integration in the learning process. In addition, they showed that learner performance was not significantly different in a matching task where learners were instructed to label each item on the illustration/diagram. Furthermore, their results showed that participants in the color-coding group, on average, spent significantly less time visually searching for specific text in the diagram description and more time fixating on critical information and processing information at a deeper level. These results indicate that associations between color-coded illustrations and accompanying text can happen in a closely temporal manner, thus, resulting in better learning outcomes compared to monochrome diagrams and text.

The uses of learner signaling can vary and largely depend on the effect that the developer is trying to achieve (de Koning et al., 2009) and the context with which they are working. Conveying relationships between two or more modalities (i.e. text, sound, illustrations, etc.) often require the instructional developer to add elements to the learning material. However, by developing layout changes into learning material using the visual perception models as a guide, it stands to reason that signaling effects can be achieved to increase learning performance outcomes.

Hypotheses

Determining the effectiveness of all design methods in online learning resources is an important but large goal. However, determining the relationship between specific design methods and learning principles would be a step in providing supporting evidence for the use of specific design methods into a larger framework. The blurred line between the purposeful use of empty space (i.e. anti-horror vacui) and proximity effects as design methods and the signaling principle of multimedia learning appear to be congruent in their effectiveness within their respective domains. It is still an open question if design methods will impact a multimedia learning environment. As past research demonstrates, signaling effects can have a positive impact on performance outcomes in multimedia learning environments. While the research shows integrated text within illustrations, multi-modal material presentation, and color-coding assist learning by reducing visual search – isolated text within a larger block of the text (see image in Appendix C) remains untested as to whether it can produce signaling effects and subsequently improve learner outcomes.

Hypothesis 1. Because signaling techniques have been shown to be an effective method for providing learners a cue to important information that is relevant to accompanying illustrations (Kalyuga, Chandler, & Sweller, 2011; Mautone & Mayer, 2001; Ozcelik et al., 2009), it was predicted that learners who are provided a signaling method in a learning task would outperform learners who are not signaled in their learning task. This prediction aims to replicate past findings that support the use of signaling as an effective method for increasing learning performance (i.e. highlighted signaling > no signaling). Specifically, participants who are provided signaling cues (i.e. highlighting) in their learning task would have higher retention of the learning material and be able to provide more solutions to transfer problems than non-signaled learners (Craig, Twyford, Irigoyen, & Zipp, 2015; de Koning, Tabbers, Rikers, & Paas, 2010; van Gog, 2014).

Hypothesis 2. Past research suggests that by inducing grouping effects described by gestalt and feature integration theory through the deliberate use of empty space in a design, an effective means of signaling can be achieved. Because the human visual perception system recognizes scenes in a hierarchy with elemental group features being in the pre-attentive stage of perception, it was hypothesized that isolating text which corresponds to a particular illustration can be an effective signaling method to increase learning performance compared to the performance of the group which is provided continuous monochromatic text (i.e. isolated text signaling > no signaling). It is expected that the isolated text group will outperform participants in the continuous text group in multiple choice, retentions and transfer measures.

Hypothesis 3. Because there is little to no evidence to suggest that isolated text within a larger body of text is an effective signaling method, it is unclear if it is more effective than signaling learners via highlighting. If text isolation is effective, the learner performance outcomes will be similar to the performance seen in the highlighting group. In other words, multiple choice, retention, and transfer scores for the isolated text group should outperform the control group (hypothesis 2), but should not statistically differ from the highlighting group (i.e. highlighting = text isolation > no signaling).

Hypothesis 4. In asking participants to perform their tasks in each group, there may be an impact on cognitive workload. Because it has been suggested that visual search times are reduced (Ozcelik et al., 2009) when signaling methods are designed into the learning material, it is hypothesized that participants in the highlighting and isolated text groups will report reduced task difficulty compared to the continuous text group.

Hypothesis 5. Because the task difficulty may differ between groups (hypothesis 4), it is necessary to determine whether the task difficulty is due to the user interface or signaling manipulations. Participants will be using the same interface to interact with the learning material, but the text will be different in each group. Thus, it is hypothesized that there will be no significant differences in the user interface evaluations between groups.

CHAPTER 2

METHOD

Participants

To estimate the number of participants required to find significant results a power analysis was conducted using the reported effect size in de Koning et al., (2010). The power analysis was conducted using G*Power with a moderate effect size of $\eta^2 = .123$ and a power of 0.95. The results indicated the total $n = 114$ participants are required to find an effect with $\alpha < 0.05$. In consideration of the likelihood that there would be unusable data, the sample size was increased to $n = 150$.

The participants were 150 individuals recruited from Amazon.com's Mechanical Turk (MTurk). Each participant received $1.00 US in compensation for his or her participation in the study. Twelve participants were excluded from analyses because of blatant plagiarism identified through a simple internet search using whole sections of provided answers. There were 46 participants (18 males) in the continuous text group with an average age of $M = 40.8$, 50 participants (22 males) in the highlighted group with an average age of $M = 41.2$ (one participant chose not to comment on age or gender), and 42 participants (22 males) in the isolated text group with an average age of $M = 40.26$. A Kruskal-Wallis H test was performed on participants demographics to determine if there were any differences in gender, education, and employment status between groups – no significant differences were found ($\chi^2(2) = 1.614$, $p = 0.446$; $\chi^2(2) = 1.012$, $p = 0.603$; $\chi^2(2) = .095$, $p = 0.954$, respectively).

Sampling rationale. As with traditional convenience samples (university subject pools, community volunteers, etc.), MTurk has a fair amount of criticism in regards to being used as a means for data collection. Studies have shown that there are both advantages and disadvantages to using MTurk's labor market for various kinds of participatory research (for more information see: Goodman, Cryder, & Cheema, 2012; Landers & Behrend, 2015; Paolacci, Chandler, & Ipeirotis, 2010; Peer, Vosgerau, & Acquisti, 2014), but similar criticisms can be made for university students as research participants and not all of the issues apply to this particular study. However, one major criticism that was addressed was the issue of reduced attentional effort from participants. A recent study has demonstrated a process for selecting reliable participants by setting participant inclusion criteria with MTurk workers that have a 95% or above Human Intelligence Task (HIT) approval rating; for further information see Paolacci and Chandler (2014). Another important criticism that was addressed is the issue of participant cheating by using the internet to search for answers. A recent study has suggested to simply ask participants to not look up answers as a sufficient method to reduce dishonesty (Goodman et al., 2012).

Sampling procedures. Inclusion criteria for participant selection consisted of requirements afforded by MTurk. The first criterion was that all participants had to be based in the United States to reduce the likelihood that English was not the participant's primary language. By doing this, issues that may arise as a result of a language barrier can be mitigated. The second criterion was that participants must have a 95% or above HIT approval rating (Paolacci & Chandler, 2014).

Experimental Design

The experiment was a between-subjects design with three conditions: continuous text (CT), highlighted text (HT), and isolated text (IT). Each group was provided the same base text detailing the process of lightning formation, but the information signaling methods were manipulated.

The CT group was shown text that had no line breaks in the base-text that accompanied the 16 illustrations. Because this condition provided no signaling elements, the CT group served as a control to establish baseline learning performance of the learning task. The HT group received the same display as the control except for the featured portion of the base-text was highlighted with a light gray color that corresponded to the relevant illustration. The IT group learning task provided participants the same base text as the other two groups, but as the learning task progressed, the featured sections of the base-text were isolated within the larger block of text with empty space. In other words, when a section of text corresponded to the illustration, there was empty space both above and below that text. The text that was isolated shifted down the learning material as the learner progressed through the task with the corresponding illustration changes.

Materials. All materials and procedures have been adapted from Harp and Mayer (1998), Moreno and Mayer (1999), and Craig, Gholson, and Driscoll (2002).

Questionnaire. The questionnaire contained standard questions (see Appendix A) intended to gather demographic information like age, gender, level of education, etc. and a seven-item knowledge assessment and five item self-rating (see Appendix B) used to provide information about the participants' general knowledge of meteorology.

Learning material. The base text used in all conditions were used from Harp and Mayer (1998). The base-text contained a 521-word passage that explained how lighting is formed (see Appendix D). The text was created using factual information found in high school level textbooks and encyclopedias and was calculated to have a Flesch-Kincaid grade-level score of 9.40, which is equivalent to a freshman or sophomore in high school.

The illustrations (see Appendix E) were created based on the 16 text/narration cycles found in Moreno and Mayer (1999). Illustrations were presented as a slideshow with the ability to play as a slideshow, pause, go back, or skip forward. After 6.5 minutes, the participants moved on to the next phase of the experiment.

Knowledge assessments. The retention test required participants to answer one question/prompt – “Please explain how lightning works in the text box below." by entering their answer into the supplied text box. Participants were provided a timer that counted down from 6 minutes, before being forced to move on to the next test.

The transfer test required participants to provide as many solutions as they could to four separate open-ended questions: "What could you do to decrease the intensity of lightning?" "Suppose you see clouds in the sky, but no lightning. Why not?" "What does air temperature have to do with lightning?" "What causes lightning?". Participants were presented each question serially and given a 3-minute timer for each question before moving on to the next question.

The administered post-learning knowledge assessments were adapted from Craig et al. (2002), which asked six multiple-choice questions that vary in two levels: “What causes a flash of lightning?” and “When do downdrafts occur?” which are deep explicit questions, “The upper portion of the cloud is made up of what?” and “What part of the cloud are the positively charged particles located in?” which are shallow explicit questions, and “Why does lightning strike buildings and trees?” and “Why does it get colder right before it rains?” which are deep implicit questions.

Usability assessments. Two self-paced cognitive measures were obtained; the first was the six-item NASA – Task Load Index (TLX) which is a subjective measure of perceived cognitive workload on a 1-20 scale (see Appendix E). Along with the six subscale ratings, 15 pair-wise weights were collected for the NASATLX analysis. The second cognitive measure was the ten item System Usability Scale (SUS), that uses a 1 (strongly disagree) to 5 (strongly agree) scale to assess the usability of the user interface (see Appendix F and G).

Experimental Procedures

All participants volunteered for this study by accepting the online HIT, which was posted on MTurk. Participants first accepted the informed consent procedures before moving on. First, all participants completed the self-paced questionnaire. Second, participants were randomly assigned to one of three conditions using the random assignment feature provided by Qualtrics. Then, they were presented with an instructional screen (see Appendix C), which informed the participants that they would have 6.5 minutes to learn from illustrations with accompanying text, which would explain to them, the process of lightning formation followed by a series of tests. Participants confirmed they understood the proceeding tasks and the following screen presented their respective illustrations. Third, after participants finished with their learning materials, an instructional screen gave directions to the participants completed the retention test and the four transfer questions. Fourth, participants for the multiple choice post-test. Fifth, participants completed the SUS and NASA TLX. Finally, participants were debriefed and were provided information regarding the purpose of this study

Scoring procedures. Participant answers from the retention and transfer tests were independently scored without knowledge of participant conditions. Retention scores were established for each participant based on counting major idea units out of a possible 19 item scoring list that was from the presented illustrations and text. A single point was given for each major idea unit the participant wrote in their answer regardless of the specific words they provided. The transfer test scores were evaluated based on acceptable answers and then summed for a total transfer score; for more information on the definition of an acceptable or unacceptable answer see Moreno and Mayer (1999).

Retention and transfer scoring procedures were evaluated using Cohen's Kappa interrater reliability tests. Two scorers independently scored a random sample of 30 participants (i.e. 20% of *N)* from all groups. Retention scoring had an interrater reliability score of $\kappa = .67$ which is interpreted as a substantial amount (.61-.80) of reliability (Cohen, 1960; McHugh, 2012). Interrater reliability for transfer scores was higher at $\kappa = .75$, which is still considered substantial.

CHAPTER 3

RESULTS

Preliminary Analyses

A preliminary analysis was performed to identify if there were any differences between groups regarding participant pre-knowledge scores (i.e. the sum of the seven-item checklist and five-item self-report). A one-way analysis of variance (ANOVA) with an alpha level of .05 showed there was a significant difference in prior meteorology knowledge before exposure to the learning material ($F(2,135) = 5.06$, $p = .008$, $\eta^2_p = .07$) (see Table 1. for means and standard deviations). Post-hoc analyses using Tukey's HSD showed there was no significant difference in pre-knowledge scores between the CT and HT groups ($p = .766$), but there was a marginally significant difference between CT and IT groups ($p = .053$) and a significant difference between the HT and IT groups ($p = .007$). However, one-tailed Pearson's correlation tests showed there were no significant correlations between meteorology pre-knowledge and any of the learning assessments (see Table 1).

Table 1

Pre-Knowledge Correlations to Learning Assessments

		Retention		Transfer		Multiple Choice	
Condition	*n*	*r*	*sig.*	*r*	*sig.*	*r*	*sig.*
Continuous	46	-0.011	0.472	0.06	0.345	0.001	0.497
Highlighted	50	-0.106	0.231	-0.048	0.37	-0.008	0.479
Isolated	42	0.085	0.297	0.184	0.122	-0.05	0.376

Further analyses of the knowledge assessment data showed significant Shapiro-Wilk tests indicating these data were positively skewed in most cases. While linear based models (i.e. ANOVA, ANCOVA, etc.) are robust enough to accommodate positively skewed data, any directional tests following the initial analysis are severely impacted in reliability for both significance and power (Glass, Peckham, & Sanders, 1972). Data transformations (e.g. Log_2 transform) are suggested to be an effective method for dealing with non-normally distributed data (Feng et al., 2014). However, in this study, unsuccessful data transformations suggested these data would violate the assumptions for an ANCOVA. Therefore, subsequent non-parametric tests were used to analyze the knowledge assessments.

Table 2

Statistics for Participants' Scores of Retention, Transfer, and Multiple Choice Questions

		Pre-Knowledge	Retention			Transfer
Condition	*n*	*M (SD)*	*M (SD)*	*Mdn*	*Mean Rank*	*M (SD)*
Continuous	46	5.48 (2.65)	4.96 (3.17)	5	60.48	1.96 (1.25)
Highlighted	50	5.1 (2.7)	5.56 (3.33)	5.5	67.42	2.4 (1.41)
Isolated	42	6.81 (2.63)**	6.9 (3.75)*	7	81.86	2.48 (1.45)

Note. CT = continuous text; HT = highlighted text; IT = isolated text; * = p<.05, ** = p<.01

Retention

Analyses of retention scores for each condition indicated that the score distribution was not normally distributed indicated by a significant Shapiro-Wilk result (W = 0.947, p = .035) for the CT group and log transformations were unsuccessful in normalizing all retention score data. Because of this, a Kruskal-Wallis H test was conducted to determine if there were any significant differences in retention scores between the signaling conditions. This test showed a significant difference between groups, ($\chi^2(2) = 6.543$, $p = 0.038$), with a mean rank retention score of 60.48 for the CT group, 67.42 for the HT group, and 81.86 for the IT group (see Table 1).

Follow-up Mann-Whitney U tests were conducted to identify the pairwise differences of knowledge retention between conditions using the Bonferroni approach (i.e. post hoc $\alpha = 0.017$) to control for Type I error. The results of these tests indicated that the IT group (*Mdn* = 5) had signigicantly higher retention scores than the CT group (*Mdn* = 7), $U = 671.0$, $p = .013$, $r = .26$. Pairwise comparisons for CT vs. HT ($U = 1030.0$, $p = .38$, $r = .09$) and HT vs. IT ($U = 826.0$, $p = .078$, $r = .18$) were not significant.

An analysis of verbosity was conducted to identify if there were any significant differences between groups that could account for the differences observed in the retention scores. Past research, which uses Mayer's lightning formation material, has not focused on the amount of text generated by different learning-material manipulations. For this particular study, text generation was analyzed because manipulations which are suggested to impact a learner's ability to recall material may be expressed in the amount of learning material recall, rather than the categorical scoring of "idea units". Because of this, a sum of characters was calculated for each participants' retention assessment answer and used for further analysis. Tests of normality and homogeneity showed the character analysis scores distributions were not normal indicated by significant Shapiro-Wilk results (CT: $W = 0.878$, $p < .000$; HT: $W = 0.885$, $p < .000$; IT: $W = 0.957$, $p < .000$), but the Levene's tests were not significant. Because of this, a Kruskal-Wallis H test was conducted to determine if there were any significant differences of text generation between the signaling conditions. This test showed a significant difference between groups, ($\chi^2(2) = 8.309$, $p = 0.016$), with a mean rank retention score of 66.59 for the CT group, 60.19 for the HT group, and 83.77 for the IT group.

Follow-up Mann-Whitney U tests were conducted to identify the pairwise differences of text gerneration between conditions using the Bonferroni approach (i.e. post hoc $\alpha = 0.017$) to control for Type I error. The results of these tests indicated that the IT group (*Mdn* = 652) produced signigicantly more text than the HT group (*Mdn* = 473), $U = 699.0$, $p = .006$, $r = .29$. Pairwise comparisons for CT vs. HT ($U = 1035.5$, $p = .401$, $r = .09$) and CT vs. IT ($U = 717.5$, $p = .038$, $r = .22$) were not significant.

Table 3

Descriptive Statistics of Retention Assessment Character Counts

Condition	*n*	*M (SD)*	*Median*	*Min.,Max*	*Mean rank*
Continuous	46	532.96 (258.56)	481.5	111,1576	66.59
Highlighted	50	501.96 (272.87)	473	121,1623	60.19
Isolated	42	642.79 (268.41)*	652	188,1405	83.77

Note. * = p<.05

Transfer

Analyses of transfer scores for each condition indicated that the score distribution was not normal for all conditions indicated by significant Shapiro-Wilk results (CT: $W = 0.893$, $p = .001$; HT: $W = 0.942$, $p = .016$; IT: $W = 0.933$, $p = .016$). Because log transformations were unsuccessful in normalizing all transfer score data, a Kruskal-Wallis H test was conducted to determine if there were any significant differences in transfer scores between the signaling conditions. This test showed no significant difference between groups ($\chi^2(2) = 4.253$, $p = 0.119$).

Multiple Choice

Analyses of the multiple choice scores for each condition indicated that the score distributions were not normal for all conditions indicated by significant Shapiro-Wilk test results (CT: $W = 0.901$, $p = .001$; HT: $W = 0.932$, $p = .006$; IT: $W = 0.936$, $p = .021$) and log transformations were unsuccessful in normalizing the data, while Levene's tests showed the heterogeneity of group distributions. Analyses of the multiple choice data were performed using two strategies. The first method used was congruent with past analyses performed by Craig et al., (2002) using the sum of correct answers and showed no significant results ($\chi^2(2) = 2.524$, $p = 0.28$). For a more granular analysis, the three different question types – explicit shallow, explicit deep, and implicit deep were analyzed separately between groups. A Kruskal-Wallis H test was conducted on each question type to determine if there were any significant differences in multiple choice scores between conditions. This test showed no significant difference between groups for the explicit shallow ($\chi^2(2) = 5.315$, $p = 0.07$), explicit deep ($\chi^2(2) = 0.876$, $p = 0.645$), and implicit deep ($\chi^2(2) = 1.698$, $p = 0.428$) questions.

Table 4

Descriptive Statistics for Multiple Choice Questions

		Explicit Shallow		Explicit Deep		Implicit Deep	
Condition	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Continuous	46	0.64	0.39	0.37	0.36	0.58	0.32
Highlighted	50	0.64	0.4	0.43	0.32	0.49	0.34
Isolated	42	0.81	0.29	0.43	0.42	0.56	0.37
p-value		0.07		0.645		0.428	

NASA TLX

In order to determine if there was a significant difference between groups in the amount of cognitive workload, a one-way ANOVA was performed. Shapiro-Wilk and Levene's tests were non-significant indicating the assumptions of normality and homogeneity were met. A one-way ANOVA was performed at an alpha level of .05, which showed there were no significant differences of cognitive workload between groups ($F(2,135) = 1.48$, $p = .231$, $\eta^2_p = .02$).

Table 5

Means and Standard Deviations for Usability Measures

		NASA TLX		SUS		
Condition	*n*	*M*	*SD*	*M*	*SD*	*Median*
Continuous	46	11.68	2.78	66.14	24.31	70
Highlighted	50	12.49	2.37	72.45	20.91	77.5
Isolated	42	12.45	2.55	70.06	20.25	70

SUS

Analyses of SUS scores for each condition indicated that the score distribution was not normal for all conditions indicated by significant Shapiro-Wilk results (CT: $W = 0.939$, $p = .019$; HT: $W = 0.927$, $p = .004$; IT: $W = 0.94$, $p = .028$). Because log transformations were unsuccessful in normalizing all SUS score data, a Kruskal-Wallis H test was conducted to determine if there were any significant differences in SUS scores between the signaling conditions. This test showed no significant difference between groups ($\chi^2(2) = 1.356$, $p = 0.508$).

CHAPTER 4

DISCUSSION

This current study aimed at identifying the viability of implementing empty space within a learning material layout as the primary means to signal learners of temporally important information. It was hypothesized that participants who were provided text signaling (i.e. highlighting and text isolation) in their learning material would show increased learning performance compared to the group with was provided traditional continuous text. It was also hypothesized that the signaled groups would report a reduced cognitive workload compared to the control group. Because signaling learners to important information has been shown to reduce visual search time (Ozcelik et al., 2009), it indirectly suggests there would a decrease in extraneous processing by mitigating the need to hold information in working memory during visual search in the learning task.

The results of this study provide mild evidence to suggest the viability of using text isolation as an effective signaling method. There was partial support for the first prediction suggested by differences in the retention scores between groups, but not with transfer and multiple choice knowledge assessments. There was no significant evidence in support of the cognitive workload prediction, which would have suggested whether learner signaling decreased cognitive workload compared to non-signaled learners.

Learning Measures

Retention. In evaluating the replication hypothesis, analyses of retention scores between the continuous text (CT) and the highlighted text (HT) groups showed there were no statistically significant differences in knowledge retention. While the results of the analyses were not significant, the mean differences indicated a trend towards learners having higher amounts of knowledge retention for the highlighted text group. One possible explanation for this result is that highlighting, as a dynamic signaling method, may not have been the optimal signaling method for this particular study (van Gog, 2014). Boucheix & Lowe (2010) conducted a study teaching participants how internal mechanisms in a piano make a sound. Their results showed that two separate, but similar, dynamic signaling methods compared to a non-signaling group, only one group showed significant improvement in learner mental models. However, their eye-tracking analyses suggested that both methods were effective at signaling to learners based on similar fixation numbers. This suggests that in this current study, highlighting may have been effective at signaling learners to temporally important information, but was not the optimal method for forming a more accurate mental model of how lighting is formed. It is plausible that using color-coding signaling with learning material that consists of images and text, similar to the method used in Kalyuga et al., 2011, would have been a more effective means for signaling learners in this study.

In consideration of the perceptual research regarding stimuli that targets pre-attentive processing through proximity grouping, text isolation was expected to provide learners with an effective signaling method during the learning task (i.e. H_2). The retention score analyses support the hypothesis that text isolation is an effective method for signaling learners. Participants in the IT group were observed to have higher knowledge retention compared to the control group (CT). Furthermore, the isolated text group (IT) participants produced, on average, nearly half of a standard deviation more text in their retention assessment answers. These results taken together suggest participants in the IT group were better able to recall and express their learned material in the same amount of time as the control group which could be interpreted as better mental model formation. However, it is important to note there was a significantly higher amount of participants in the IT group with high levels of pre-knowledge about meteorology according to the metrics used by Moreno & Mayer (1999). Attempts at statistically mitigating these effects through an analysis of covariance were not possible without violating the assumptions of the statistical test. It is important to note that the non-significant correlations between the pre-knowledge measures and the learning assessments may suggest the pre-knowledge metrics themselves are flawed. More research is needed to assess whether the pre-knowledge assessments indicate any prior meteorology knowledge that could interfere a clear interpretation of the learning assessment analyses using Mayer's lighting formation material. Also, further research is required to evaluate whether the results of these tests are accurate and representative of the effectiveness of text isolation or if it is a coincidence produced by the random assignment process.

The final cognitive learning research question aimed at identifying how effective text isolation is at signaling learners compared to an established signaling design method like highlighting (i.e. H_3). While the retention post hoc analyses showed there was no significant difference between the HT and IT groups, the group means suggest a trend towards showing text isolation as being more effective. Non-significant results in the replication analyses complicate the evaluation to assess whether text isolation is better, worse, or equal compared to an established and effective signaling method. Further research is needed that a) replicates an established signaling method and b) can mitigate any confounding variables when comparing the novel isolated text-signaling method.

Transfer. While the results of the transfer score analyses for all three cognitive learning hypothesis showed no significant differences, these results were not unexpected and appear to be congruent with the Cognitive Theory of Multimedia Learning. Incorporating a signaling method into the instructional design is particularly beneficial when extraneous material is present and cannot be removed. Thus, the signaling design method assists the learner in the selection of important words and images processed in working memory. Transfer of knowledge happens at a later stage of processing learning-material, specifically in the integration or updating of new and prior knowledge structures (van Gog, 2014). However, it is important to note the importance of downstream effects that proper signaling methods can have on learning. Without proper selection of important words and images during a learning task, updating existing knowledge structures would not be possible or worse, the formation of incorrect knowledge structures could result.

Multiple-choice. The results from the multiple choice questions suggest there were no differences between all three conditions of group scores. Similar to the transfer analyses, these results were also not entirely unexpected. The three types of questions were designed as being increasing in difficulty to assess the depth of learning (Craig et al., 2002), where explicit shallow corresponds to recall and explicit and implicit deep correspond the depth of knowledge transfer. If the transfer of knowledge does not occur, correct answers to explicit deep and implicit deep questions are left to chance. However, the explicit shallow questions were expected to show significant differences, but analyses across all conditions showed a p-value of .07. While the analyses for the explicit shallow questions were non-significant, the means and standard deviations trends suggests nearly identical results in knowledge retention between the CT and HT group, but the IT group showed improvement in their recall ability which would account for the marginally significant p-value of .07 (see Table 3). However, it is important to note that while the knowledge transfer scores appear to be congruent with CTML, variations in the explicit and implicit deep questions were expected to remain relatively flat. More research is needed to assess if the deep-level assessment scores were affected by the signaling manipulations.

Usability Measures

NASA TLX. With cognitive resources being finite and optimal essential processing being the target for improved learning performance outcomes in multimedia learning, it was expected that signaling learners during a learning task would mitigate extraneous processing and increase essential processing. However, it is difficult or near impossible to directly measure essential processing, but the NASA TLX has been shown, in some instances, of being a suitable stand-in by measuring extraneous processing which is associated with the mental effort subscale in the NASA TLX (Kalyuga, 2011).

One possible reason that no significant differences were found is that the version of the NASA TLX that was used to evaluate the fourth hypothesis was the original version produced by NASA intended for the aviation domain. It is not uncommon for researchers to use modified versions of the NASA TLX, but they are all not validated and are usually paired with other measures to assess the desired subjective variable (Hart, 2006). One common, but alternate, version is the Raw TLX (RTLX), which is generally the same metric as the original without the weighting process. However, Hart (2006) reports in a NASA TLX meta-analysis that the RTLX can be either more, less, or equally sensitive. In this current study, the validated measure (original version) was chosen to ensure the measure of extraneous processing was captured.

An alternative explanation for the results of this study is that the NASA TLX was not sensitive enough to capture only the extraneous processing inherent in this study's learning material and a modified version that is context specific should have been used. Although, Scheiter, Gerjets, Vollmann, & Catrambone (2009) suggest that learners cannot usually distinguish between measures that capture generative, essential, and extraneous processing. Thus measuring mental workload, with any aggregative NASA TLX version may be capturing all three kinds of mental workload inherent in the learning task and not extraneous processing alone because in calculating the composite overall weighted workload results in a single score. However, deductive conclusions could be made regarding which kind of workload had a greater effect if an experiment has observed learning differences between conditions.

SUS. In this study, the System Usability Scale (SUS) was used to investigate if there were any significant differences between groups regarding the user interface of the multimedia learning material. Specifically, the SUS would provide a measure to determine if there was any increase in mental workload was inherent in the learning material provided by the NASA TLX across conditions or if the increases were due to the user interface of the learning material. As predicted, analyses of the SUS scores showed no differences between groups. Although, interpretation of the scores suggest the user interface for the learning material has a rather low usability score. Bangor, Kortum, & Miller (2008) suggest that systems which score just above 70 are minimally sufficient in their usability, whereas better systems score in the high 70 to low 80 range, while superior systems have a score higher than 90. This study observed that the CT group reported a SUS score below the minimum threshold while the HT and IT groups were acceptable. This suggests that further product development would be necessary for a basic learning material presentation (i.e. no signaling), while the signaling design methods (i.e. highlighting and text isolation) would be passable in their development into a learning product. However, the non-significant SUS analyses observed in this study's results could suggest that the SUS (and to some extent, the NASA TLX) was used in evaluating a system that was too simple. The SUS is particularly useful when evaluating large, complex, and/or diverse systems and conducted by usability professionals (Ackerman, Parush, Nassar, & Shtub, 2016; Bangor et al., 2008) largely in the product evaluation fields. It is important to note that without clear results from the NASA TLX, it cannot be entirely determined that the non-significance of SUS score comparisons captured what they were intended to capture which is the usability of a system, or in this instance, a simple, purpose-built website, used to learn about lightning formation.

Sampling Issues & Limitations

In this current study, there were some unexpected observations regarding the sample of participants that may have significantly influenced the pre and post knowledge-assessment analyses. Possibly the most influential and unexpected factor was the 62 participants that scored above a six (out of 12) on the pre-knowledge assessment. This is perhaps the most concerning issue with this particular study because there was a statistically significant higher number of participants in the IT group who scored above six. The retention score analyses showed the IT group significantly outperformed the control group and marginally outperformed the HT group, but these effects could possibly be due to a number of high pre-knowledge participants in the IT group. Further research is needed to assess if the group differences in knowledge retention are due to the high amount of meteorology pre-knowledge or if participants learned from the experiment task. This issue could be addressed by implementing a high/low pre-knowledge split of participants that use restricted random group assignment to force equal sample sizes across groups (Shadish, Cook, & Campbell, 2002).

Alternatively, all high knowledge participants could have been susceptible to the expert reversal effect. In short, the expert reversal effect is where a person with high knowledge can perform worse on a learning task designed for people with low-knowledge because an existing robust knowledge structure of a topic can interfere with the integration of repetitive or simplified information of the same topic (van Merriënboer, Kester, & Paas, 2006). This explanation could account for the uncharacteristic and widely variable data distributions that were found throughout this study's analyses. Past experiments using Mayer's lightning formation material presumably control for this by excluding participants' data that score above a six on the pre-knowledge assessment (Harp & Mayer, 1997; Mautone & Mayer, 2001; Moreno & Mayer, 1999). However, there is little to no evidence to support that having a high score on the pre-knowledge assessment has a direct relationship to the outcome of the post-treatment knowledge assessments.

Another issue that was discovered was the sample used in this study was comprised of adults older than expected. Paolacci et al. (2010) observed the average age of MTurk users to be thirty-six years old, while other studies have seen samples with a lower average age (Chiu & Egner, 2015; Craig et al., 2015; Goodman et al., 2012;

Paolacci & Chandler, 2014). In more controlled environments, like a laboratory setting, for example, participants are primarily younger and are recruited through introductory psychology subject pools. With these tightly controlled environments, confidence in internal validity can be quite high, but this confidence comes at the expense of ecological validity (Aguinis & Edwards, 2014; Landers & Behrend, 2015). In this particular study, the observed results are not as well defined as what would have potentially been observed in a tightly controlled lab setting. Thus, more research is needed with a higher degree of control to clearly identify if text isolation can induce increased learning performance compared to a control group and other signaling methods.

Conclusions

This present study attempted to determine if using text isolation as a signaling method could be a viable pedagogical design method in a web-based multimedia learning format. While the results show there were minimal differences between conditions in the majority of post-test knowledge assessments, the data suggests a promising trend towards showing viability. Future research conducted in a more tightly controlled environment is needed. This research will be able to fully determine if the signaling method is effective at what is was proposed to do – help learners learn more effectively. Specifically, implementing eye-tracking metrics into the research protocols would be particularly beneficial to the evaluation of text isolation as a signaling method. Because signaling methods can work in different ways (i.e. organizing, conveying information, guiding attention) (de Koning et al., 2009), learning metrics and human performance data, like eye-tracking, should be gathered in parallel. However, as with this study, using a representative population will also be informative as to the applicability of pedagogical design methods.

Instructional designers who are creating the next generation of learning material have a plethora of technologies to deliver their lessons through. However, not all pedagogical design principles are effective across technological platforms and research is needed to determine what methods should be used in specific applications. This study is one step of many in determining the limitations of a more broad research area that focuses on the numerous combinations of different kinds of technologies and instructional designs (van Gog, 2014). Furthermore, this present study contributes to the, oftentimes, ambiguous results that are found in studies using diverse samples that are more representative of a global population. Because of this, the future directions for isolated text signaling could assist in developing instructional design guidelines. These guidelines would be particularly useful for the development of the next generation of learning formats. Ultimately, this type of design method could be useful in providing easy-to-implement signaling that supports the development of effective e-learning resources for the hearing impaired and color-blind.

REFERENCES

Ackerman, R., Parush, A., Nassar, F., & Shtub, A. (2016). Metacognition and system usability: Incorporating metacognitive research paradigm into usability testing. *Computers in Human Behavior*, *54*, 101–113. doi: 10.1016/j.chb.2015.07.041

Aguinis, H., & Edwards, J. R. (2014). Methodological wishes for the next decade and how to make wishes come true. *Journal of Management Studies*, *51*(1), 143–174. doi: 10.1111/joms.12058

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839. doi: 10.1038/nrn1201

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, *24*(March 2015), 574–594. doi: 10.1080/10447310802205776

Boucheix, J.-M., & Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction*, *20*(2), 123–135. doi: 10.1016/j.learninstruc.2009.02.015

Butcher, K. R. (2014). The Multimedia Principle. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 174–205). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139547369.010

Chiu, Y.-C., & Egner, T. (2015). Inhibition-Induced Forgetting. *Psychological Science*, *26*(1), 27–38. doi: 10.1177/0956797614553945

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210. doi: 10.1007/BF01320076

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46. doi: 10.1177/001316446002000104

Coser, L. a., & Elias, N. (1978). *The Civilizing Process: The History of Manners. Contemporary Sociology* (Vol. 7). doi: 10.2307/2065007

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features and redundancy. *Journal of Educational Psychology*, *94*(2), 428–434. doi: 10.1037/0022-0663.94.2.428

Craig, S. D., Twyford, J., Irigoyen, N., & Zipp, S. A. (2015). A Test of Spatial Contiguity for Virtual Human's Gestures in Multimedia Learning Environments. *Journal of Educational Computing Research*, *53*(1), 3–14. doi: 10.1177/0735633115585927

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a Framework for Attention Cueing in Instructional Animations: Guidelines for

Research and Design. *Educational Psychology Review*, *21*(2), 113–140. doi: 10.1007/s10648-009-9098-7

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*, *20*(2), 111–122. doi: 10.1016/j.learninstruc.2009.02.010

EACEA, & European Commission. (2014). *Modernisation of Higher Education in Europe: Access, Retention and Employability 2014*. *Education, Audiovisual and Culture Executive Agency* (Vol. 2). Brussels. doi: 10.2797/72146

Elias, N. (1994). State Formation and Civilization. In *The Civilizing Process: The History of Manners and State Formation and Civilization* (pp. 265-334-421).

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, *26*(2), 105–9. doi: 10.3969/j.issn.1002-0829.2014.02.009

Florax, M., & Ploetzner, R. (2010). What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. *Learning and Instruction*, *20*(3), 216–224. doi: 10.1016/j.learninstruc.2009.02.021

Gardner, H. (1987). *The Mind's New Science* (1st ed.). United States: BasicBooks.

Glass, G. V, Peckham, P. D., & Sanders, J. R. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research*, *42*(3), 237–288. doi: 10.3102/00346543042003237

Goodman, J. K., Cryder, C. E., & Cheema, A. (2012). Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224. doi: 10.1002/bdm.1753

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology*, *89*(1), 92–102. doi: 10.1037/0022-0663.89.1.92

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, *90*(3), 414–434. doi: 10.1037/0022-0663.90.3.414

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Human Factors and Ergonomics Society Annual Meting*, 904–908. doi: 10.1037/e577632012-009

Hartley, J. (2004). Designing instructional and informational text. *Handbook of Research in Educational Communications and Technology*, 917–947.

Jin, S.-H. (2013). Visual design guidelines for improving learning from dynamic and interactive digital text. *Computers & Education*, *63*, 248–258. doi: 10.1016/j.compedu.2012.12.010

Kahneman, D., & Henik, A. (1977). Effects of visual grouping on immediate recall and selective attention. In S. Dornič (Ed.), *Attention and performance VI: proceedings of the Sixth International Symposium on Attention and Performance* (Vol. 6, pp. 307–332). Stockholm, Sweden: L. Erlbaum Associates.

Kalyuga, S. (2011). Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educational Psychology Review*, *23*(1), 1–19. doi: 10.1007/s10648-010-9150-7

Kalyuga, S., Chandler, P., & Sweller, J. (2011). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, *25*(SUPPL. 1). doi: 10.1002/acp.1773

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the Lawfulness of Grouping by Proximity. *Cognitive Psychology*, *35*(1), 71–98. doi: 10.1006/cogp.1997.0673

Landers, R. N., & Behrend, T. S. (2015). An Inconvenient Truth: Arbitrary Distinctions Between Organizational, Mechanical Turk, and Other Convenience Samples. *Industrial and Organizational Psychology*, *8*(2), 142–164. doi: 10.1017/iop.2015.13

Lidwell, W., Holden, K., Butler, J., & Elam, K. (2010). *Universal Principles of Design*. *Universal principles of design: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design.* Beverly, MA: Rockport Publishers.

Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, *93*(2), 377–389. doi: 10.1037/0022-0663.93.2.377

Mayer, R. E. (1999). Multimedia aids to problem-solving transfer. *International Journal of Educational Research*, *31*(7), 611–623. doi: 10.1016/S0883-0355(99)00027-0

Mayer, R. E. (2014a). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 43–71). Cambridge University Press.

Mayer, R. E. (Ed.). (2014b). *The Cambridge Handbook of Multimedia Learning* (2nd ed.). Cambridge University Press.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282. doi: 10.11613/BM.2012.031

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, *91*(2), 358–368.

doi: 10.1037/0022-0663.91.2.358

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, *92*(1), 117–125. doi: 10.1037/0022-0663.92.1.117

Mortelmans, D. (2005). Visualizing Emptiness. *Visual Anthropology*, *18*(1), 19–45. doi: 10.1080/08949460590900034

Murray, T. (2004). Educational Software Authoring Tools. *Educational Technology*, *44*(October), 10–16.

Nuland, S. E. Van, Eagleson, R., & Rogers, K. A. (2016). Educational Software Usability : Artifact or Design ? The Economic Potential of Online Learning, *0*, 1–10. doi: 10.1002/ase.1636

Oyama, T. (1961). Perceptual Grouping As a Function of Proximity. *Perceptual and Motor Skills*, *13*, 305–306. doi: 10.2466/pms.1961.13.3.305

Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Computers and Education*, *53*(2), 445–453. doi: 10.1016/j.compedu.2009.03.002

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science*, *23*(3), 184–188. doi: 10.1177/0963721414531598

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, *5*(5), 411–419. doi: 10.2139/ssrn.1626226

Park, I., & Hannafin, M. J. (1993). Empirically-based guidelines for the design of interactive multimedia. *Educational Technology Research and Development*, *41*(3), 63–85. doi: 10.1007/BF02297358

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023–1031. doi: 10.3758/s13428-013-0434-y

Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Bulletin*, *129*(5), 643–673. doi: 10.1037/0033-2909.129.5.643

Scheiter, K., Gerjets, P., Vollmann, B., & Catrambone, R. (2009). The impact of learner characteristics on information utilization strategies, cognitive load experienced, and performance in hypermedia learning. *Learning and Instruction*, *19*(5), 387–401. doi: 10.1016/j.learninstruc.2009.02.004

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and Quasi-Experimental for Generalized Designs Causal Inference. *Handbook of Industrial and Organizational Psychology*, *223*, 623. doi: 10.1198/jasa.2005.s22

Treisman, A. M. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology. Human Perception and Performance*, *8*(2), 194–214. doi: 10.1037/0096-1523.8.2.194

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. doi: 10.1016/0010-0285(80)90005-5

van Gog, T. (2014). The Signaling (or Cueing) Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 263–278). Cambridge, MA.: Cambridge University Press.

van Merriënboer, J. J. G., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology*, *20*(3), 343–352. doi: 10.1002/acp.1250

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. a., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. doi: 10.1037/a0029333

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung*, *4*(1), 301–350. doi: 10.1007/BF00410640

Wertheimer, M. (1938). Laws of organization in perceptual forms. In *A source book of Gestalt psychology.* (pp. 71–88). London: Kegan Paul, Trench, Trubner & Company. doi: 10.1037/11496-005

Wertheimer, M. (1959). *Productive Thinking - Enlarged Edition*. (M. Wertheimer, Ed.) (1st ed.). New York, New York, USA: Harper & Brothers.

Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., … Merfeld, D. M. (2012). *Sensation & Perception* (3rd ed.). Sunderland, MA: Sinauer Associates, Inc.

Zucker, S. W., & Davis, S. (1988). Points and endpoints: a size/spacing constraint for dot grouping. *Perception*, *17*(2), 229–247. doi: 10.1068/p170229

APPENDIX A

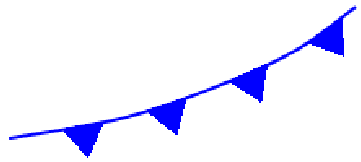
DEMOGRAPHICS QUESTIONNAIRE

1. Please enter your age in the text box
    a. If you do not wish to answer, please enter ‘999’
2. Please select your gender
    a. Male
    b. Female
    c. Other
    d. I do not wish to specify
3. Please select your highest level of education
    a. No schooling completed
    b. High School Diploma or GED
    c. Some college credit; no degree
    d. Trade/technical/vocational training
    e. Associate’s Degree
    f. Bachelor’s Degree
    g. Master’s Degree
    h. Doctorate Degree
    i. I do not wish to specify
4. Employment Status: Are you currently
    a. Employed for wages
    b. Self-employed
    c. Out of work and looking for work
    d. Out of work but not currently looking for work
    e. A homemaker
    f. A student
    g. Military
    h. Retired
    i. Unable to work
    j. I do not wish to specify

APPENDIX B

PRE-KNOWLEDGE CHECKLIST

1. Please check the boxes next to the items that apply to you
    a. I regularly read the weather maps in the newspaper or online
    b. I know what a cold front is
    c. I can distinguish between cumulous and nimbus clouds.
    d. I know what a low pressure system is.
    e. I can explain what makes the wind blow.
    f. I know what this symbol means



    g. I know what this symbol means

APPENDIX C

CONDITION INSTRUCTIONS

Continuous Text

Introduction

Shortly, you will be asked to click a link that will take you to another tab. In this new tab, you will be provided images and text that teach you about how lightning is formed.

You will have 6.5 minutes to learn how lightning is formed, but you will be able to move on after 5 minutes has elapsed. Please take as much time as you need up to the 6.5minute allotment to learn as much as you can.

After the 6.5 minutes have passed, you will be asked to return to the survey tab in your browser to answer a series of questions about how lightning is formed.

Details

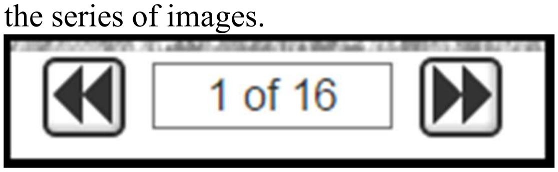
To the left of the screen, you will see images that show how lightning is formed.

To the right of the screen, you will see text that details how lightning is formed.

All images correspond to the text that is provided.

Navigation Instructions

Below the images, there are Back and Forward buttons that allow you to scroll through the series of images.



Highlighted Text

Introduction

Shortly, you will be asked to click a link that will take you to another tab. In this new tab, you will be provided images and text that teach you about how lightning is formed.

You will have 6.5 minutes to learn how lightning is formed, but you will be able to move on after 5 minutes has elapsed. Please take as much time as you need up to the 6.5minute allotment to learn as much as you can.

After the 6.5 minutes have passed, you will be asked to return to the survey tab in your browser to answer a series of questions about how lightning is formed.

Details

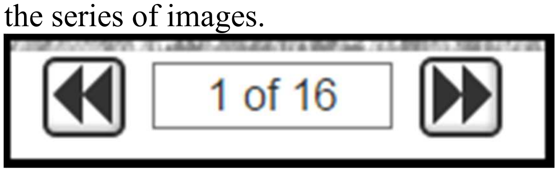
To the left of the screen, you will see images that show how lightning is formed.

To the right of the screen, you will see text that details how lightning is formed.

All images correspond to the highlighted text that is provided.

Navigation Instructions

Below the images, there are Back and Forward buttons that allow you to scroll through the series of images.



Isolated Text

Introduction

Shortly, you will be asked to click a link that will take you to another tab. In this new tab, you will be provided images and text that teach you about how lightning is formed.

You will have 6.5 minutes to learn how lightning is formed, but you will be able to move on after 5 minutes has elapsed. Please take as much time as you need up to the 6.5minute allotment to learn as much as you can.

After the 6.5 minutes have passed, you will be asked to return to the survey tab in your browser to answer a series of questions about how lightning is formed.

Details

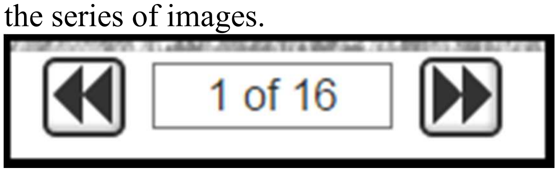
To the left of the screen, you will see images that show how lightning is formed.

To the right of the screen, you will see text that details how lightning is formed.

All images correspond to the isolated text that is provided.

Navigation Instructions

Below the images, there are Back and Forward buttons that allow you to scroll through the series of images.



APPENDIX D

LIGHTNING FORMATION BASE TEXT

The Process of Lightning

Lightning can be defined as the discharge of electricity resulting from the difference in electrical charges between the cloud and the ground.

When the surface of the earth is warm, moist air near the earth's surface becomes heated and rises rapidly, producing an updraft. As the air in these updrafts cools, water vapor condenses into water droplets and forms a cloud. The cloud's top extends above the freezing level. At this altitude, the air temperature is well below freezing, so the upper portion of the cloud is composed of tiny ice crystals.

Eventually, the water droplets and ice crystals in the cloud become too large to be suspended by updrafts. As raindrops and ice crystals fall through the cloud, they drag some of the air from in the cloud downward, producing downdrafts. The rising and falling air currents within the cloud may cause hailstones to form. When downdrafts strike the ground, they
spread out in all directions, producing gusts of cool wind people feel just before the start of the rain.

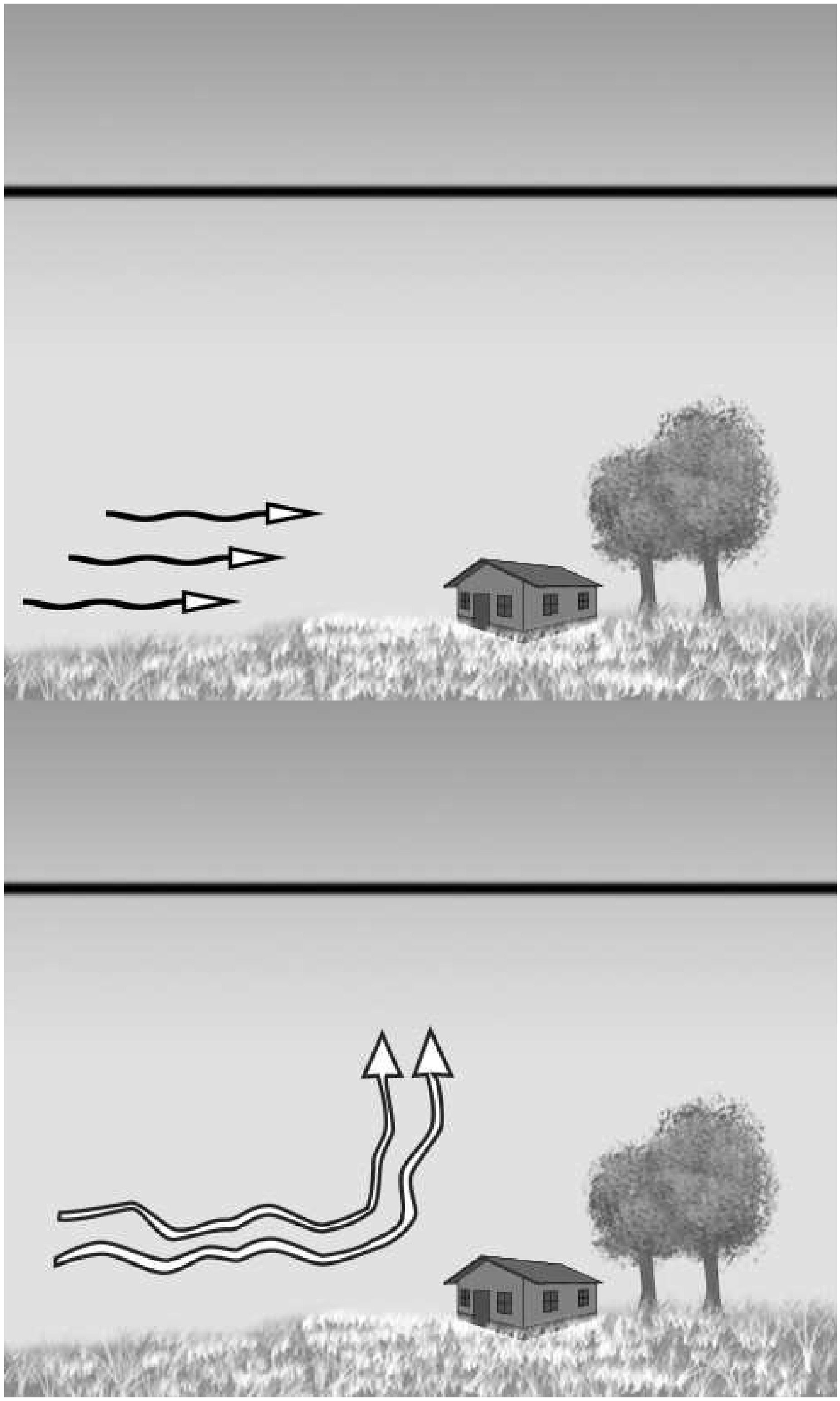
Within the cloud, the moving air causes electrical charges to build, although scientists do not fully understand how it occurs. Most believe that the charge results from the collision of the cloud's light, rising water droplets and tiny pieces of ice against hail and other heavier, falling particles. The negatively charged particles fall to the bottom of the cloud, and most of the positively charged particles rise to the top.

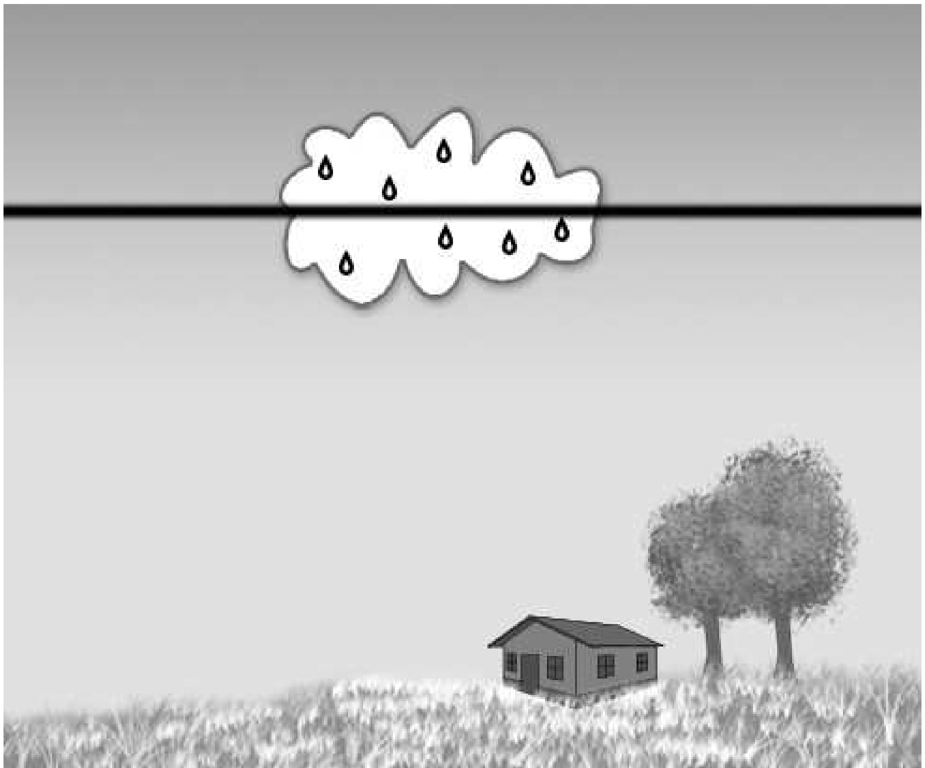
The first stroke of a cloud-to-ground lightning flash is started by a stepped leader. Many scientists believe that it is triggered by a spark between the areas of positive and negative charges within the cloud. A stepped leader moves downward in a series of steps, each of which is about 50 yards long, and lasts for about 1 millionth of a second. It pauses between steps for about 50 millionths of a second. As the stepped leader nears the ground, positively charged upward-moving leaders travel up from such objects as trees and buildings, to meet the negative charges. Usually, the upward moving leader from the tallest object is the first to meet the stepped leader and complete a path between the cloud and earth. The two leaders meet generally about 165 feet above the ground. Negatively charged particles then rush from the cloud to the ground along the path created by the leaders. It is not very bright and usually has many branches.

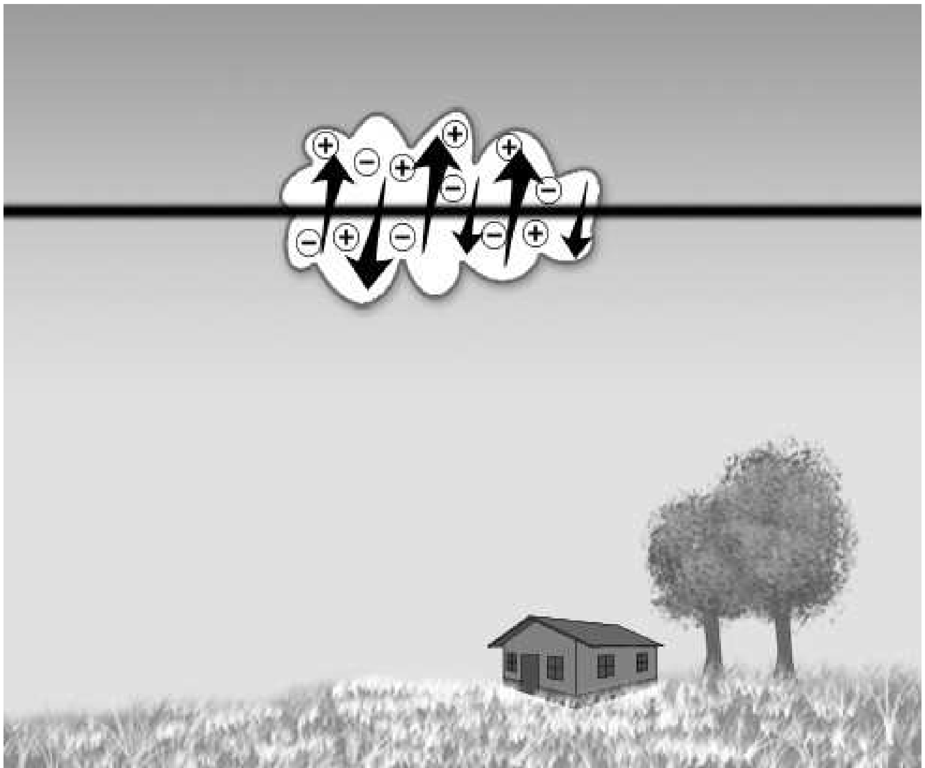
As the stepped leader nears the ground, it induces an opposite charge, so positively charged particles from the ground rush upward along the same path. This upward motion of the current is the return stroke and it reaches the cloud in about 70 microseconds. The return stroke produces the bright light that people notice in a flash of lightning, but the current moves so quickly that its upward motion cannot be perceived. The lightning flash usually consists of an electrical potential of hundreds of millions of volts. The air along the lightning channel is heated briefly to a very high temperature. Such intense heating causes the air to expand explosively, producing a sound wave we call thunder.

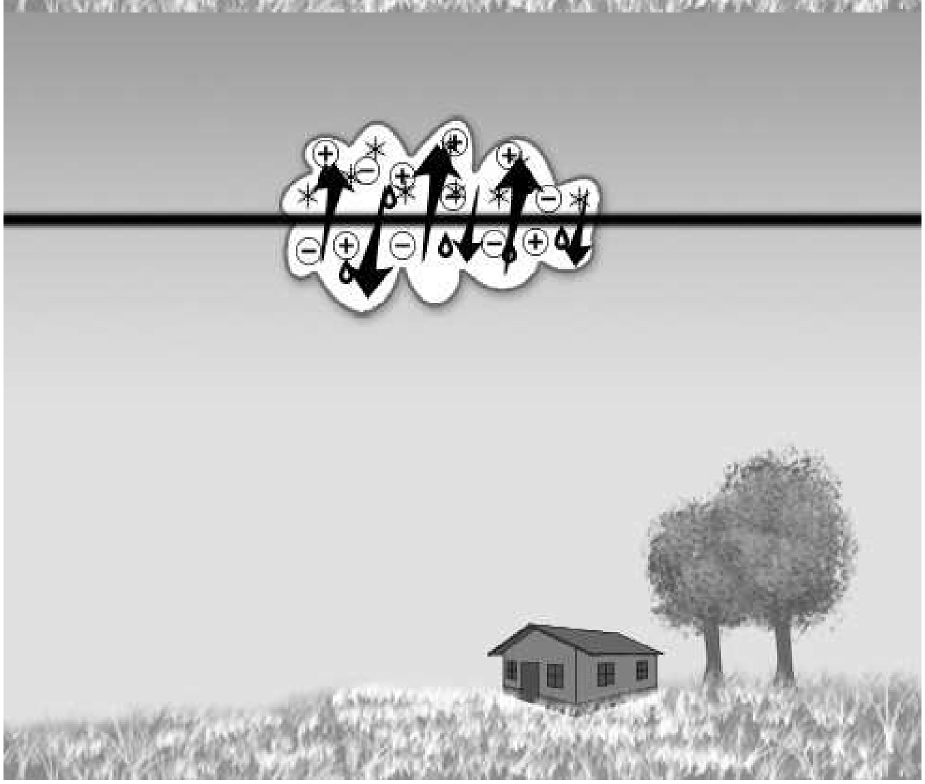
APPENDIX E

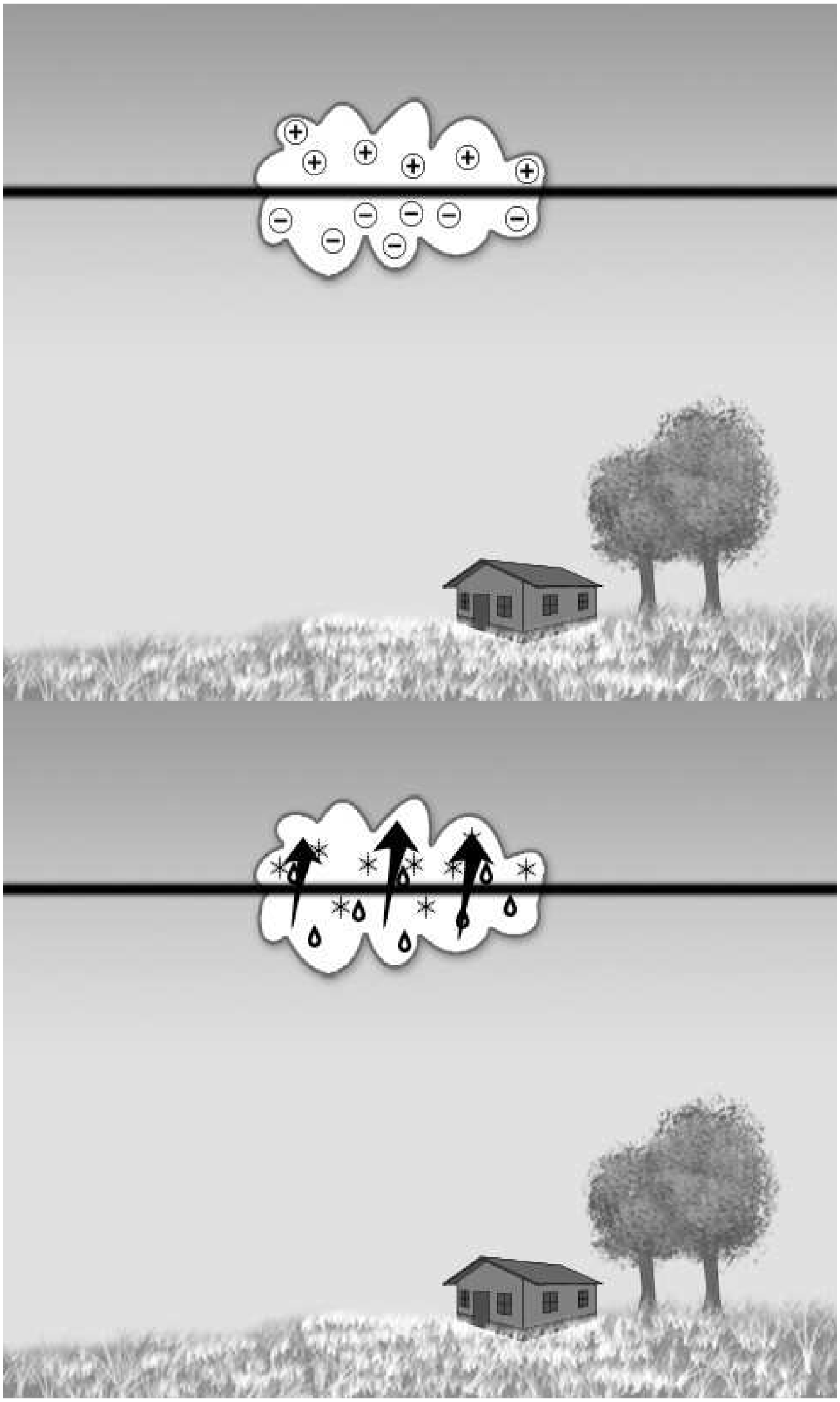
LIGHTNING FORMATION IMAGES

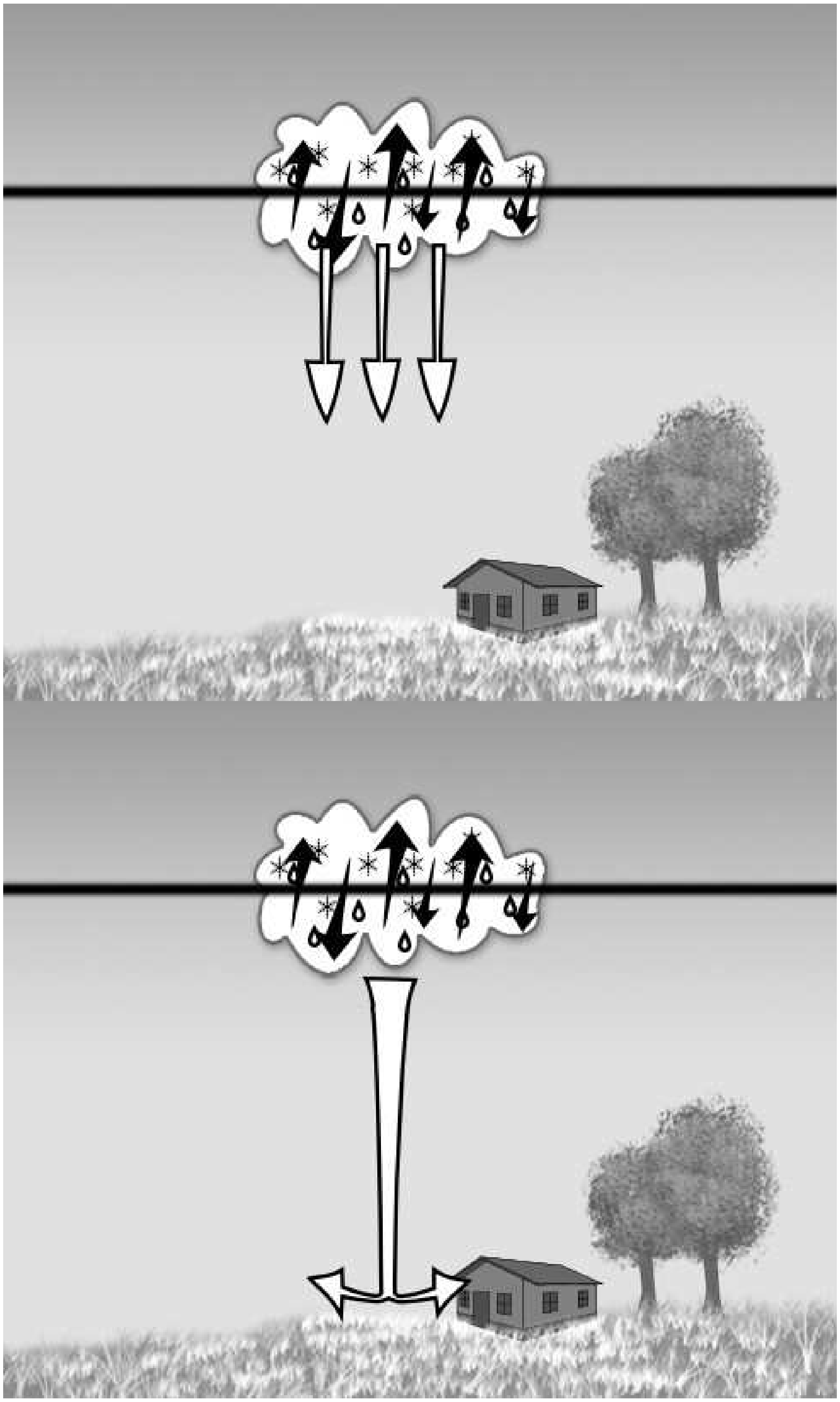


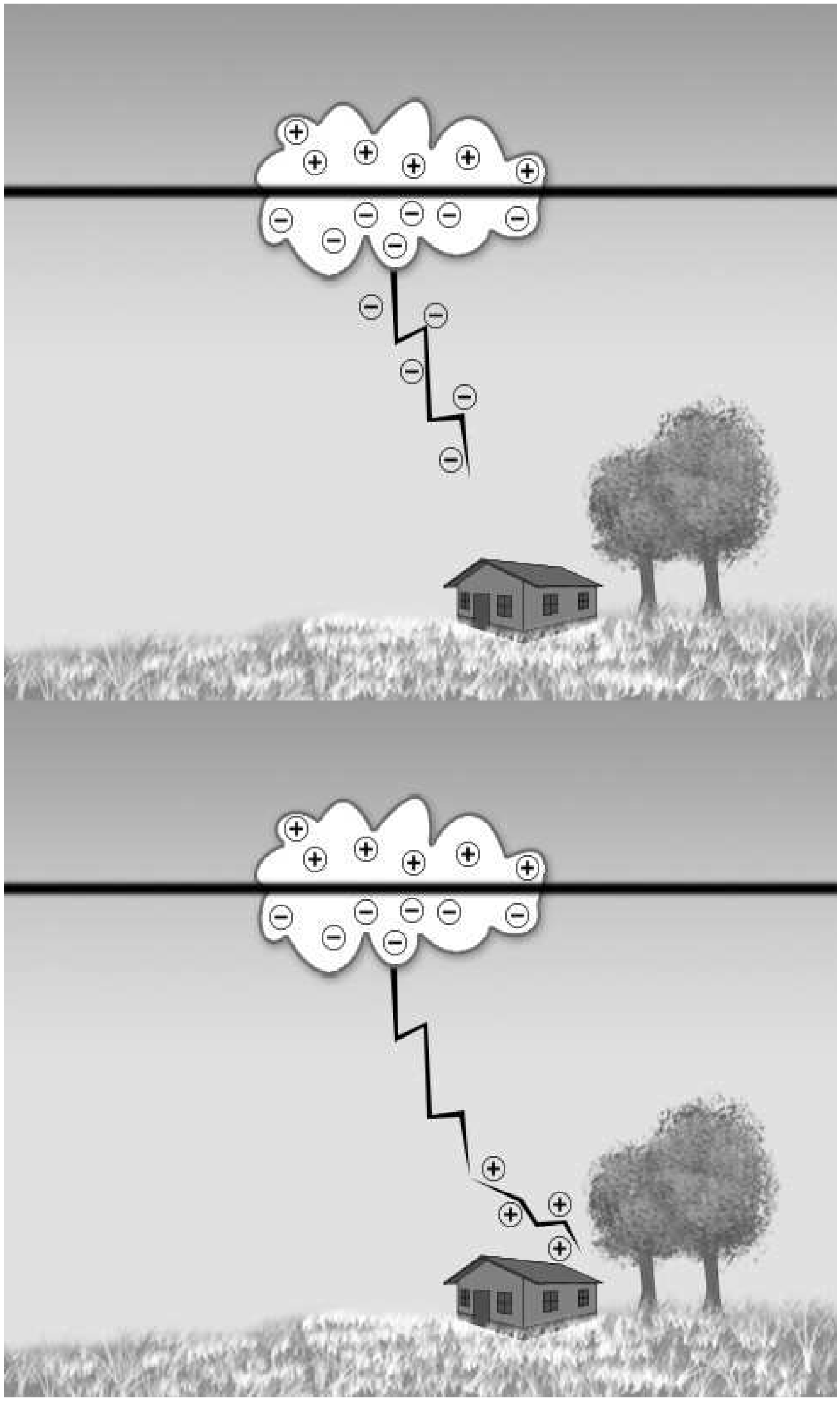


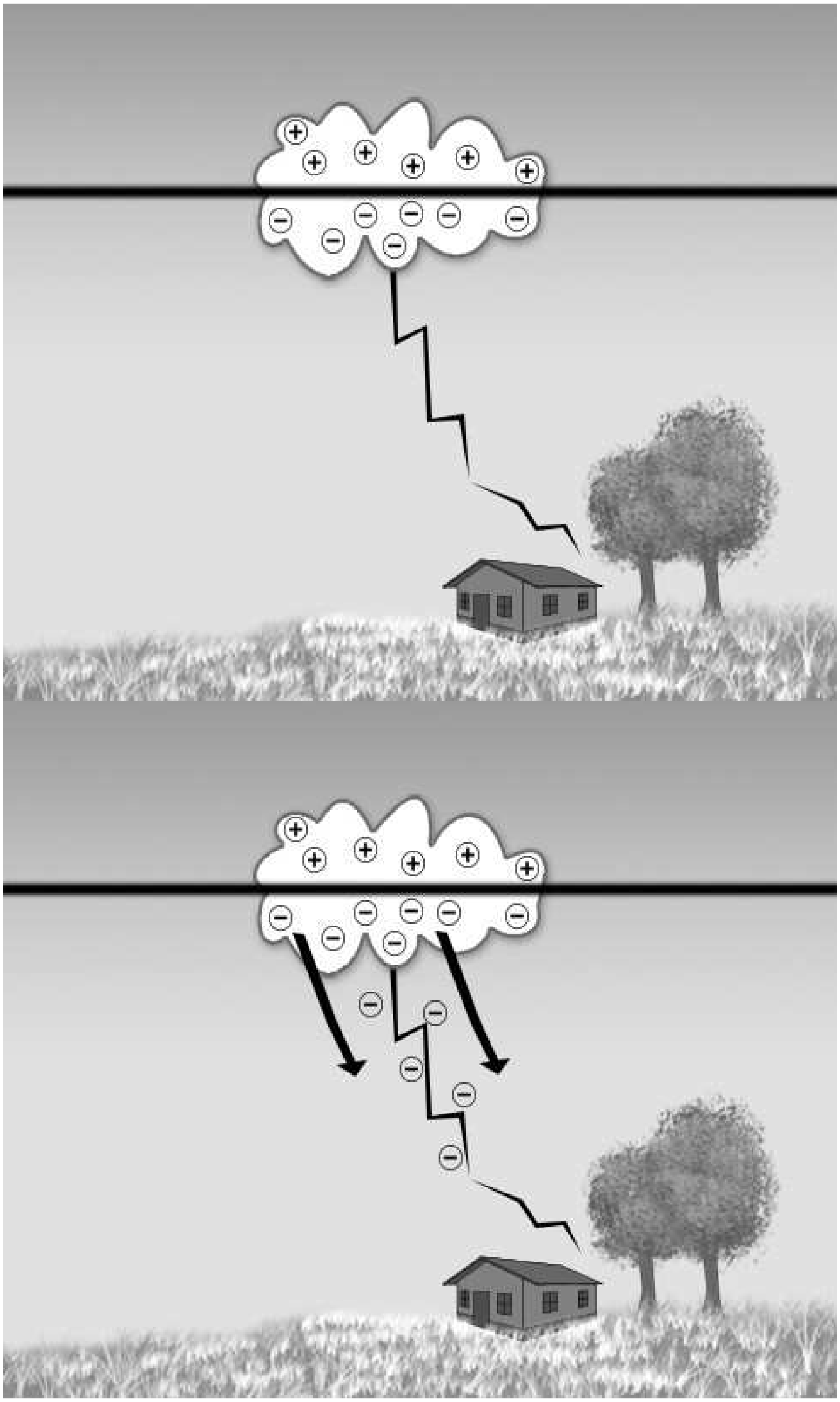


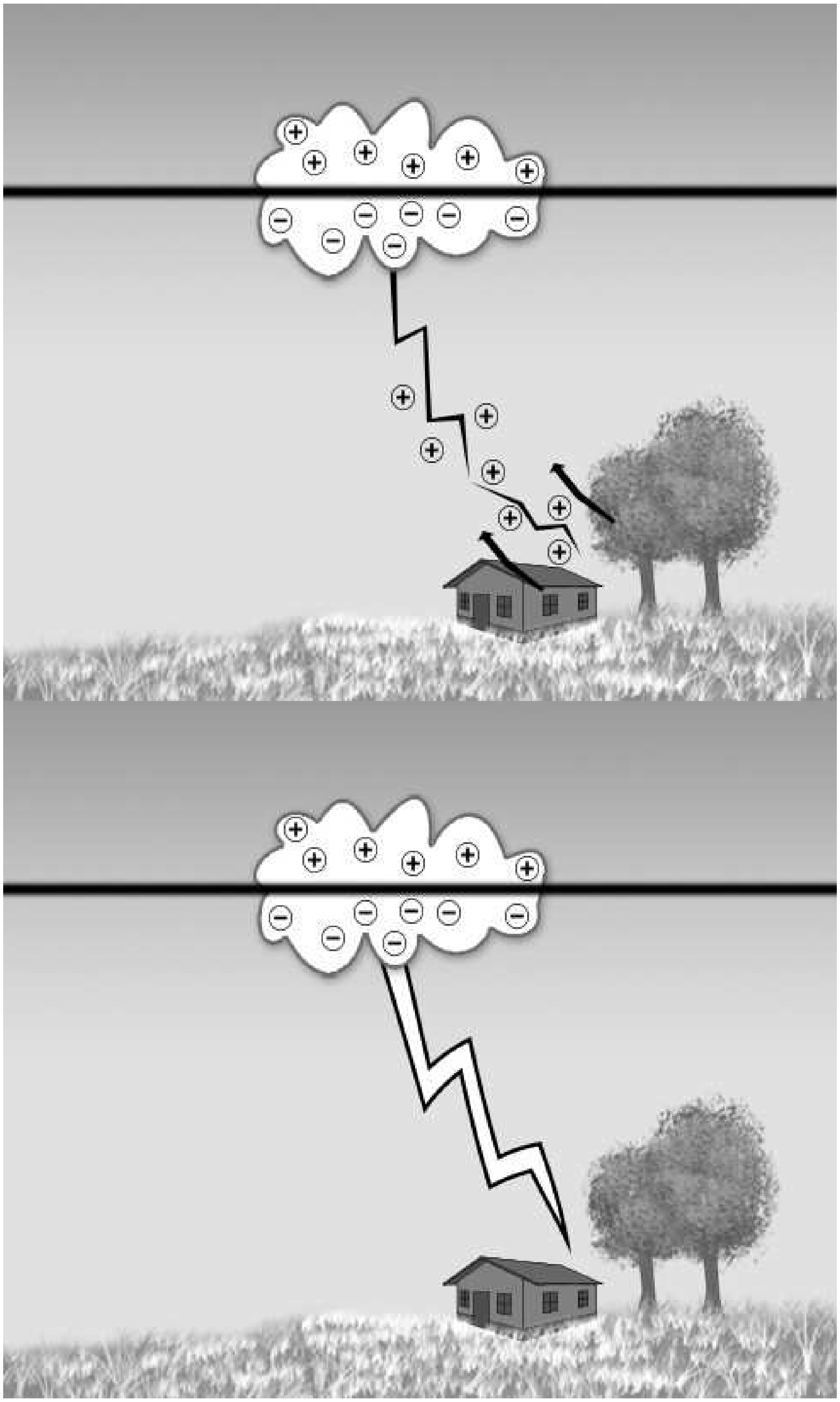












APPENDIX F

NASA TASK LOAD INDEX

Rating Scale

Please move the sliders to the appropriate level for each question

1. Mental Demand
    a. How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?
2. Physical Demand
    a. How much physical activity was required? Was the task easy or demanding, slack or strenuous?
3. Temporal Demand
    a. How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?
4. Overall Performance
    a. How successful were you in performing the task? How satisfied were you with your performance?
5. Frustration Level
    a. How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?
6. Effort
    a. How hard did you have to work (mentally and physically) to accomplish your level of performance?

Assessment Weights

1. Which was more important to your experience?
    a. Effort
    b. Performance
2. Which was more important to your experience?
    a. Temporal Demand
    b. Frustration
3. Which was more important to your experience?
    a. Temporal Demand
    b. Effort
4. Which was more important to your experience?
    a. Physical Demand
    b. Frustration
5. Which was more important to your experience?
    a. Performance
    b. Frustration
6. Which was more important to your experience?
    a. Physical Demand
    b. Temporal Demand
7. Which was more important to your experience?
    a. Physical Demand
    b. Performance
8. Which was more important to your experience?
    a. Temporal Demand
    b. Mental Demand

9. Which was more important to your experience?
    a. Frustration
    b. Effort
10. Which was more important to your experience?
    a. Performance
    b. Mental Demand
11. Which was more important to your experience?
    a. Performance
    b. Temporal Demand
12. Which was more important to your experience?
    a. Mental Demand
    b. Effort
13. Which was more important to your experience?
    a. Mental Demand
    b. Physical Demand
14. Which was more important to your experience?
    a. Effort
    b. Physical Demand
15. Which was more important to your experience?
    a. Frustration
    b. Mental Demand

APPENDIX G

SYSTEM USABILITY SCALE

1. I think I would like to use this system frequently
    a. Strongly Disagree
    b.
    c.
    d.
    e. Strongly Agree
2. I found the system unnecessarily complex
    a. Strongly Disagree
    b.
    c.
    d.
    e. Strongly Agree
3. I thought the system was easy to use
    a. Strongly Disagree
    b.
    c.
    d.
    e. Strongly Agree
4. I think that I would need the support of a technical person to be able to use this system
    a. Strongly Disagree
    b.
    c.
    d.
    e. Strongly Agree
5. I found the various functions in this system were well integrated
    a. Strongly Disagree
    b.
    c.
    d.
    e. Strongly Agree
6. I thought there was too much inconsistency in this system
    a. Strongly Disagree
    b.
    c.
    d.
    e. Strongly Agree
7. I thought there was too much inconsistency in this system
    a. Strongly Disagree
    b.
    c.
    d.
    e. Strongly Agree
8. I found the system very cumbersome to use
    a. Strongly Disagree
    b.

c.
d.
e. Strongly Agree

9. I felt very confident using the system
    a. Strongly Disagree
    b.
    c.
    d.
    e. Strongly Agree
10. I needed to learn a lot of things before I could get going with this system
    a. Strongly Disagree
    b.
    c.
    d.
    e. Strongly Agree

APPENDIX H

CONSENT FORM

CONSENT FORM
Text Isolation as a Signaling Method for Learners

RESEARCHERS
Dr. Scotty D. Craig, Ph.D and Josh Chin have invited your participation in a research study. Dr. Craig is an Assistant Professor and Josh Chin is a graduate student in the Human Systems Engineering Program and at Arizona State University.

STUDY PURPOSE
Studies have shown that signaling methods, which are used in web-based learning materials, are effective at helping to increase learning performance in certain situations. This study aims to investigate the viability of a novel method for signaling learners to important information.

DESCRIPTION OF RESEARCH STUDY
If you decide to participate, then you will join a study involving research of learning with different signaling methods in a multimedia setting. Your participation in this online study will last for 60 minutes. You will be asked to take a brief questionnaire, read information about the formation of lightning and then complete another series of questionnaires. You will be paid $1.00US for your participation in this and approximately 150 people will be participating in this study.

RISKS
There are no known risks from taking part in this study.

BENEFITS
While the main benefit of your participation will be the better understanding of a novel signaling method for multimedia environments, you will also have the opportunity to improve your understanding of the formation of lightening.

CONFIDENTIALITY
All information obtained in this study is strictly confidential. The results of this research study may be used in reports, presentations, and publications, but the researcher will not identify you or your individual results. In order to maintain confidentiality of your records, Scotty D. Craig, PhD will ensure no identifying information will be collected that will link the data to the individual it was collected from. All data will be kept electronically and will be deleted after a 5 year period from the date of publication as is customary with the field.

WITHDRAWAL PRIVILEGE
Participation in this study is completely voluntary. It is ok for you to say no. Even if you say yes now, you are free to say no later, and withdraw from the study at any time.

COSTS AND PAYMENTS
The researcher wants your decision about participating in the study to be absolutely voluntary.
Yet they recognize that your participation may pose some time inconvenience. In order to compensate you, you will be paid $1.00US through Amazon Mechanical Turk.

VOLUNTARY CONSENT
Any questions you have concerning the research study or your participation in the study, before or after your consent, will be answered by Scotty D. Craig, Ph.D at Santa Catalina Hall, Ste. 150, 7001 E. Williams Field Road, Mesa, Arizona 85212. Phone: 480-727-4723

If you have questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk; you can contact the Chair of the Human Subjects Institutional Review Board, through the ASU Office of Research Integrity and Assurance, at 480-965 6788.

Remember, your participation is voluntary. You may choose not to participate or to withdraw your consent and discontinue participation at any time without penalty

Clicking on the button below indicates your consent to participant in this research study.

-