2015

Configural Displays: The effects of salience on multi-level data extraction

Audrey Fok
University of Central Florida

CONFIGURAL DISPLAYS: THE EFFECTS OF SALIENCE ON MULTI-LEVEL DATA EXTRACTION

by

AUDREY WINGSZE FOK
B.S. Maryville University of Saint Louis, 2009
M.A. University of Central Florida, 2014

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2015

Major Professor: Mustapha Mouloua

ABSTRACT

Displays are a useful tool for users and operators to understand information quickly. Configural displays are effective in supporting focus and divided attention tasks through the use of emergent features. Emergent features are highly salient and are generally used to support divided attention task However, due to the salience of emergent features, a potential performance costs to focused attention tasks arises with configural displays. To address this cost, semantic mapping has been used to map salience techniques to information needed by focus attention tasks to increase their saliency (Bennett & Walters, 2001; Bennett et al., 2000). Semantic mapping is the process of mapping the domain constrains to the display, which in turn is mapped to the users capabilities and limitations to understand that domain data. The objective of this dissertation is to extend the use of semantic mapping to address potential performance costs of configural displays for hierarchical domains using the scenario-based training (SBT) instructor domain.

Two studies were conducted to examine the effects of salience application and salience type on data extraction accuracy and response time performances at low-level, mid-level, high-level, and a remediation task. The first study examined the effects of one salience technique mapped to the display. This study employed a 2(low or mid application) X 3(baseline, color techniques, and alphanumeric techniques) mixed model design in which 63 participants completed 3 blocks of 32 trails each using displays with the salience techniques mapped to either low- or mid-level data. Results from the first study showed that salience type had a significant impact on multi-level data extraction performance, but interactions were not found. The second study changed the manipulation of application and mapped two salience techniques display at the same time, using either the same technique or a combination of different techniques. The same experimental design was utilized and 65 participants completed study 2. Results of study 2 showed that different application resulted in greater improvements of performance and specific salience combinations were found better support data extraction performance. Across study analyses were also performed and revealed that more salience is not better than less salience. Instead it is the specific mapping of salience type and application which improves performance the most. Overall, these findings have major implications for theories of semantic mapping, attention and performance, and display design of hierarchical domains.

ACKNOWLEDGMENTS

I would like to thank Dr. Mustapha Mouloua, my adviser and committee chair, for his extraordinary mentorship and unwavering support during my graduate career. He is truly the embodiment of a great mentor: he has limitless patience, a creative mind, genuine compassion, and invested welfare and scientific development of his students. Under his tutelage, I have learned to think critically, write scientifically, and be the researcher I am today.

Additionally, I would like to thank the other members of my committee, Dr. Peter Hancock, Dr. James Szalma, and Dr. James Pharmer, for their guidance and understanding throughout this long process. They have been an outstanding source of encouragement and scientific expertise.

I would also like to thank my friends and coworkers at NAWCTSD who have provide encouragement, support, and always willing to help me throughout this journey. I am grateful for all the feedback, scientific discussions, and words of inspiration. Without their support, this dissertation would not have gone as smoothly.

Finally, I would like to express my deepest gratitude to my parents for their continuous love, support, and encouragement throughout my life, and particularly during trying times in my graduate studies. To my brother, for setting the academic standard so high and challenging me to achieve goals I didn’t know I could. To my fiancé, Allan, for all his support and love throughout this process and continuous belief in me, even when I did not believe in myself. I could not have finished this dissertation without each of you. Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION AND BACKGROUND

Problem Statement

Displays are used to aid users in understanding complex information. Through pattern recognition, users can comprehend information more quickly and with less workload via the display (Bennett & Flach, 1992; Szalma, 2011; 2002). How information is presented on displays is important because users need to gather and understand the appropriate information to decide on a course of action that may be time sensitive. For example, process control operators must be able to quickly extract information from a display to make timely adjustments to the system (Bennett & Flach, 2011). Data extracted from the display are used for the operator's focused and divided attention tasks. Focused attention tasks are tasks associated with obtaining information on individual variables while divided attention tasks involve obtaining information about the interaction or relationship between variables (Wickens & Carswell, 1995; Bennett & Flach, 1992; 2011; Holt, Bennett, & Flach, 2011). Displays need to be effective in facilitating the decision-making process. The efficacy of different types of displays in supporting focus and divided attention tasks has been extensively investigated (See Bennett & Flach, 1992 for a review). These displays include separable displays, objective displays and configural displays. Separable displays are displays where individual variables are represented individually. Bar graphs are an example of a separable display because each bar on the graph depicts one variable. Studies have shown that separable displays are effective for focus attention tasks because they represent variables independent of each other. However, when variables need to be integrated for divided attention tasks, separable displays may not be optimal for visualizing this data (Wickens & Carswell, 1995). Instead, research suggests that object displays are preferable for integrated data (Wickens & Carswell, 1995). Object displays use geometric forms to depict data, enabling the integration of variables and the presentation of relationships. In addition, other studies indicate configural displays are also effective in supporting focus and divided attention tasks (Bennett, Payne, Calcaterra, & Nittoli, 2000). Briefly, configural displays produce emergent features by visualizing each individual variable while allowing the relationship between the individual variables to emerge (a more detailed description of configural displays will be provided below). Emergent features formed from the interaction of individual variables have a higher visual prominence than the individual variables themselves (Bennett & Flach, 2011). Due to the highly salient nature of these emergent features, the components of the emergent features become less salient and therefore more difficult to extract from the configural display (Bennett & Flach, 1992). Salience is defined as "how well a particular visual feature stands out relative to other features that are present" (Bennett & Flach, 2011, p. 184). Therefore, information depicted by the emergent features is more easily extracted from the display compared to the information contained within their elemental parts.

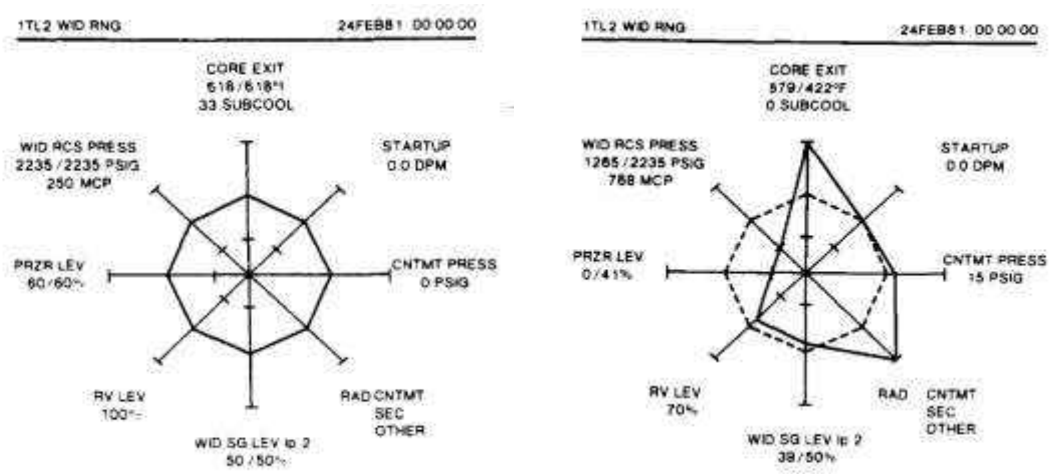
Bennett and colleagues (2000; 2001) used Semantic Mapping to map design techniques to the display elements in order to increase their salience, thereby increasing the efficiency of extracting data visualized by these elements and offsetting the potential performance cost of configural displays. Semantic mapping is the process of mapping domain constraints to the display, which in turn is then mapped to the user's capabilities and limitations in understanding that domain data (Bennett & Flach, 1992). These techniques were applied to a display showing the process control work domain, which includes two levels of data: individual variables and integrated data (Bennett & Walters, 2001; Bennett et al., 2000). However, other domains may require more levels of information to be represented (e.g., individual variables and multiple levels of integrated data). The central focus of this dissertation will be to examine how salience techniques can be mapped to multi-level domain data (and the corresponding display) using the Semantic Mapping approach to counterbalance the potential cost associated with configural displays of multi-level data domains.

To investigate these questions, a fundamental understanding of configural displays and the differences in their inherent salience will be established. Secondly, the Semantic Mapping Principle and its application in counteracting configural display costs will be also be described. Finally, the research questions proposed in the literature review will be applied to and tested within the Scenario-Based Training (SBT) Instructor Work Domain. Thus, the SBT Instructor Work Domain will be described and how the domain data is mapped to a configural display will be shown.

Configural Displays

Configural displays produce higher-order visual properties (i.e., emergent features) (Bennett & Flach, 2011). Configural displays come in many different varieties. In bar graph configural displays, individual bars of data configure to produce emergent features of linearity (for example, a line connecting the maximum or top of each bar depicts the relationship of the individual variables; Bennett & Flach, 2011). Object displays can also be considered configural displays if emergent features arise from the interaction of individual constituents. In object configural displays, the individual measures are represented by analog values (such as points on a graph) and these "configure" to produce higher order geometrical forms (Bennett & Flach, 2011). Bennett and Flach (2011) describe the concept of a configural display in which "pixels will configure to produce meaningful, low-level patterns such as lines and curves (i.e., graphical atoms). Graphical atoms (e.g., two lines) will configure to produce graphical fragments (i.e., emergent features) with higher order emergent properties (e.g., angles)... [and] collections of graphical fragments will produce analog, geometric forms" (p. 128). Each level of data in the displays interacts to form emergent features that are more salient than the individual components (e.g., pixels interact to produce emergent features of lines which subsequently interact to form more salient angles). Pomerantz and Pristach (1989) describe emergent features as "relations between more elementary line segments, relations that can be more salient to human perception than are the line segments themselves" (p. 636). Emergent features are inherently more salient due to their holistic properties, which have higher visual prominence than their elements (Pomerantz & Pristach, 1989; Bennett & Flach, 1992). In the configural display concept provided by Bennett and Flach (2011), geometric forms are more salient than the graphical fragments (e.g., angles), which are in turn more salient than lines, which are more salient than pixels, etc.

Woods, Wise, and Hanes (1981) describes an example of a configural display which maps numerous individual variables onto a single octagonal-shape geometric form (see Figure 1). Emergent features of symmetry and linearity are produced from the individual points on the graph. Changes in the value of the individual variables causes changes in the collective geometrical form. The distortions in the geometrical form reflect different states of the system and can be readily observed as symmetry differences in the geometrical form (the emergent features). In Figure 1a, when the state of the system is normal, linearity and symmetry produces the octagonal shape. However, changes to the state of the system lead to distortions of symmetry in the display (Figure 1b). These distortions are highly salient and easily distinguishable from the octagonal shape produced by the normal state.



a) b)

Figure 1. Configural object display (Woods, Wise, & Hanes, 1981)

Emergent Features

As previously discussed, emergent features are higher order visual properties produced in configural displays from display stimulus dimensions (e.g., symmetry; Bennett & Flach, 2011). Emergent features are more salient than their individual elements. Emergent features were initially described in a study of the discriminability of line segments either alone or within a configural context (Pomerantz, Sager, & Stoever, 1977). The stimuli in the study that did not have context, [“)”], resulted in significantly slower response times compared to stimuli presented with context, [“))”] (Pomerantz et. al., 1977).This study illustrated the configural superiority effect whereby combined features resulted in faster discriminability than the individual components themselves (Pomerantz et. al., 1977). Configurations that produce this configural superiority effect were termed emergent features, as these novel features only emerged upon the addition of context to individual parts (Pomerantz et al., 1977). Furthermore, “improvement in perception when context is added implies that the novel features happen to be more discriminable than the features by the targets without context" (Pomerantz et al., 1977, p. 433). The inherent salience of emergent features allows them to be easily discriminated from each other (Bennett & Flach, 2011). For example, the emergent feature of parallelism ["))"] and the emergent feature of symmetry ["()"] are readily distinguishable (Bennett & Flach, 2011). However, not all emergent features possess the same level of salience. Within a given configural display, many types of emergent features may arise. Bennett and Flach (2011) describe how nested, hierarchical geometries can be arranged within configural displays such that the most salient emergent feature is the closed and symmetrical geometric shape itself. Within the same display, the area of the shape can also be utilized as another emergent feature. These emergent features are typically the most salient. Next are the emergent features with intermediate levels of salience where portions of the shape collectively constitute the full geometric shape. Lastly, the elementary emergent features such as the lines and angles of the shape would have the lowest level of salience (Bennett & Flach, 2011). The variability in the salience of the emergent features within a configural display is a potential cost of the display.

Potential Performance Cost of Configural Displays

The complexity of salience levels associated with the elemental and emergent features of configural displays increases the difficulty of extracting information represented by the less salient elements compared to the more salient emergent features (Bennett & Flach, 1992). Bennett and Flach (2011) explain this occurrence with the following analogy: an individual tree with no other surrounding trees is easy to find but a tree surrounded by other trees (i.e., the tree within a forest) is more difficult to locate than the solitary tree. Although the forest is more salient (i.e., more easily seen) than a single tree, it is substantially more difficult to locate the single tree. By analogy, information depicted by dots on the configural display are less salient than information depicted by the shapes that are formed from the interactions between dots and therefore require more effort and time to extract from the display. The potential additional performance costs of extracting these less salient data will impact user performance when using the display. Researchers have implemented the Semantic Mapping Principle (which will be reviewed in more detail in the next section) as a strategy to offset these additional potential performance costs; however, these design strategies were only empirically examined on tasks of extracting two levels of data: individual variables data and integrated data (Bennett & Walters, 2001; Bennett, Payne, Calcaterra, & Nittoli, 2000; Bennett, Toms, & Woods, 1993). The research in this dissertation will extend these strategies by determining effective techniques to offset the potential performance cost of configural displays for hierarchical domains with more than two data levels.

Semantic Mapping Principle

The Semantic Sapping Principle is an approach to display design that focuses on the integration of information to support effective decision making (Bennett & Flach, 2011). This approach emphasizes user understanding of the work domain through the display rather than understanding the visual aspects of the display itself (Bennett & Flach, 2011). The efficacy of this principle is dependent on the mappings between the human user, display, and domain. Through the process of semantic mapping, constraints of the domain (e.g., the goals, functions, and requirements of the work domain) are mapped to a display, which is then mapped to user perception and pattern recognition capabilities and limitations to understand that domain data (Bennett, Nagy, & Flach, 2012). Sanderson and colleagues (1989) state that the goal of display design is to map process invariants to the display so that they are easily visible. When this is achieved, less demanding processing such as pattern recognition can be used rather than more demanding process, like computation (Sanderson et. al., 1989). Through the semantic mapping approach proper mapping of the domain to the display reduces operator workload through the use of perceptual processing rather than cognitive processing (Sanderson, Pipingas, Danieli, & Silberstein, 2003). Furthermore, researchers have determined that ecological interfaces and the semantic mapping approach facilitates pattern recognition, which is quick, accurate, and low in mental workload (Sanderson et. al., 2003).

Effective semantic mapping depends on the quality of the mapping between the user, display, and domain (Bennett, Nagy, & Flach, 2001). Each component introduces its own set of constraints (Bennett & Flach, 2011). Constraints from the domain include the goals, which then determine the tasks to be completed. The cognitive and perceptual capabilities and limitations of the user also present constraints. All these constraints must be accounted for when the domain is mapped to the display. Previous studies indicate that the use of semantic mapping in display design is effective for supporting both focus and divided attention tasks (Bennett, Toms, & Woods 1993; Holt et. al., 2011; Bennett et. al., 2000; Sanderson et. al., 1989). For example, In Figure 1, Woods, Wise, and Hanes (1981) mapped sensor values from power plant systems onto a configural object display to produce emergent features representing critical relationships. This display facilitated divided attention tasks due to the salient emergent features that reflected accurate relationships within these systems (Woods et. al., 1981). In addition, Sanderson, et. al. (1989) found that the use of a configural bar graph resulted in significantly better system fault detection compared an object display. Furthermore, semantically mapped configural displays outperformed bar graph displays in supporting divided attention tasks (Bennett, Toms, & Woods

1993). The results from studies examining the benefits of configural displays for focus attention tasks is more mixed. Researchers have noted that the use of configural geometric forms may add additional costs (Wickens & Carwell, 1995; Bennett & Flach, 1992). However, studies have also found ways for configural displays to support focus attention tasks (Bennett et. al., 1993; Bennett et. al., 2000; Bennett & Walters, 2001). These studies will be discussed further in a later section.

There are two main steps in semantic mapping (Bennett & Flach, 1992). The first step is to determine the semantics of the domain. Designers must initially define the information of the domain (i.e., domain data) required by operators and users of the display to complete required tasks and the relationship among tasks. This information can be determined through work domain analysis such as Rasmussen's (1986) Abstraction Hierarchy. The second step in the semantic mapping approach involves mapping the domain data to the display in a manner that produces emergent features. Each step will be described in more detail below.

Work Domain Analysis

Defining the semantics of the domain, which includes the domain data and its structure, is done through a work domain analysis. For example, data can be structured in a hierarchical format consisting of various levels of data that interact with each other to form additional, new levels of data (see Figure 2 for an example). Many levels of data can exist within a domain; all of these levels would have be determined at this stage of the semantic mapping approach. The work domain analysis typically used in this approach is Rasmussen's Abstraction Hierarchy (Bennett & Flach, 2011).

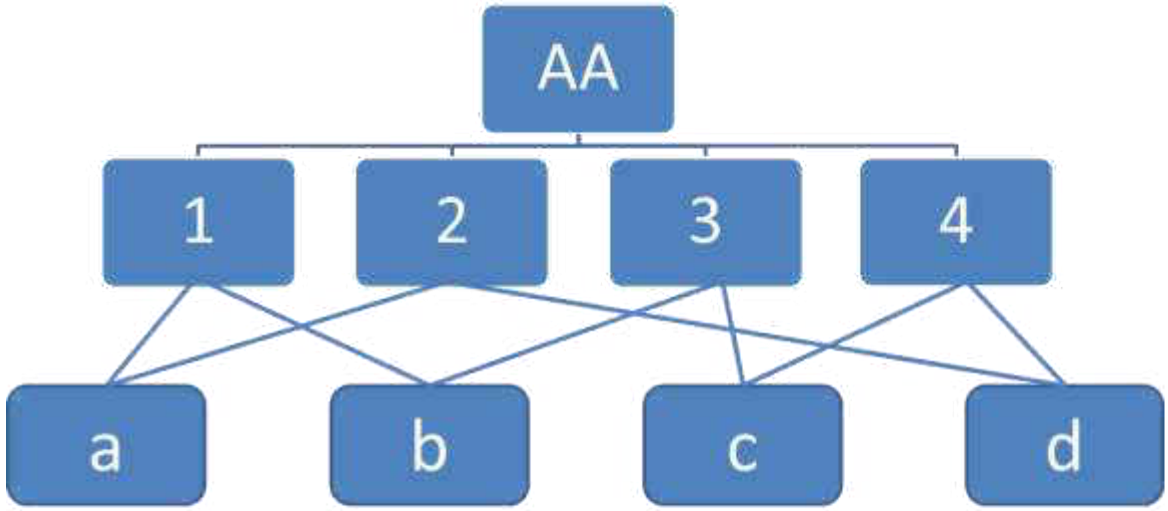


Figure 2. Example of hierarchical data.

Abstraction Hierarchy

The Abstraction Hierarchy is a framework for describing and deconstructing components of complex work environments that facilitates the understanding of the interaction between the domain and display (Rasmussen, 1985). This theoretical framework describes the work domain within a nested hierarchy of functional constraints. The hierarchy represents “mean-ends relationships” that can be utilized from the bottom-up or top-down (Rasmussen, 1986). From the bottom-up, the operator can understand how each component or function is used and how it supports higher level purposes. When the hierarchy is used top-down, the operator can recognize how the purpose and goals of the system are achieved by its components.

There are five levels in this hierarchy: Functional Purpose, Abstraction Function, General Function, Physical Function, and Physical Form. The Functional Purpose is the highest level and states the overall goals, purpose, and objectives of the system. Below this level is the Abstraction Function which describes the priorities, laws, and constraints of the system which cannot be broken (Burns & Hajdukiewicz, 2001). Additionally, the Abstraction Function explains how the system goals can be achieved. Rasmussen (1986) describes this level as "the semantic content of the physical signals and hence, the overall organizing principle" (p. 19). The next level, General Function, includes the functions and processes required to achieve the goal state. Task analyses generally belong to this level (Bennett & Flach, 2011). The remaining levels of the Abstraction Hierarchy pertain to the hardware of the system. The level of Physical Function describes the type of systems that will accomplish the functions determined at the General Function level. The last level, Physical Form, considers the physical details of the system, such as the appearance and configuration of the system. The operator of the system may shift between these levels of the hierarchy when appropriate (Vicente, 1999).

Mapping

The second step in the semantic mapping approach to display design is mapping the domain data to the configural display. In this crucial step, the domain constraints and relationships (determined from the work domain analysis) are mapped to the features of the visual display (Bennett & Flach, 1992; 2011). To map hierarchically structured data, low level data is visualized on the display as individual variables that interact to produce emergent features (representing the relationships between these low-level data variables; Bennett & Flach, 2011). For example, Figure 3 illustrates one way how hierarchical data can be mapped to a configural object display: the first level of data (a through d) is represented by dots; the second level of data, relationship data 1 through 4 (the collective interactions of a, b, c, and d), is represented by lines produced by the configuration/interaction of these dots; and highest level of data AA (the collective interactions of 1, 2, 3 and 4) is represented by the rectangle shape (formed by the configuration of lines from the second level of data).

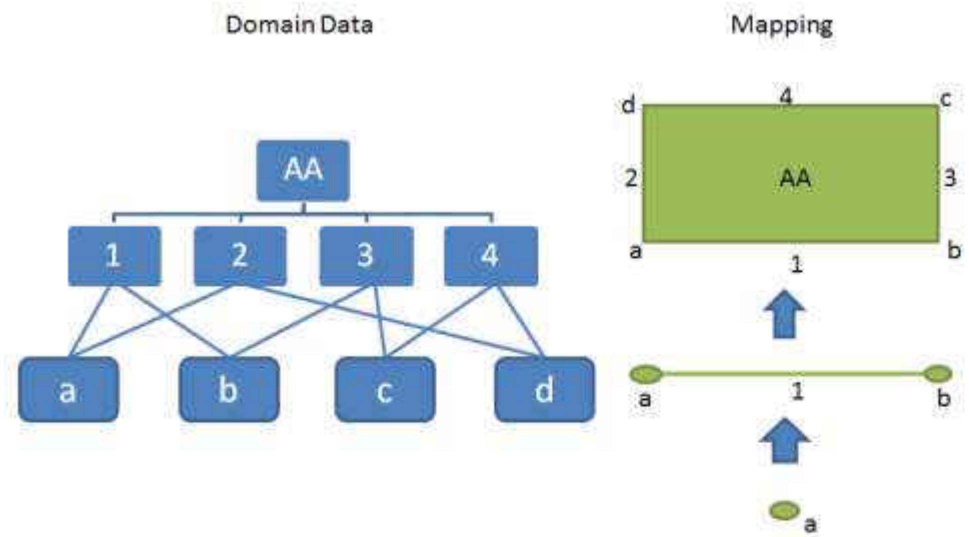


Figure 3. Hierarchical semantic mapping

The process whereby levels of data are mapped to elements of the configural object display is critical. Regardless of the type of display, mapping can significantly impact performance of data extraction (Bennett & Flach, 2011). Studies have demonstrated which mapping processes constitute good versus poor mappings as well as the associated benefits/costs (MacGregor & Slovic, 1986; Coury, Boulette, & Smith, 1989; Sanderson et. al., 1989). MacGregor and Slovic (1986) provide an excellent example of the importance of mapping. Two studies were conducted in which in the first study MacGregor and Slovic (1986) compared different several types of displays (bar graph display, configural face display, polar coordinate display, and deviation display) that required participants to use multiple pieces of information to predict when a runner would complete a marathon. They found that the configural face display resulted in significantly better performance and they determined which information cues were more critical for predicting the correct outcome. In the second study, mapping was manipulated within the same face display context; the critical information cues were either mapped to more salient emergent features (considered a well mapped display) or less salient features (poorly mapped display). The authors found that well-mapped face displays resulted in more accurate predictions of marathon times compared to poorly mapped face displays (MacGregor & Slovic, 1986). This study supports an important aspect of the semantic mapping principle: the mapping of the data to the display is more crucial for better task performance (e.g., focus and divided attention tasks) than the type of display. The same display type could result in good or poor performance depending on the mapping of domain data to the display.

In another study illustrating the importance of mapping in designing effective displays, Buttigieg & Sanderson (1991) used three different display types (bar graph, house object display, triangle object display) to examine how well mapped or poorly mapped emergent features affected user performance of these dynamic displays. Participants monitored displays for global failures and local failures. The authors found that independent of display type, well mapped emergent features were the most effective for detecting global failures (Buttigieg & Sanderson, 1991). The influence of well mapped emergent features on detecting local failures, however, was less conclusive.

A common theme that emerged from studies comparing well mapped displays versus poorly mapped displays is that when more salient emergent features depicted critical informational cues, performance was significantly improved (Holt et. al., 2011; Sanderson et. al., 2003). Generally, when the most salient features of the display represented less critical information, performance (in terms of information extraction and using the extracted data to make decisions) was negatively impacted (Bennett & Flach, 2011). Therefore, it is imperative that the most critical relationships in the domain data are mapped to the most salient feature of that display (Bennett & Flach, 2011). However, the pieces of data that form these critical relationships are more difficult to extract from the display than the relationships themselves (Bennett & Flach, 1992). This is the cost of configural displays. In figure 3, the dots depicted configure to produce more salient emergent features (i.e., lines), which then configure to produce even more salient emergent features (i.e., shapes). As each level of data is mapped to the display, the visual representation of that level increases in salience such that the highest level of data is most salient, the data level below that is the second most salient, and so forth. Figure 4 below illustrates the correlation between salience and data levels. Configural displays have the capability to depict several levels of data at once, which would then have different corresponding levels of salience; thus, certain pieces of data may be more difficult to extract from these displays than others. This configural display cost has been studied and display design techniques offset these costs have been proposed (Bennett et. al., 2000; Bennett & Walters, 2001), which are summarized in more detail below.

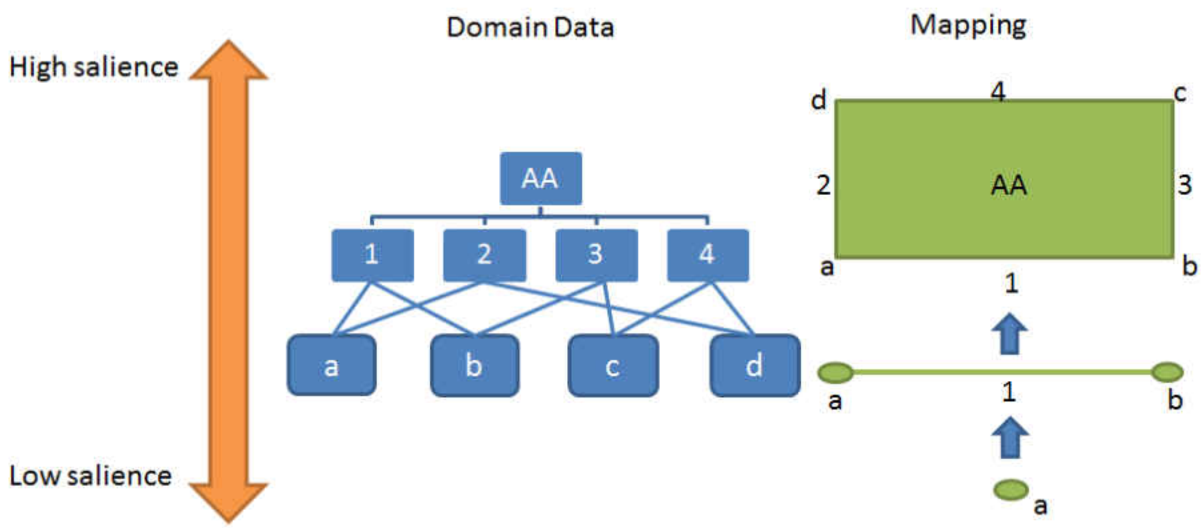


Figure 4. Salience relation to hierarchical semantic mapping

Mapping Salience Techniques to Offset Configural Displays Costs

Researchers have proposed that the cost of extracting low level data from configural displays was due to the inherent difference in salience between emergent features and the elements that constitute them (Bennett & Flach, 1992). They hypothesized that in order to improve low level data extraction performance, the salience of the visual features depicting the low level data on the display had to increase (Bennett & Flach, 1992; Bennett et. al., 1993). Several studies have tested the effect of different salience techniques on the extraction of low level data from displays (Bennett et. al., 2000; Bennett & Walters, 2001). Bennett et. al. (1993) proposed four different techniques that could be applied to low-level data to increase their salience on the display: color, scales, bar graph extenders, and alphanumeric values. These techniques were examined initially as a combination (Bennett et. al., 2000) and then individually

(Bennett & Walters, 2001). When compared to a baseline display with no salience techniques, the display using all four salience techniques significantly improved accuracy and response time for low-level data extraction (i.e., focus attention tasks; Bennett et. al., 2000). Bennett and Walters (2001) then investigated how these four salience techniques on their own or in varying factorial combinations impacted data extraction performance for focus attention tasks. This study showed that three of the four salience techniques improved low level data extraction performance (Bennett & Walters, 2001). The design techniques implementing bar graph extenders, scales, and alphanumeric values all improved timing and accuracy performance (with alphanumeric values increasing performance the most; Bennett & Walters, 2001). However, adding colors did not significantly improve data extraction performance relative to the baseline display. In this study, the color provided categorical information but not the quantitative information required by the constraints of the task and thus did not improve low level data extraction (Bennett & Walters, 2001). Bennett and Flach (2011) later concluded from these results that increasing the visual salience of low level data on the display was not sufficient to improve extraction (e.g., adding colors increased salience but did not improve performance). Instead, they suggested that these salience techniques must also be appropriately mapped to the constraints of the task. That is, the low level data extraction tasks in the above studies involved participants reporting a quantitative value and therefore only the techniques that were well mapped to this task increased performance (Bennett & Walters, 2001). The bar graph extenders, scales, and digital value techniques all increased salience by providing quantitative data which mapped well to the quantitative task. Furthermore, alphanumeric values produced quantitative data that eliminated all mental estimates, and therefore mapped to the task constraints the most appropriately (Bennett & Flach, 2011).

The previous studies described the application of salience techniques to the nuclear power plant process control domain and the experimental tasks involved extracting individual variables as well as integrated data (Bennett et., al., 1993; Bennett et. al., 2000; Bennett & Walters, 2001). In these studies, the process control domain contained two levels of data that were required to extract individual valuables and integrated data. These studies showed that when salience techniques were applied to the first level of data (in a two-level domain data hierarchy), techniques that were well-mapped to tasks constraints were effective in improving extraction of first level data from a configural object display (Bennett & Walters, 2001). Additionally, configural object displays applying a composite of salience techniques (e.g., colors, bar graph extenders, alphanumeric values, and scales) to the first level of data in a two-level domain data hierarchy improved performance of first level data extraction (Bennett et al., 2000). However, domains may contain multiple levels of integrated data and thus there are potential performance costs for extracting data at all of these levels except for the highest level (which is the most salient). It is not known how these salience techniques will affect data extraction performance at all of these levels within the display.

Multi-level Data Extraction

As discussed above, in domains containing multiple levels of data that are mapped to the configural display, data from the first data level (the lowest level) is more difficult than extracting data from the second level; additionally, extracting data from the second level is more difficult than extracting data at the third level of data. The data extraction at each level is impacted by the salience of the data above it; that is, data level extraction performance is heavily influenced by the salient emergent features it produces. As summarized above, researchers have attempted to improve low level data extraction performance by increasing the salience of these data and mapping them appropriately to task constraints. However, these studies have only demonstrated the efficacy of these techniques on the process control domain when two levels of data were observed (Bennett et. al., 1993; Bennett et. al., 2000; Bennett & Walters, 2001). As such, there remains many open questions regarding the impact of salience techniques on data extraction. It is unknown how these techniques would improve data extraction from domains containing more than two levels of data. For example, in a domain containing three levels, how would these techniques impact performance of data extraction at all levels? At which level of data should salience techniques be applied so that data extraction performance is supported at all levels? Which salience techniques are most effective in improving performance in these multi-level domains? An example of a multi-level domain is the Scenario-Based Training (SBT) instructor work domain. In these domains, three data levels are required: instructors must account for low level data (e.g., timing and accuracy measures of specific tasks), mid-level data (e.g., overall task performances as measured by the combination of timing and accuracy data), and high level data (e.g. trainee performance as measured by multiple tasks performance). Based on the continuum of salience, the highest level of data is most salient and applying salience techniques to this level would only further increase the inherent gap in salience between the highest level and the levels below. Studies have also shown that increasing salience of the lowest level of data will increase data extraction at that level (Bennett et al., 2000; Bennett & Walters, 2001). However, within a three-level domain, there is likely to be a cost associated with data extraction at the middle level (because the most salient, high level data impacts data extraction at this middle level). While applying salience techniques may improve extraction of middle level data, these techniques may potentially affect the extraction of the lowest level of data (again, by increasing the salience gap between the middle and lowest levels of data). Thus, the application of salience techniques must be evaluated such that extraction is not hindered at any one level and is optimized for every level of data. The research proposed in this dissertation will empirically examine at which levels of data salience techniques should be applied to a configural display within a multi-level domains. Furthermore, this research will evaluate which salience techniques are the most effective for improving data extraction at these various levels.

Context: Scenario-Based Training Instructors

Scenario-Based Training

Training is invaluable in the execution of tasks for many fields including aircraft pilots, emergency response, and the military. However, live field training can be costly and does not provide the control necessary for trainees to practice certain skills. Simulation training, on the other hand, is less costly as it does not require the same human resources and equipment associated with live field training. Simulation training also enables trainees to practice their skills within safe, reproducible, controlled environments. Simulation training presents advantages over conventional classroom teaching as well because it requires trainees to assume an active role in the simulated environment and gain experiences rather than just passively observing within a classroom (Nicholson, Fiore, Vogel-Walcutt, & Schatz, 2009). One of the most common strategies used in simulation training is Scenario-Based Training (SBT). In SBT, the trainees complete training in the form of controlled exercises that present realistic environmental stimuli and feedback (Cannon-Bowers, Burns, Salas, & Pruitt, 1998). Each SBT scenario is based on pre-determined learning objectives; specific events and environments are incorporated into each training scenario which best help trainees achieve these learning objectives (Oser, Cannon-

Bowers, Salas, and Dwyer, 1999). Instructors in each scenario assess the trainee's performance and determine whether the trainee has met the training objectives. The instructors also provide feedback on the trainee's performance and learning during or after the scenario session.

Scenario-Based Training has been applied to many simulation training systems. For example, Tactical Action Officers (TAO) use a trainer called the TAO Sandbox which provides training scenarios and real-time automated feedback (Milham, Pharmer, & Fok, 2013b). Other SBT simulators, such as the P-8 Whole Task Trainer (WTT), are not equipped with automated feedback features. Thus, instructors must manually provide feedback to complete the SBT process.

The Instructor's Role

Regardless of the trainer used, the roles of the instructor are paramount to the effectiveness of the SBT process. An instructor's tasks can be divided into two categories: during the scenario and post scenario. During the scenario session, instructors may have both instructing and/or role-playing tasks, which vary depending on the training systems. Instructors may be required to play the role of active crew members: listening to communications, moving entities on the maps, and conducting themselves as a fellow team member would. Depending on the training event, instructors may play a number of different roles during the scenario session. Instructing tasks during the scenario session, on the other hand, require the instructor to assess, diagnose, and provide remediation (i.e., feedback) based on a trainee's performance (Milham, Pharmer, & Fok, 2013a). One of the main goals of the instructor is to determine when he/she needs to provide feedback, which is dependent on the assessment and diagnosis the trainee's performance. Assessment requires the instructor to monitor all of the trainee's actions and determine if any errors were made. After assessment, the instructor must diagnose the trainee's performance. Diagnosis entails judging or grading the trainee's performance relative to the goals of the SBT scenario and determining which aspect of the performance needs real-time feedback. For example, if the trainee frequently sends tardy reports, then the proper diagnosis would be that the trainee performs poorly on making timely reports, perhaps due to the fact that the trainee is unaware of the need for timely reports. The instructor would then provide remediation or feedback in real-time to the trainee advising him/her to send reports on time. Real-time feedback is a significant part of training and has been shown to be effective in improving performance (Azevedo & Bernard, 1995; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kulik & Kulik, 1988; Milham et. al., 2013a).

Alternatively, the instructor can reserve giving feedback until after the scenario session, during a debriefing session known as After Action Review (AAR). Tasks performed during this session fall into the second category of instructor tasks, post scenario. During AAR, the instructor will discuss the trainees' performance and address areas where trainees need improvement. At this time, instructors may also provide examples of appropriate courses of action that the trainee should have performed during the scenario session. Collectively, these role-playing and instruction duties place a tremendous workload burden on the instructor (especially if there are multiple trainees per SBT session), which will be discussed below.

Instructor Workload

Workload is defined as the effort needed to complete a task or series of tasks in specific environments (Cain, 2007). Instructional simulations during SBT can induce heavy workload burdens on the instructors. These exercises often demand extensive instructor participation

(Salas, Rhodenizer, & Bowers, 2000; Ross, Phillips, Klein, & Cohn, 2005; Oser, Gualtieri, Cannon-Bowers, & Salas, 1999). Many key components of simulation/scenario-based training are not provided by the simulation technology but instead must come from the instructor (Schatz, Oakes, Folsom-Kovarik, & Dolletski-Lazar, 2012). Instructors must also manage many aspects of the training (Schatz, Oakes, Folsom-Kovarik, & Dolletski-Lazar, 2012). Collectively, these roles greatly increase the instructor's workload burden during adaptive SBT, which may affect task performance (Raby & Wickens, 1994; Morris & Leung, 2006).

While there are systems that can automatically assess performance and report these assessments to the instructor (see Milham et. al., 2013a and Milham, Pharmer, & Fok, 2015), he/she is still required to constantly monitor these assessments to determine the overall performance of the trainee. However, instructors cannot focus on a single signal display; they must toggle between this screen and the display for role-playing, the map screen, the evaluation display (or a paper for recording these evaluations), and any other miscellaneous screen(s) required by that training exercise. A summary display which presents an "at-a-glance" picture of the trainee's performance would therefore be very useful in alleviating this workload burden. This display would provide a quick snapshot of trainee performance and help indicate which aspects of the training require feedback, and this is the intended use case scenario for the displays designed for this research study.

CHAPTER 2: THE CURRENT STUDY

The current study utilizes a configural display to examine how salience techniques applied to the low and mid-levels of a semantically mapped, three level data domain impacts data extraction at all levels of data. Using the SBT Instructor work domain mapped to a polar coordinate configural display, the effects of salience techniques will be measured across two variables: the level of data at which these techniques are applied to and which particular salience technique (or a combination of techniques) is used. The study will examine the effects of these techniques on: 1) the ability to extract low, mid, and high levels of data, and 2) the ability of the instructor to know when remediation is needed. In the next section, a work domain analysis to develop and evaluate display design techniques (e.g., salience techniques) that can support SBT Instructors will be presented. That section will be followed by a description of how the domain data (determined from the work domain analysis) will be mapped to the configural display. Finally, the research objectives and hypotheses will be stated.

Scenario Based Training Instructor Work Domain Analysis

A work domain analysis of the SBT Instructor work domain is shown in Table 1. This abstraction hierarchy describes the levels of abstraction for an Instructor Operating System (IOS) Performance Summary Display. Because a SBT instructor must be aware of when the trainee is performing poorly such that remediation is required, the goal of the system will be to determine when trainee performance is not within the desired performance range. This is the first level of abstraction. The second level of abstraction describes the abstract functions of the SBT Instructor work domain. At this level, instructors must determine when to provide remediation by monitoring several variables and the relationships between these variables. In this case, the variables include the trainee's overall performance, his/her process learning objectives performances, and his/her course of action performances. The next abstraction level, the General Work Activities and Functions, outlines the functions that the system must complete to achieve the goals described. For the SBT Instructor work domain, the system functions include collecting and recording the trainee's performance and making automated assessments. Additionally, the system must summarize the performance of the trainee. These summaries are determined by the relationships between overall performance and learning objectives performances, which are determined from the relationships described in Table 1. The next level, Physical Function, describes the data the system will collect and assess: the trainee's action and interaction with the training system. The last level is Physical Form. This level describes the physical appearance of the data; for this study the domain data consists of the trainee location and physical appearance of the training system.

Table 1. Scenario-Based Training Instructor Work Domain Analysis

Abstraction Hierarchy		
Levels	Definition	SBT Instructor Domain
Goals, Purpose	Goals of system	Relation between optimal performance and actual trainee performance
Priority Measures and Abstract Functions	Underlying laws and principles. Priorities that must be achieved	Overall trainee performance TAO performance = C+R+TP1+TP2 Relation between overall performance and learning objectives performance. Learning Objectives performance Communication (C) performance = O+U Resource Usage (R) performance = O+T Team Participation (TP1) performance = P+U Tactical Procedure (TP2) performance = P+T Relation between objectives performance and COA performance. Trainee's Course of Action performance Provide Navigational Orders (O) Provide Status Updates (U) Follow Mission Plan (P) Manage Threats (T)
General Work Activities and Functions	Functions required to achieve the goal state. Processes involved	Collection of performance measures Performance assessment
Physical Function	What is required to accomplish a given function?	Trainee's actions and interactions with the training system
Physical Form	Physical Details; Appearance, Location	Trainee location, training system physical appearance

The domain data required for instructor assessment, diagnosis, and remediation are arranged in a hierarchical format, as described in Figure 5. There are three levels of domain data in this hierarchy (course of action performance, learning objectives performance, and overall TAO performance) that must be considered by the instructor prior to determining if trainee performance is acceptable. For the purposes of this study, the first level of domain data in this hierarchy, the course of action performance, will be referred to as the lowest level of data. The second and third levels of data in the hierarchy, the learning objectives performance and the TAO performance, will be considered the mid and high levels of data, respectively.

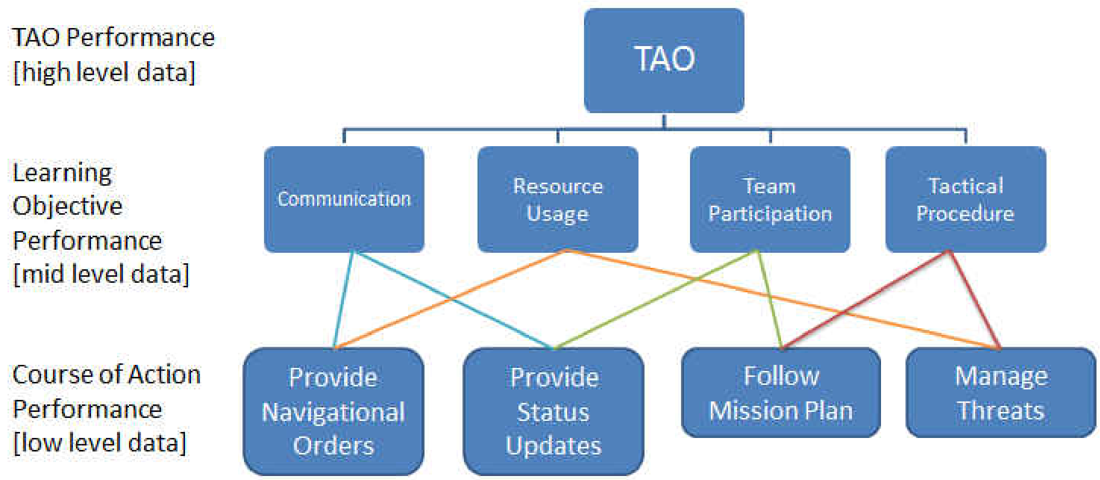


Figure 5. Scenario-Based Training Instructor Hierarchical Data

Mapping SBT Instructor Domain Data to a Configural Display

The SBT Instructor work domain analysis was used to map the domain data to a configural display in accordance with the semantic mapping approach to display design. A polar coordinate configural display was chosen to represent the requisite data for instructor assessment, diagnosis, and remediation. This display was chosen for its inherent nested format which would map well with hierarchal SBT instructor domain data. Figure 6 shows the polar coordinate configural display with the mapped SBT Instructor domain data. The course of action (COA) performance data are represented by the points on the horizontal and vertical lines of the polar coordinate display. The two lines intersect perpendicularly; the center of their intersection has the value of zero. The lines are on a continuum of 0% (center) to 100% (one end point) whereby the end points of the horizontal line are the Manage Threat and Status Update course of action performance measures, and the end points of the vertical line are the Follow Mission Plan and Navigational Orders COA performance measures. Interactions of these COA performances form learning objectives; these relationships are represented on the display by the lines formed between two course of action performances. For example, the COA performances of Status Updates and Navigational Orders together form the learning objectives performance of Communications. On the display shown in Figure 6, if both the Status Update and Navigational Order performances are set at 100%, then the line adjoining the two COA performances represents the learning objective performance of Communication and this line forms a right triangle bounded by the vertical and horizontal axes. The high level data, the TAO performance, is visualized as the shape formed by the learning objectives representations (i.e., the lines). If the COA performances are all set to 100%, then the geometric shape collectively formed by the learning objectives performance relationships will be a rhomboid or diamond shape (Figure 6). This geographic shape will vary depending on learning objectives performance relationships, which are in turn determined by the COA performance measures.

To summarize, the COA performance measures (low level data) are depicted as the dots on the polar coordinate configural display. These dots configure to form lines (a less salient emergent feature) that represents learning objectives performances (mid-level data). The lines configure to produce the shape (a more salient emergent feature) that illustrates the overall TAO performance (high level data).

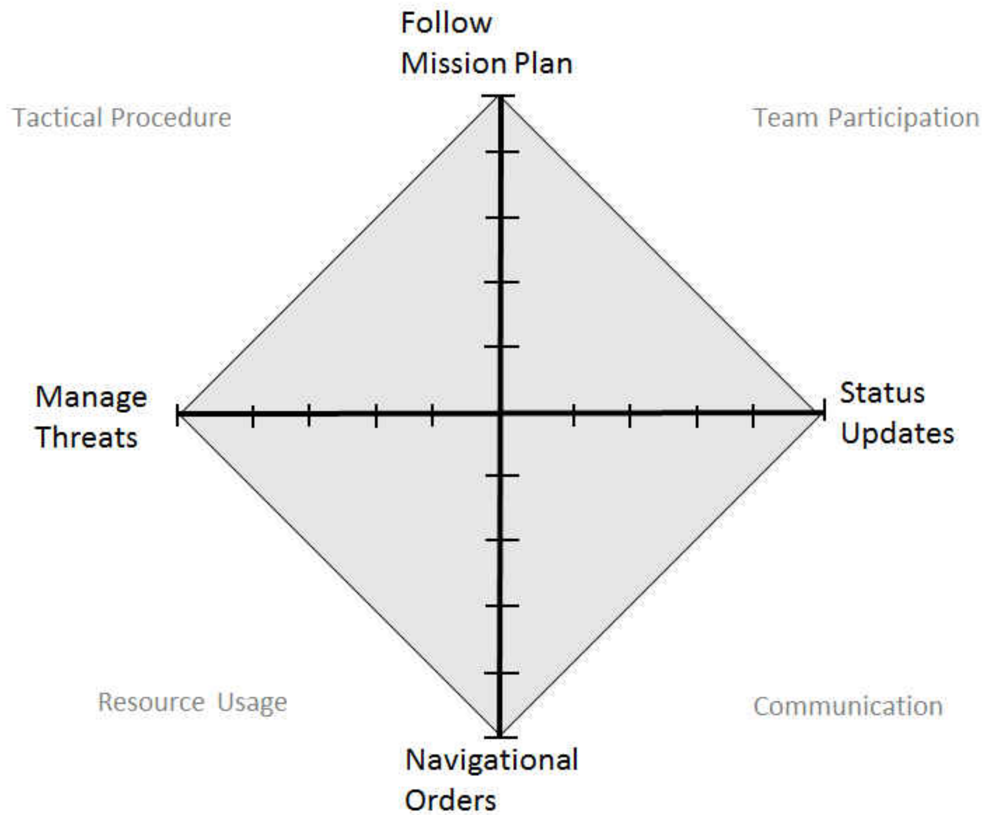


Figure 6. Mapping of SBT Instructor work domain data to configural display.

Criteria for evaluation of the domain data are shown in Table 2. The COA performance data for Follow Mission Plan, Status Updates, Navigational Orders, and Manage Threats, are expressed as percentages on the display. A good performance in any COA that is rated as or greater than 80%, and anything under 80% is considered a poor performance. If at least one COA performance is rated below 80%, the corresponding learning objectives performance is classified as poor. A learning objectives performance is considered good when all of the underlying COA performances are equal to or greater than 80%. The TAO performance data indicates whether the TAO passed or failed the task. If at least one learning objectives performance is rated as poor, that constitutes a task failure; all learning objectives performances must be graded as good for the TAO to pass. Finally, in the remediation task, remediation should be provided when at least one COA is poor. These evaluation criteria were given to participants for the extraction of low, mid, and high level data using the display.

Table 2. Evaluation Criteria

Data Level	Domain Data	Poor Performance	Good Performance
Low	COA performance	0-79%	80-100%
Mid	Learning Objective Performance	At least one COA Performance is between 0-79%	All COA Performances are equal or greater than 80%
High	TAO	Fail when at least one Learning Objective is at poor performance	Pass when all Learning Objectives are at good performances
Low	Remediation Task	Provide remediation when at least one COA performance is poor.	

Research Objectives

To empirically examine the effects of salience techniques on multi-level data extraction performance, two studies will be conducted. The first study will investigate the effects of color and alphanumeric values techniques mapped to low or mid-level data on the display. The second study will determine the effects of color and digital value techniques mapped to low and mid-level data on the display in different combinations. These two studies and their objectives are described below.

Study 1

The objective of the first study is to examine the effects of salience techniques applied to low or mid-levels of data on the extraction of multi-level data from a configural display. The first study extends the work of Bennett and colleagues (2001), who investigated the impact of different salience techniques applied to low-level data on the display. Their work, along with other studies, have applied salience techniques to displays representing the work domain of process control, which has two levels of data (Bennett et al., 2000; Bennett & Walters, 2001). This first study will apply color and alphanumeric values salience techniques to one of two levels of data (low- or mid-level data) instead of only one level and examine their impact on multi-level data extraction and instructor task performance. The color and alphanumeric values salience techniques have been shown to improve accuracy performance of extracting low-level data without hindering accuracy performance of divided attention tasks in displays where two data levels were mapped (Wickens & Andre, 1990; Hansen, 1995; Bennett et. al., 1993; Bennett et al., 2000; Bennett & Walters, 2001). The following hypotheses are categorized by the study measures.

Study 1 Data Extraction Performance Hypotheses

Hypothesis 1a: Both salience techniques mapped to low level data will improve response time and accuracy performance of extracting low-level data compared to the baseline display.

Hypothesis 1b: Alphanumeric techniques will improve response time and accuracy performance of low-level data extraction compared to color techniques.

Hypothesis 1c: Both salience techniques mapped to mid-level data will improve response time and accuracy performance of extracting mid-level data compared to the baseline display.

Hypothesis 1d: Color techniques will improve response time and accuracy performance of mid-level data extraction compared to alphanumeric techniques.

Study 1 Remediation Task Performance Hypotheses

When color or alphanumeric values salience techniques are applied to low level data, I hypothesize that:

Hypothesis 2a: Mapped salience techniques will improve remediation performance in response time and accuracy compared to the baseline display.

Hypothesis 2b: Color salience techniques will improve remediation performance in response time and accuracy compared to alphanumeric salience techniques.

The remediation task requires low-level data extraction; thus I do not expect differences in response time and accuracy performance between the color or alphanumeric value salience techniques when each are applied to mid-level data. This is because salience techniques applied to mid-level data does not map to the task constraints.

Study 1 Subjective Workload Hypotheses

Workload is of particular interest in the SBT instructor work domain and finding support for displays that reduce instructor workload is crucial. Therefore I hypothesize that:

Hypothesis 3: Salience techniques mapped to the display will reduce subjective workload compared to the baseline display.

Study 2

The objective of this study is to build upon the findings from study 1 and apply salience techniques to both low and mid-level data on the polar coordinate configural display. The combinations of color and alphanumeric values salience techniques applied to each level of data will be manipulated. Salience techniques applied to the display will be either: 1.) the same across both low- and mid-levels of data, or 2.) different across the two levels of data. Additionally, comparisons will be made across the two studies.

Study 2 Data Extraction Performance Hypotheses

Hypothesis 4a: Both salience techniques mapped to low and mid-levels data will improve response time and accuracy performance of multi-level data extraction compared to the baseline display.

Hypothesis 4b: Different salience techniques mapped to low and mid-level data will improve multi-level data extraction performance compared to same salience techniques

Hypothesis 4c: Mapping alphanumeric techniques to low level data and colors techniques to mid-level data will improve response time and accuracy performance of multi-level data extraction compared to other same and different combinations of salience techniques.

Study 2 Remediation Task Performance Hypotheses

Hypothesis 5a: Mapped salience techniques will improve remediation performance in response time and accuracy compared to the baseline display.

Hypothesis 5b: Remediation performance in response time and accuracy will improve when color techniques are mapped to low-level data, regardless of which salience technique is applied to mid-level data.

Study 2 Subjective Workload Hypotheses

Hypothesis 6a: Different salience techniques applied to the display will reduce subjective workload compared to when two of the same salience techniques are applied to the display.

Hypothesis 6b: When alphanumeric techniques are mapped to low level data and colors techniques to mid-level data, subjective workload will significantly reduce compare to other same and different salience combinations.

Across studies Hypotheses

Study 1 examined the application of salience techniques to either low or mid-level data on the display while Study 2 applied salience techniques across both low and mid-level data on the display (either the same technique to both levels or a combination of two techniques). However, are two salience techniques better than one? Due to the potential cost of configural displays, extracting mid-level data would be impacted by the salience of high-level data. Using the semantic mapping approach to address these potential performance costs, I hypothesize that:

Hypothesis 7a: Mapping salience techniques to two data levels will improve response time and accuracy performance of multi-level data extraction compared to mapping one salience techniques to the display.

Hypothesis 7b: Mapping different salience techniques to two levels of data will improve response time and accuracy performance of multi-level data extraction compared to mapping one salience techniques to the display.

Hypothesis 7c: Mapping alphanumeric techniques to low-level data and colors techniques to mid-level data will improve response time and accuracy performance of multi-level data extraction compared mapping one salience techniques to the display.

CHAPTER 3: STUDY 1

Study 1 Method

Study Design

A 2x3 mixed factorial design was used for study 1. The between-subjects variable was salience application (low and mid) and the within-subjects variable was salience technique (baseline, color, and alphanumeric values).

Study Conditions

To determine the effects of salience techniques on multi-level data extraction when applied to different levels of data, the following conditions shown in Table 3.

Table 3. Conditions for Study 1.

		Salience Technique Type		
		Baseline	Color	Alphanumeric Values
Salience Application	Low-Level	None	Color at low	Alphanumeric at low
	Mid-Level	None	Color at mid	Alphanumeric at mid

Participants

From previous studies that implemented the same salience techniques (such as Bennett et al., 2000; Bennett & Walters, 2001), the effect of these techniques is expected to be medium in size. For a medium effect size ($f = .25$), alpha at 0.05, power of .90, and 6 conditions in the experiment, the required sample size for study 1 (2 between x 3 within mixed model design) is 54 participants (Faul, Erdfreld, Buhner & Lang, 2009).

86 participants were recruited through the University of Central Florida SONA system and were given course credit for their participation in study 1. Six participants were removed due to experimenter and system error. 17 participants were removed because they did not obtained 80% accuracy or higher on the knowledge test. 63 participants met the knowledge test accuracy criteria and were included in the data analyses. Participants were randomly assigned to the low or mid salience application condition. 36 of the participants were female, and 27 participants were male. Ages range from 18-25 years old (*M*=18.27, *SD*=1.05). All participants had normal or correct-to-normal vision and color vision. None of the participants had any military or SBT experience. All participants were treated in adherence to the American Psychological Association (APA) guidelines.

Experimental Tasks and Materials

Testbed

A stimulus-response software (Open Sesame; Mathôt, Schreij, & Theeuwes, 2012) was used to present the polar coordinate configural display to the participant. The software also collected responses and recorded response time. The displays shown to participants included different salience techniques applied to data levels as specified by the experimental conditions stated in Table 3. In the baseline conditions (Figure 7), no added salience techniques (color and alphanumeric values) were applied to the configural display. Colors and alphanumeric techniques were applied to either low or mid-level data. Numeric values were used to the express the performance of process elements in the condition where alphanumeric techniques were applied to low-level data (Figure 8). As automated performance assessment is available in training systems, red and green colors were used to signify good and poor performances, respectively, when colors techniques were applied to low or mid-level data (see Figure 9 and Figure 11, respectively). Because the responses are categorical in conditions where alphanumeric techniques were applied to mid-level data, the symbols plus ("+") and minus ("-") were used to indicate good or poor performances (Figure 10).

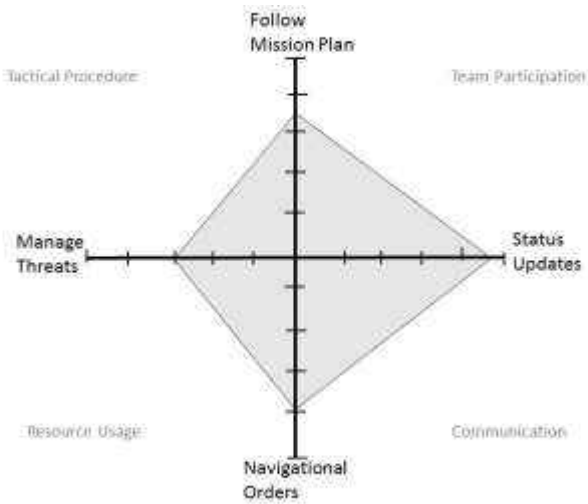


Figure 7. Baseline Condition

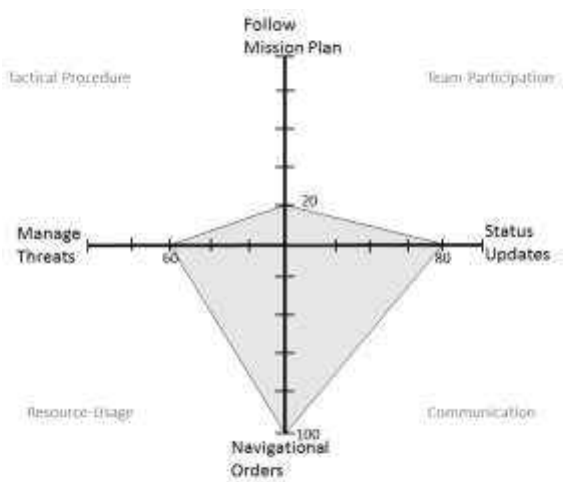


Figure 8. Alphanumeric values techniques applied to low-level data

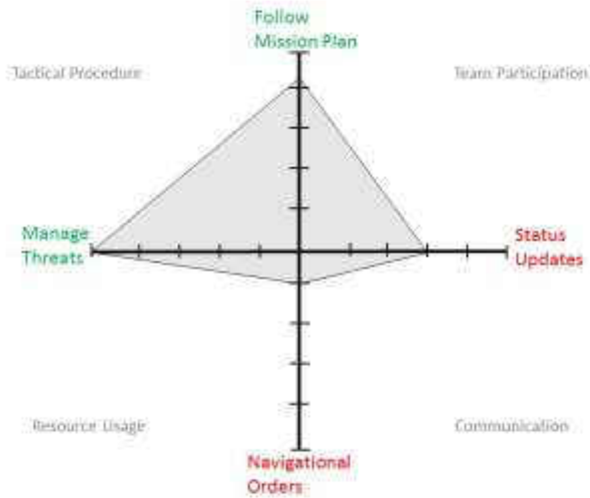


Figure 9. Color technique applied to low-level data

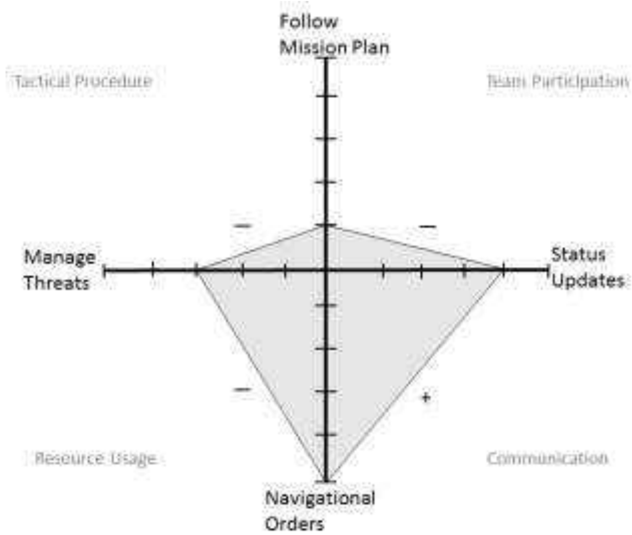


Figure 10. Alphanumeric values technique applied to mid-level data

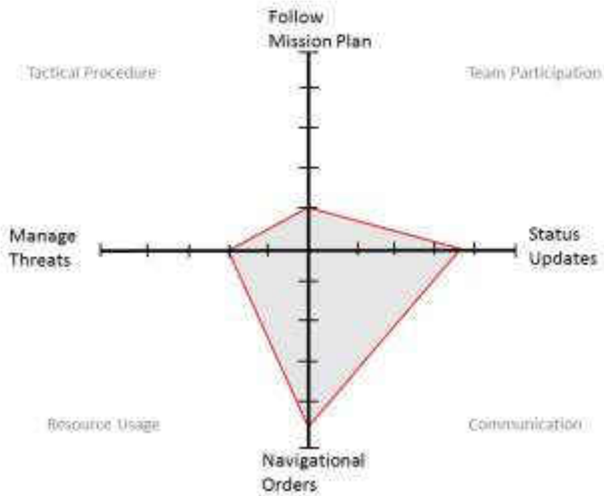


Figure 11. Color technique applied to mid-level data

Experimental Tasks

Participants were presented with selected displays and asked to answer questions relating to the display shown. For each condition, participants answered four types of questions related to specific tasks. For tasks involving extraction of low-level data, participants were asked questions about COA performance. For tasks involving extraction of mid-level data, participants were asked questions about learning objectives performance. For tasks involving extraction of high-level data, participants were asked questions about the overall performance. Additionally, for instructor-related tasks, participants were asked whether they wanted to provide remediation or not. Table 4 provides examples of the types of questions related to specific tasks and the possible responses.

Table 4. Task examples

Tasks	Question Examples	Responses available
Course of Action Performance (low-level data)	What is the performance of Manage Threats?	1) 0-79% 2) 80-100%
Learning Objectives Performance (mid-level data)	What is the performance of Communication?	1) Good Performance 2) Poor Performance
Overall Performance (high-level data)	What is the performance of the TAO?	1) Good Performance 2) Poor Performance
Feedback (instructor task)	Would you provide feedback?	1) Yes 2) No

<u>Performance Measures:</u>

Response time and accuracy were used as measures of performance for each of the four tasks. These parameters have been used in previous display research as a gauge for how well participants could understand and use the display (Bennett, Toms, & Woods, 1993; Bennett et al., 2000; Bennett & Walters, 2001; Holt, 2013).

<u>Subjective Measures</u>

Demographics questionnaire: A questionnaire containing queries about demographic information and computer/display history was provided to participants to assess their experience level with displays.

Workload Questionnaire: The NASA TLX was used to measure a participant's subjective workload. The NASA TLX has been found to be sensitive tool for measuring levels of workload (Hart & Staveland 1988; Nygren 1991; Hill et al. 1992), particularly when it is used in experimental settings (Byers et al., 1988; Hill et al., 1992). This tool allows participants to rate their perceived workload across 5 different dimensions (physical demand, temporal demand, performance, effort, and frustration). Previous IOS studies have shown that perceived workload has a direct impact on instructor task performance (Milham et. al., 2013a; Milham et. al., submitted). Thus, a participant's subjective workload was used to measure how the different displays contributed to his/her perceived workload.

Usability questionnaire: The Mouloua Usability Questionnaire (MUQ) was used to as a subjective measure of usability. Participants were asked to rate 27 items regarding the perceived usability of the display based on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). The questionnaire assessed perceived usability on 9 dimensions: simplicity, usefulness, functionality, consistency, proficiency, satisfaction, behavior, needs for improvement, and mental models.

<u>Individual Differences Measures:</u>

Individual differences may have potential influences to performance of when using displays and interfaces (Szalma, 2009). Therefore, the following individual differences measures were included to assess for their impact on performance.

LAG Task (*or n-back task)*: The LAG task from Steinhauser and Dehne (2014) was used to measure working memory capacity. This task was developed based on the working memory task described in Shelton, Metzger, and Elliott (2007) that used common 4-5 letter English words. Participants were presented with a list of words with each word appearing on the computer screen for one second. The list varied in length and participants were asked to recall either the last word presented (lag 0), the word before the last presented (lag 1), or two words before the last word presented (lag 2). A total of 24 lists were presented. Correct answers were weighted such that lag 2 words were given 3 points, lag 1 words were given 2 points, and lag 0 words were given 1 point for each correct response

Mental Rotation Test (*MRT):* A redrawn version of Vandenberg and Kuse (1978) Mental Rotation Task (MRT) was used to measure spatial ability (Peters et al., 1995). Participants were given 12 questions and were asked to identify which two out of four possible 3-dimensional geometric figures matched a target figure. Scores ranged from 1 to 24 and participants only received points if both figures were correctly identified.

Trail Making Test (*TMT*): The trail making test was used as a measure of cognitive flexibility (e.g. task shifting) and executive functioning. Participants were given parts A and B, but only part B was used in the analyses. Time to complete the trail was recorded.

Waterloo Gestalt Closure Task: The modified version of the Waterloo Gestalt Closure Task from Steinhauser and Dehne (2013) was used assess object recognition, a process which requires holistic processing. Participants were presented with either images of distorted objects or random lines for one second each. Participants were presented with 100 images and after each image, they were asked if they saw an object and to identify the object if they saw one. One point was given for every correctly identifying an object or non-object.

Procedure

Upon arrival to the laboratory, participants were given an informed consent form, screened for any color vision deficiencies, and then asked to complete the demographics questionnaire. Participants were then randomly assigned to one of the between-subjects conditions (low-level data or mid-level data). Participants were counterbalanced to complete either the individual differences tests or displays tasks first. For the individual differences tests, participants completed the MRT, TMT, LAG, Gestalt measures in counterbalancing order. For the displays tasks, participants were first provided with a 20 minute training on the SBT instructor domain, how to use the displays, and how complete the experimental tasks. After the training, participants complete the knowledge test to check their understanding of the instructional material. Participants were then given a practice session to familiarize themselves with and understand the display. Following the practice session, the participant completed three experimental conditions: the baseline block, a block with alphanumeric techniques, and a block with color techniques. All conditions were counterbalanced. Within each block, participants were asked to extract low, mid, or high level data and the remediation task. This set of questions was repeated 8 times within each block such that participants were given 32 trials per session with a particular display. In total, participants completed 96 trails for the 3 blocks. Between each block, participants completed the NASA TLX and MUQ pertaining to the block they had just finished. After completion of the displays tasks and individual differences tests, participants were debriefed and released.

Study 1 Results and Discussion

Random Assignment Checks

To check the effectiveness of random assignment, a series of t-tests were conducted using demographics information, individual differences, and accuracy and response time performance from the practice session and baseline as the dependent measure. No significant differences between groups were found in any of the dependent measures in Table 5. None of the participants had any military or SBT experience and therefore t-tests could not be completed for those measures.

Table 5. Random assignment checks for various dependent measures (study 1).

Variable	Mean Difference	Standard Error Difference	*t*	*p*
Gender	-.05	.13	-.36	.72
Age	.36	.26	1.36	.18
Handedness	-.002	.06	-.03	.97
Year in School	.10	.16	.66	.51
Computer Experience	.11	.17	.64	.54
Display Experience	.24	.36	.66	.51
Configural Display Experience	.21	.16	1.30	.20
Knowledge Test	-.27	1.8	-.15	.88
Practice Accuracy	-.07	.04	-1.87	.07
Practice Response Time	-340.54	358.33	-.95	.35
Baseline Accuracy	-.04	.03	-1.53	.13
Baseline Response Time	-413.97	230.71	-1.59	.12
MRT	-1.32	1.39	-.95	.35
LAG	-.01	.06	-.13	.85
TMT	-2.76	4.42	-.63	.53
Gestalt	-.84	1.70	-.49	.62
Military Experience	0	0		
SBT Experience	0	0		

Data Extraction Performance

Approach to hypotheses testing

The effects of salience type and data level salience application were examined using a 2x3 mixed model MANCOVA with the practice session as the covariates and the dependent variables of low-level data extraction, mid-level data extraction, high-level data extraction, and the remediation task. These dependent variables were shown to be moderately correlated, indicating that the use of the multivariate analysis is a good fit. Table 6 and Table 7 show the intercorrelations for accuracy and response time dependent measures.

Only the last 10 practice trails were included in practice accuracy and response time covariate. The practice session consisted of 32 trails. In order to determine if there were differences in performance between the beginning of the practice session and the later part of the practice session, the first ten trails were compared with the last ten trails in the practice session. Significant differences were found in both accuracy [$t(62)=-2.838$, $p=.006$] and response time [$t(62)=5.016$, $p<.001$]. Participants were more accurate in the last 10 trails (M=85.08%, SD=16.2%) of the practice session than the first 10 trails (M=79.21%, SD=17.2%). Additionally, participants had must faster response times in the last 10 trails (M=4110.90, SD=1420.78) compared to the first 10 trails (M=5747.13, SD=2057.09). The difference between the first 10 trails and last 10 trails of the practice session indicates that participants were becoming acclimated to the using the system during the first part and that training actually occurred in the later part.

The practice session accuracy and response time scores were significantly correlated to the dependent performance measures of low-level data extraction, mid-level data extraction, high-level data extraction, and the remediation task. The assumption of homogeneity of regression slopes was met for the two covariates. Additionally, there were no interactions between the covariates and the experimental manipulations.

Table 6. Intercorrelations for accuracy dependent measures by salience type and covariate

Accuracy Measures	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. Base Low	92.29	13.81	-											
2. Base Mid	90.61	11.52	.51**	-										
3. Base High	91.36	12.35	.58**	.42**	-									
4. Base Remediation	84.50	17.82	.53**	.62**	.56**	-								
5. Color Low	92.86	9.97	.34**	.33**	.35**	.42**	-							
6. Color Mid	92.26	10.63	.43**	.35**	.33**	.38**	.47**	-						
7. Color High	98.41	5.26	.28*	.38**	.47**	.24	.41**	.23	-					
8. Color Remediation	91.87	15.09	.09	.21	.34**	.38**	.33**	.17	.12	-				
9. Alpha Low	93.06	13.43	.56**	.37**	.38**	.31*	.53**	.50**	.41**	.06	-			
10. Alpha Mid	92.06	13.34	.54**	.46**	.50**	.48**	.57**	.56**	.43**	.33**	.67**	-		
11. Alpha High	93.45	10.73	.21	.42**	.37**	.45**	.38**	.26*	.48**	.23	.41**	.65**	-	
12. Alpha Remediation	89.88	13.63	.16	.25*	.05	.11	.20	.06	.05	.10	-.01	.23	.33**	-

* $p<.05$. **$p<.01$.

Table 7. Intercorrelations for response time dependent measures by salience type and covariate

Response Time Measures	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. Base Low	3843.02	1042.13	-											
2. Base Mid	4632.63	1559.40	.59**	-										
3. Base High	3497.36	1309.11	.60**	.75**	-									
4. Base Remediation	3147.41	932.53	.61**	.72**	.67**	-								
5. Color Low	4121.56	1442.23	.54**	.59**	.48**	.49**	-							
6. Color Mid	4093.13	1492.68	.43**	.76**	.67**	.58**	.63**	-						
7. Color High	2592.13	1070.65	.34**	.50**	.56**	.35**	.68**	.63**	-					
8. Color Remediation	2399.89	1035.52	.24	.54**	.45**	.44**	.72**	.73**	.80**	-				
9. Alpha Low	4433.83	1517.76	.52**	.64**	.53**	.52**	.56**	.61**	.42**	.40**	-			
10. Alpha Mid	4477.34	1344.76	.34**	.62**	.55**	.46**	.42**	.67**	.39**	.38**	.76**	-		
11. Alpha High	3043.33	1027.08	.45**	.65**	.70**	.64**	.47**	.55**	.55**	.39**	.56**	.59**	-	
12. Alpha Remediation	2923.41	893.32	.39**	.50**	.60**	.61**	.40**	.55**	.57**	.48**	.56**	.64**	.64**	-

* *p*<.05. ***p*<.01.

Hypotheses testing

A 2x3 mixed model MANCOVA was conducted to examine salience type and salience application to data levels on the dependent variables of low-level data extraction accuracy, mid-level data extraction accuracy, high-data level extraction accuracy, and remediation accuracy using the practice session accuracy as the covariate. The practice session accuracy covariate was found to be significant $F(4,57)=8.026$, $p<.001$, $\eta_p^2 = 0.360$. There was no significant interaction of salience type and application $(8,53)=1.014$, $p=.437$, $\eta_p^2 = 0.133$. Additionally, there was no significant main effect for the between-subjects condition of application $F(4,57)=1.88$, $p=.125$, $\eta_p^2 = 0.117$. There was a significant main effect for the within-subject variable of salience $F(8,53)=9.01$, $p<.001$, $\eta_p^2 = 0.576$. The accuracy means and standard deviations for all four dependent measures are listed in Table 8. Specific significance for each dependent measure are listed below in their respective section.

Table 8. Percent Correct Means and Standard Deviations for Low and Mid Conditions on Accuracy Performance Dependent Measures

Dependent Measures	Low Application Mean (SD) n=31	Mid Application Mean (SD) n=32	Total Mean (SD) N=63
Low-level Data Extraction Accuracy			
Base	90.15 (13.60)	94.36 (13.91)	92.29 (13.81)
Color	93.95 (11.12)	91.80 (8.76)	92.86 (9.97)
Alpha	93.86 (15.77)	92.19 (10.88)	93.06 (13.43)
Mid-level Data Extraction Accuracy			
Base	89.74 (11.54)	91.45 (11.63)	90.61 (11.52)
Color	90.32 (11.51)	94.14 (9.52)	92.26 (10.63)
Alpha	89.52 (16.80)	94.53 (8.36)	92.06 (13.34)
High-level Data Extraction Accuracy			
Base	89.11 (12.81)	93.53 (11.68)	91.35 (12.35)
Color	97.18 (7.00)	99.61 (2.21)	98.41 (5.26)
Alpha	90.73 (11.62)	96.09 (9.22)	93.45 (10.73)
Remediation Accuracy			
Base	80.88 (16.64)	88.02 (18.48)	84.50 (17.82)
Color	92.34 (11.03)	91.41 (18.36)	91.87 (15.09)
Alpha	89.92 (11.37)	89.84 (15.70)	89.88 (13.63)

*Note: Higher numbers indicate better accuracy performance

The mixed model MANCOVA was also used to examine salience type and application to data levels on the response times for the four dependent measures (low-level data extraction, mid-level data extraction, high-level data extraction, and the remediation task) with practice session response time as the covariate. The practice session response time covariate was significant $F(4,57)=15.101$, $p<.001$, $\eta_p^2=0.514$. There was no significant interaction of salience type and application $F(8,53)=.637$, $p=.743$, $\eta_p^2=0.088$ Additionally, there was no significant main effect for the between-subjects condition of application $F(4,57)=1.786$, $p=.144$, $\eta_p^2=0.111$. There was a significant main effect for the within-subject variable of salience $F(8,53)=2.716$, $p=.014$, $\eta_p^2=0.291$. The response time means and standard deviations for all four dependent measures are listed in Table 9. Specific significance for each dependent measure are listed below in their respective section.

Table 9. Means and Standard Deviations for Low and Mid Conditions on Response Time Performance Dependent Measures (millisecond)

Dependent Measures	Low Application Mean (SD) n=31	Mid Application Mean (SD) n=32	Total Mean (SD) N=63
Low-level Data Extraction Response Time			
Base	3694.01 (1167.46)	3987.37 (899.65)	3843.12 (1042.13)
Color	3662.85 (1172.11)	4565.95 (1554.44)	4121.56 (1442.23)
Alpha	4167.29 (1559.94)	4692.04 (1454.45)	4433.83 (1517.76)
Mid-level Data Extraction Response Time			
Base	4332.64 (1430.25)	4923.42 (1645.20)	4632.63 (1559.40)
Color	3806.91 (1166.93)	4370.39 (1725.44)	4093.13 (1492.68)
Alpha	4264.92 (1192.11)	4683.12 (1467.19)	4477.34 (1344.76)
High-level Data Extraction Response Time			
Base	3299.92 (1054.26)	3688.63 (1508.37)	3497.36 (1309.11)
Color	2416.61 (962.53)	2762.16 (115.58)	2592.13 (1070.65)
Alpha	2962.05 (966.91)	3122.06 (1091.81)	3043.33 (1027.08)
Remediation Task			
Base	2945.04 (938.03)	3343.44 (898.38)	3147.40 (932.53)
Color	2116.19 (710.34)	2674.73 (1223.86)	2399.89 (1035.52)
Alpha	2728.90 (732.06)	3111.85 (1001.46)	2923.41 (893.32)

*Note: Lower numbers indicate better response time performance

Low-level Data Extraction Performance

Accuracy. Univariate tests of the within-subjects variable of salience revealed significant differences for low-level data extraction accuracy $F(2,120)=5.33$, $p=.006$, $\eta_p^2=0.082$. For low-level data extraction accuracy, there were no significant contrasts between the different salience types even though the overall univariate test was significant. Trends of the analysis show that participants were generally more accuracy with the alphanumeric displays, followed by the color display, and least accurate with the baseline display for low-level data extraction (see Figure 12). Significant differences for the between subjects variable of application on low-level data extraction accuracy was not found, $F(1,60)=1.136$, $p=.291$, $\eta_p^2=0.019$. Additionally, there was not a significant interaction effect of salience type and application on low-level data extraction accuracy performance $F(2,120)=1.164$, $p=.316$, $\eta_p^2=0.019$.

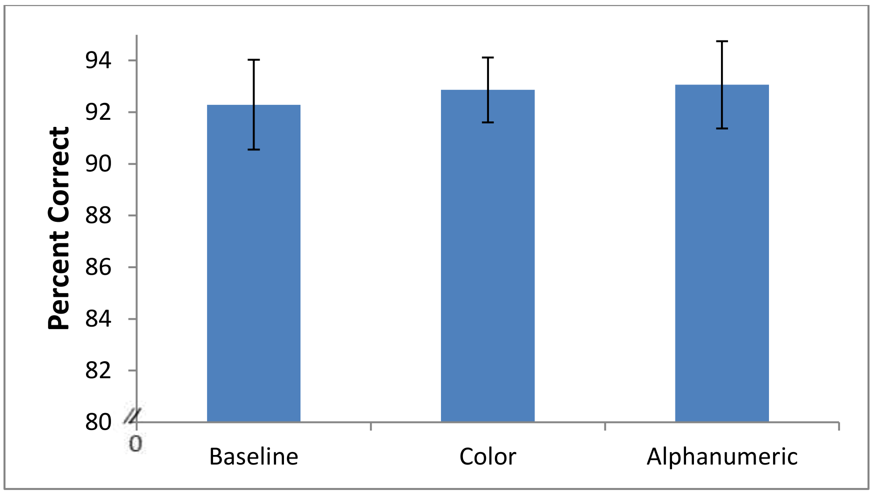


Figure 12. Low-level data accuracy performance.

Response time. Univariate tests of the within-subjects variable of salience revealed significant differences for low-level data extraction response time $F(2,120)=4.707$, $p=.011$, $\eta_p^2=0.073$. Contrasts for low-level data extraction response times revealed significant differences between the baseline (M=3843.12, SD=1042.13) and alphanumeric displays (M=4433.83, SD=1517.76). Participants had significantly faster response times using the baseline than the alphanumeric display ($p<.05$). Other contrasts for low-level data extraction response times were reaching significance, but not statistically significant. Figure 13 show the response time means for different salience type on low-level data extraction. Significant differences for the between subjects variable of application on low-level data extraction response time was not found,

$F(1,60)=3.572$, $p=.064$, $\eta_p^2=0.056$. Additionally, there was not a significant interaction effect of salience type and application on low-level data extraction response time performance $F(2,120)=2.026$, $p=.136$, $\eta_p^2=0.033$.

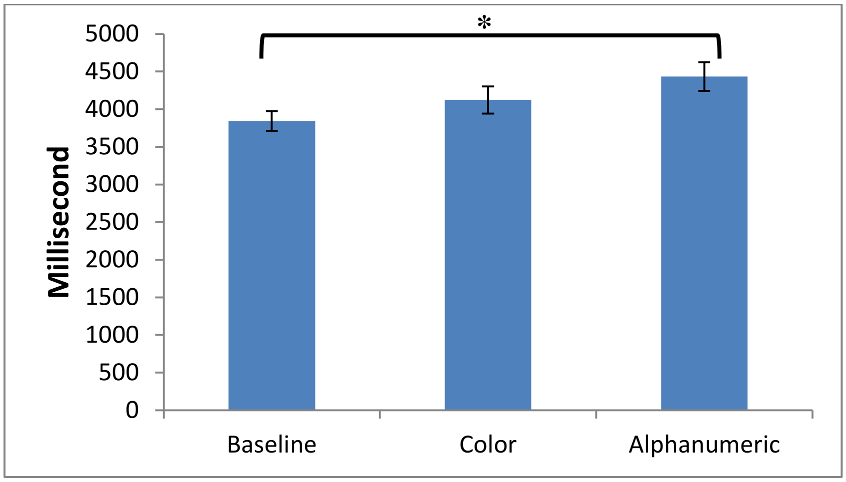


Figure 13. Low-level data extraction response times for different salience types.

Mid-level Data Extraction Performance

Accuracy. There was no significant differences for the within-subjects of salience type [$F(2,120)=.443$, $p=.643$, $\eta_p^2=.007$] or the between subjects variable of application [$F(1, 60) = 0.283$, p = 0597, $\eta_p^2=0.005$] on mid-level data extraction accuracy. Additionally, a significant interaction effect was not found for salience type and application on mid-level data extraction accuracy performance $F(2, 120)=0.467$, $p=0.628$, $\eta_p^2=0.008$.

Response time. There was no significant differences for the within-subjects of salience type [$F(2,120)=1.929$, $p=.150$, $\eta_p^2=.031$] or the between subjects variable of application [$F(1, 60) = 1.72$, p =0.195, $\eta_p^2=0.028$] on mid-level data extraction response time. Additionally, a significant interaction effect was not found for salience type and application on mid-level data extraction response time performance $F(2, 120)=0.166$, $p=0.831$, $\eta_p^2=0.003$.

High-level Data Extraction Performance

Accuracy. Univariate tests of the within-subjects variable of salience revealed significant differences for high-level data extraction accuracy $F(2,120)=13.891$, $p<.001$, $\eta_p^2 = .188$. For high-level data extraction, participants were significantly more accurate using the color displays (M=98.41%, SD=5.26) compared to the baseline (M=9135%, SD=12.35) and alphanumeric displays (M=93.45%, SD=10.73). See Figure 14 for a graphical representation of the high-level data extraction accuracy performance by salience type. Significant differences for the between subjects variable of application on high-level data extraction accuracy was not found, $F(1,60)=1.982$, $p=.164$, $\eta_p^2 = 0.032$. Additionally, there was not a significant interaction effect of salience type and application on high-level data extraction accuracy performance $F(2,120)=.467$, $p=.628$ $\eta_p^2 = 0.008$.

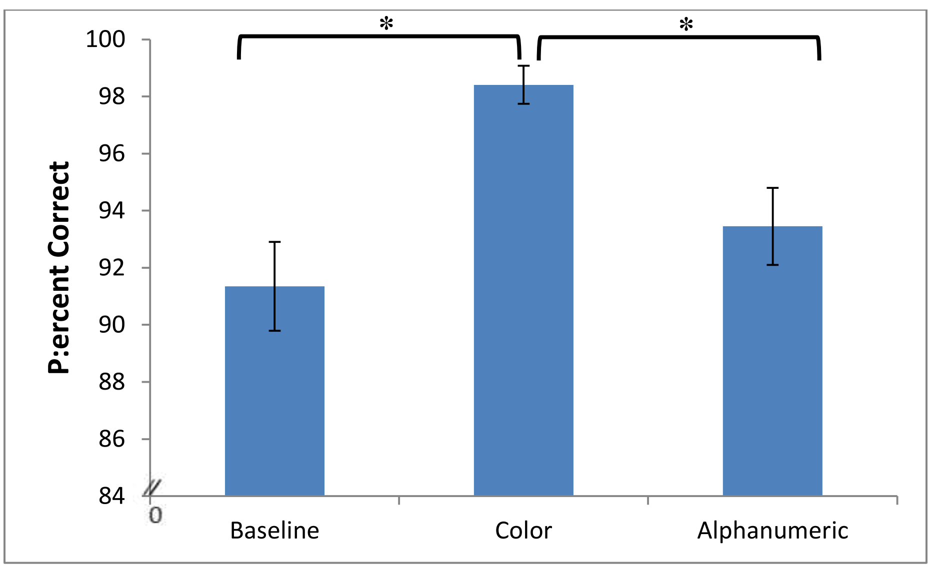


Figure 14. High-level data extraction accuracy performance by salience type

Response time. There was no significant differences for the within-subjects of salience type [$F(2,120)=.872$, $p=.416$, $\eta_p^2 = .014$] or the between subjects variable of application [$F(1,60)$ =

0.646, p = 0.425, η_p^2 =0.011] on high-level data extraction response time. Additionally, a significant interaction effect was not found for salience type and application on high-level data extraction response time performance $F(2, 120)=0.313$, $p=0.732$, η_p^2 =0.005.

Remediation Data Extraction Performance

Accuracy. Univariate tests of the within-subjects variable of salience revealed significant differences for the remediation task accuracy $F(2,120)=9.605$, $p<.001$, η_p^2 =.138. On the remediation task, participants were significantly more accurate using the color (*M*=91.87%, *SD*=15.09) and alphanumeric display (*M*=89.88%, *SD*=13.63) than to the baseline display (84.50%, *SD*=17.82). See Figure 15 for a graphical representation of the remediation task accuracy by salience type. Significant differences for the between subjects variable of application on high-level data extraction accuracy was not found, $F(1,60)=.005$, $p=.944$, η_p^2 =0.00. Additionally, there was not a significant interaction effect of salience type and application on high-level data extraction accuracy performance $F(2,120)=.497$, $p=.610$, η_p^2 =0.008.

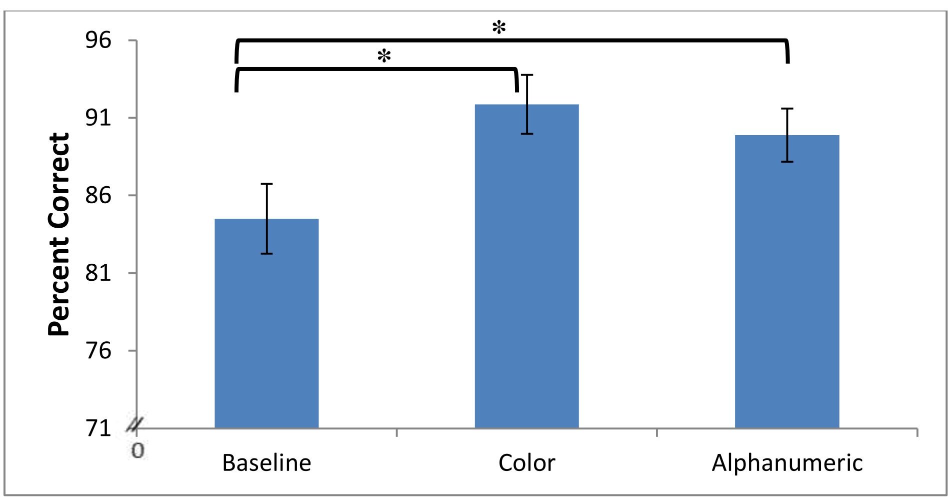


Figure 15. Remediation task accuracy by salience type.

Response time. There was no significant differences for the within-subjects of salience type [$F(2,120)=1.658$, $p=.195$, $\eta_p^2=.027$] on the remediation task response time. Additionally, a significant interaction effect was not found for salience type and application on the remediation task response time performance $F(2, 120)=0.327$, $p=0.722$, $\eta_p^2=0.005$. There was a significant difference the between subjects variable of application on mid-level data extraction response time [$F(1,60)=4.843$, p = 0.032, $\eta_p^2=0.075$], but the omnibus multivariate analysis was not significant so this result is not interpreted.

Subjective Workload

A 2 (salience application) x3 (salience type) mixed model MANOVA was conducted with the dependent measures of the six dimensions of the NASA TLX (mental, physical, frustration, temporal performance, and effort). A separate 2x3 mixed model ANOVA was conducted for the global workload measure. There was no significant omnibus effect of salience application on the workload dimension measures $F(7,56)=.752$, $p=.611$, $\eta_p^2=.075$. There was a significant omnibus effect of salience type on the workload dimension measures, $F(12,50)=1.90$, $p=.04$, $\eta_p^2=.313$. Univariate ANOVA contrasts revealed significant differences for mental [$F(2,122)=3.965$, $p=.021$, $\eta_p^2=.064$], performance [$F(2,122)=7.806$, $p=.001$, $\eta_p^2=.113$], and effort [$F(2,122)=5.483$, $p=.005$, $\eta_p^2=.082$] TLX dimensions. There was also a significant main effect for salience type for global workload $F(2,60)=8.0$, $p=.001$, $\eta_p^2=.210$.

Participants reported higher subjective mental workload for the baseline display compared to the color [$t(61)=2.64$. $p=.010$] and alphanumeric [$t(61)=2.05$, $p=.045$] displays (Figure 16). For the performance workload dimension, participants reported significantly lower performance when using the baseline display compared to the color [$t(61)=-3.92$, $p<.001$] and alphanumeric [$t(61)=-2.12$, $p=.038$] displays (Figure 17). For the effort workload dimension, participants reported significantly lower workload when using the color display compared to the baseline [$t(61)=-3.39$, $p<.001$] and alphanumeric [$t(61)=-2.07$, $p=.043$] displays (Figure 18). Additionally, participants reported significantly higher global workload when using the baseline displayed compared to the color [$t(61)=3.81$, $p<.001$] and alphanumeric [$t(61)=2.71$, $p=.009$] displays (Figure 19). Table 11 lists the means and standard deviations.

Table 10. Means and (standard deviations) for TLX measures by salience type (study 1).

Salience Type	Mental*	Physical	Temporal	Performance*	Effort*	Frustration	Global*
Baseline	34.68 (27.18)	14.92 (16.60)	26.19 (23.34)	79.05 (23.48)	36.75 (26.17)	24.92 (26.59)	26.59 (16.78)
Color	27.94 (23.13)	13.65 (16.37)	22.62 (21.96)	86.90 (17.35)	28.81 (24.34)	18.97 (20.89)	20.95 (14.19)
Alphanumeric	30.48 (24.59)	13.49 (14.30)	23.57 (19.72)	83.97 (20.36)	34.29 (27.79)	21.11 (20.51)	23.19 (14.68)

*Denotes significance at $p<.05$

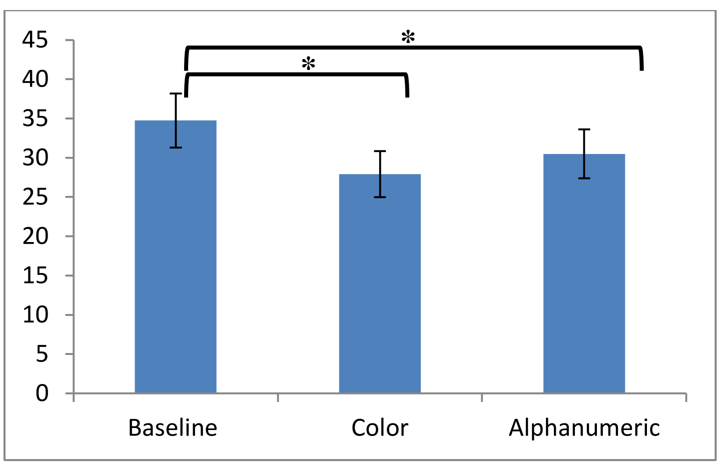


Figure 16. TLX mental workload for salience type.

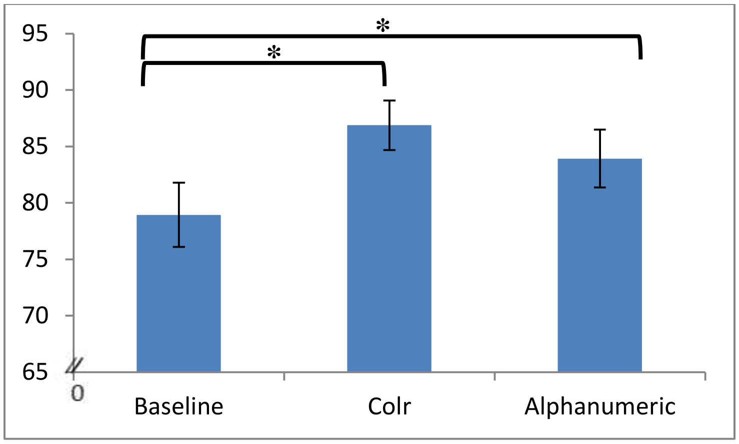


Figure 17. TLX performance for salience type.

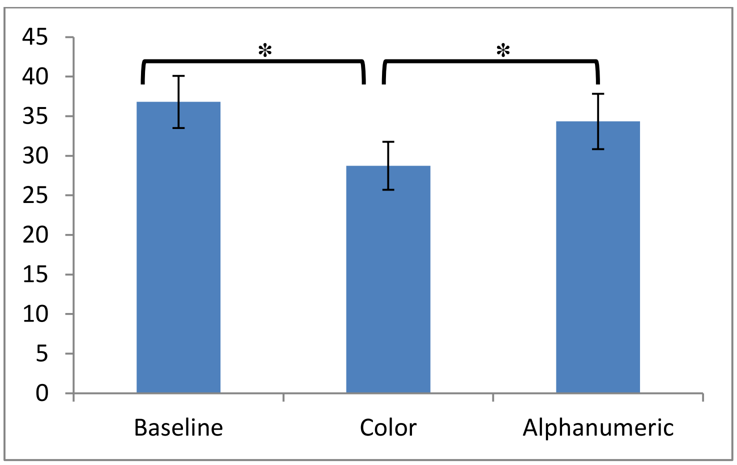


Figure 18. TLX Effort for salience type.

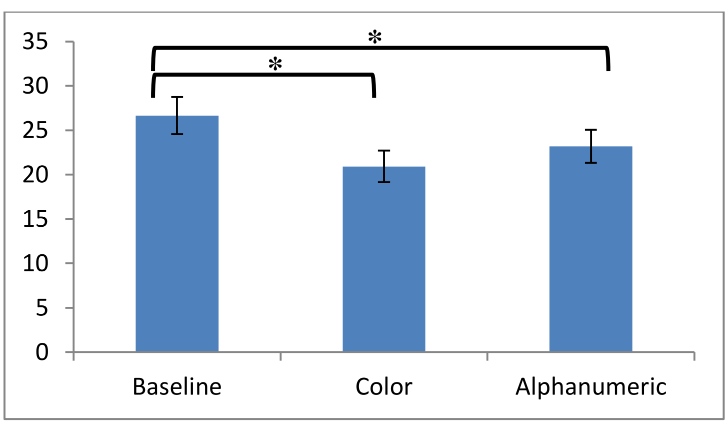


Figure 19. TLX Global for salience type.

There was a significant interaction effect of salience application and salience type on the workload dimension measures $F(12,50)=1.95$, $p=.030$, $\eta_p^2=.319$. Further analyses revealed significant interaction effects for the dependent measures of performance [$F(2,122)=3.469$, $p=.034$, $\eta_p^2=.054$] and effort [$F(2,122)=7.543$, $p=.001$, $\eta_p^2=.110$]. A significant interaction effect was also found for global workload, $F(2,60)=4.77$, $p=.012$, $\eta_p^2=.137$.

When salience techniques were applied to low level data, the baseline display resulted in lower TLX performance ratings compared to color [$t(61)=-4.52$, $p<.001$] and alphanumeric techniques [$t(61)=-2.64$, $p=.010$] applied to low-level data (Figure 20). Color applied to low-level data decreased TLX effort ratings compared to baseline [$t(61)=-4.84$, $p<.001$] and alphanumeric techniques [$t(61)=-3.63$, $p=.001$] applied to low-level data (Figure 21). Color techniques also significantly decreased global TLX workload compared to the baseline [$t(61)=-4.83$, $p<.001$] and alphanumeric techniques[$t(61)=-2.15$, $p=.036$] when applied to low-level data. Additionally participants rated alphanumeric techniques applied to low-level data to have significantly less global workload than the baseline display [$t(61)=-3.12$, $p=.003$] (Figure 22)..

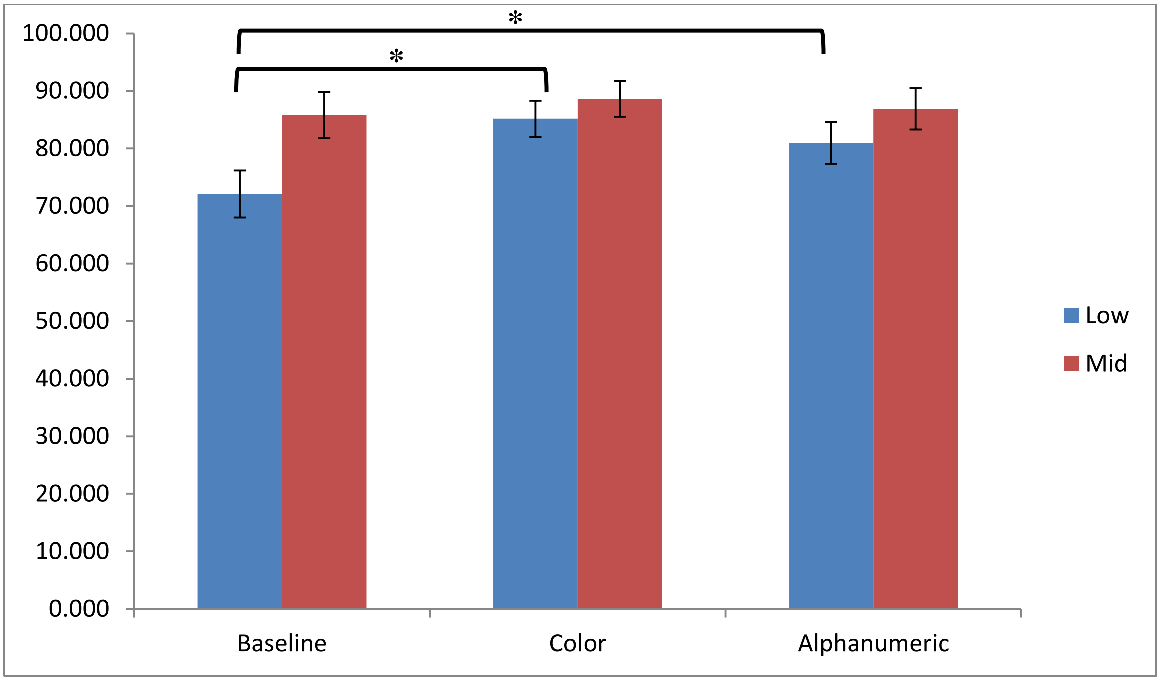


Figure 20. TLX performance ratings for salience type and salience application

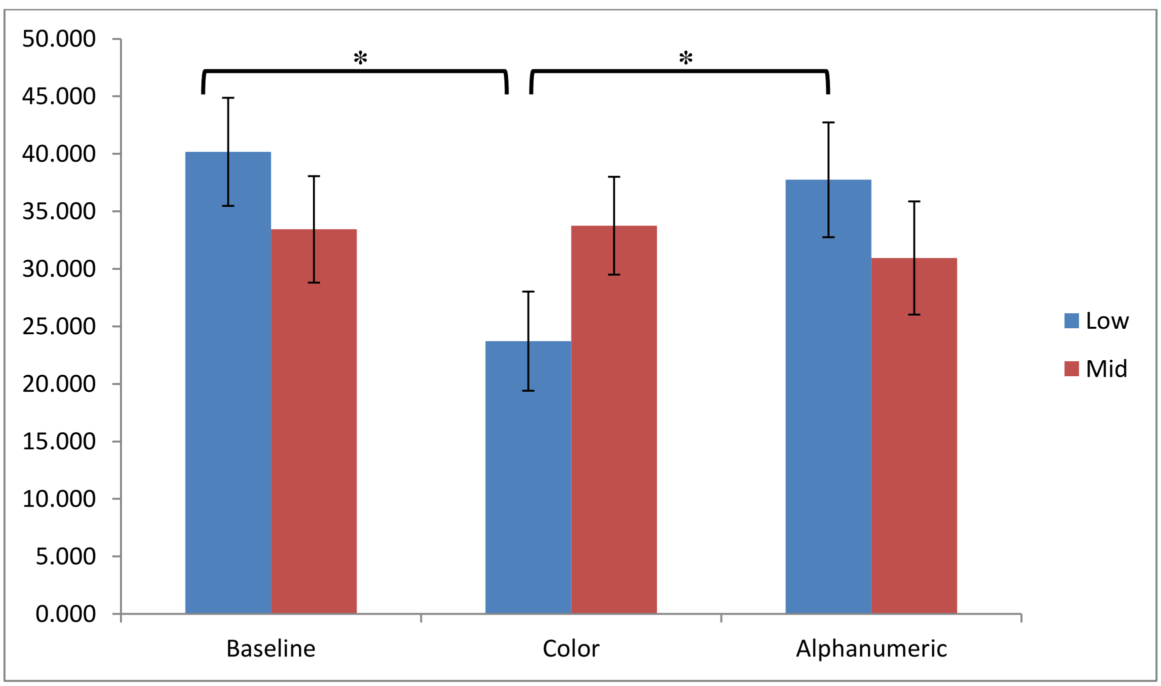


Figure 21. TLX effort ratings for salience type and salience application

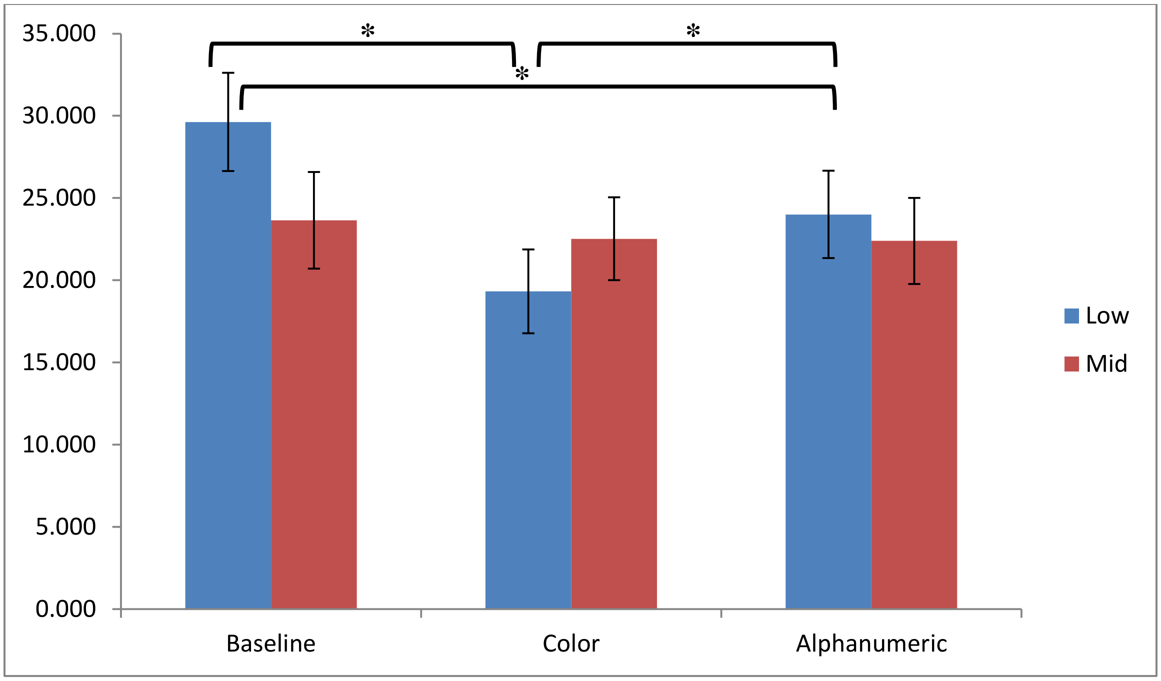


Figure 22. TLX global ratings for salience type and salience application

Usability

A 2x3 mixed model MANOVA was conducted to examine the effects of salience application and salience type on the dependent measures of the MUQ sub-dimensions (simplicity, usefulness, functionality, consistency, proficiency, satisfaction, behavior, improvement, and mental model). There was no significant interaction effect of salience application and salience type on the MUQ dimension $F(18,43)=1.481$, $p=.145$, $\eta_p^2=.383$. Additionally, there was no significant main effect of the between-subjects variable of salience application on MUQ measures [$F(9,52)=.489$, $p=.876$, $\eta_p^2=.078$], and no significant main effect of the within-subjects variable of salience type on the MUQ measures [$F(18,43)=.740$, $p=.752$, $\eta_p^2=.237$]. Due to the non-significant results of the multivariate tests, the subsequent univariate tests will not be discussed.

A separate 2x3 mixed model ANOVA was conducted to examine the global usability measure. There was no significant effect of salience type on global usability $F(2,60)=1.72$, $p=.188$, $\eta_p^2=.054$. Additionally, there was no significant main effect of salience application on global usability, $F(2,61)=.63$, $p=.43$, $\eta_p^2=.010$. There was a significant interaction effect on global usability, $F(2,60)=3.94$, $p=.025$, $\eta_p^2=.116$. Subsequent analyses revealed that for low application, when color was applied, usability was rated higher compared to the baseline [$t(61)=3.05$, $p=.003$] and alphanumeric values [$t(61)=2.06$, $p=.044$]. No significant differences were found between salience types when applied to mid-level data, $p>.05$.

Individual Differences

Individual differences measures of MRT, TMT, LAG, and Gestalt were examined to determine if they had any relationship to the performance of data extraction above and beyond the experimental factors and effects reported in the data extraction performance section. Each measure was evaluated using the General Linear Model separately. A 2(salience application) X 3(salience type) mixed MANCOVA with the practice session and individual differences measures as the covariates was computed for accuracy and response time performance for the dependent measures of low-, mid-, high-level data extraction, and the remediation task. Interactions were computed with overall performance scores.

Results of the analyses are listed in Table 11 and Table 12. If the multivariate result was not significant, the univariate result was not report or interpreted. For accuracy performance, the only individual differences measure found to have a significant impact was the TMT by salience type. TMT by salience type had significant differences on mid- and high-levels of data extraction, Subsequent correlations for both dependent measures were found to be negative indicating that those were faster on the TMT were more accurate on mid- and high-level data extraction, but only when using alphanumeric techniques. For response time performance, only TMT has significant impact for low- and high-level data extraction, as well as the remediation task. Participants who were faster on the TMT were also faster at low level data extraction, high level data extraction, and the remediation task.

Table 11. Study 1 Effects of individual differences on accuracy performance

Individual Differences Measure	Omnibus Multivariate Test	Performance Dependent Variables			
		Low	Mid	High	Remediation
MRT	F(4,53)=1.47, p=.225, η_p^2=.10	—	—	—	—
MRT x Application	F(4,53)=.311, p=.79, η_p^2=.031	—	—	—	—
MRT x Salience	F(8,49)=.801, p=.61, η_p^2=.116	—	—	—	—
MRT x Application x Salience	F(8,49)=1.38, p=.23, η_p^2=.185	—	—	—	—
TMT	F(4,54)=.92, p=.46, η_p^2=.064	—	—	—	—
TMT x Application	F(4,54)=.90, p=.47, η_p^2=.062	—	—	—	—
TMT x Salience	F(8,50)=2.81, p=.01, η_p^2=.310	F(2,114)=2.04, p=.13, η_p^2=.035	F(2,114)=5.58, p=.005, η_p^2=.089 Base: r=-.44, p=.73 Color: r=.03, p=.82 Alpha: r=-.33, p=.008	F(2,114)=3.33, p=.039, η_p^2=.055 Base: r=-13, p=.30 Color: r=-.80, p=.53 Alpha: r=-.34, p=.007	F(2,114)=1.08, p=.34, η_p^2=.019
TMT x Application x Salience	F(8,50)=.72, p=.67, η_p^2=.104	—	—	—	—
LAG	F(4,54)=.94, p=.45, η_p^2=.065	—	—	—	—
LAG x Application	F(4,54)=.72, p=.58, η_p^2=.050	—	—	—	—
LAG x Salience	F(8,50)=1.26, p=.29, η_p^2=.168	—	—	—	—
LAG x Application x Salience	F(8,50)=1.88, p=.08, η_p^2=.232	—	—	—	—
Gestalt	F(4,54)=.59, p=.67, η_p^2=.042	—	—	—	—
Gestalt x Application	F(4,54)=.99, p=.42, η_p^2=.068	—	—	—	—
Gestalt x Salience	F(8,50)=.44, p=.89, η_p^2=.066	—	—	—	—
Gestalt x Application x Salience	F(8,50)=1.11, p=.38, η_p^2=.150	—	—	—	—

Table 12. Study 1 Effects of individual differences on response time performance

Individual Differences Measure	Omnibus Multivariate Test	Performance Dependent Variables			
		Low	Mid	High	Remediation
MRT	$F(4,53)=.32$, $p=.86$, $\eta_p^2=.024$	—	—	—	—
MRT x Application	$F(4,53)=.69$, $p=.60$, $\eta_p^2=.050$	—	—	—	—
MRT x Salience	$F(8,49)=.54$, $p=.82$, $\eta_p^2=.081$	—	—	—	—
MRT x Application x Salience	$F(8,49)=.56$, $p=.82$, $\eta_p^2=.082$	—	—	—	—
TMT	$F(4,54)=2.61$, $p=.045$, $\eta_p^2=.162$	$F(1,57)=4.08$, $p=.048$, $\eta_p^2=.067$ r=.26. $p=.037$	$F(1,57)=1.24$, $p=.27$, $\eta_p^2=.021$	$F(1,57)=8.30$, $p=.006$, $\eta_p^2=.127$ r=.38, $p=.002$	$F(1,57)=6.64$, $p=.013$, $\eta_p^2=.104$ r=.35, $p=.005$
TMT x Application	$F(4,54)=1.73$, $p=.16$, $\eta_p^2=.113$	—	—	—	—
TMT x Salience	$F(8,50)=1.24$, $p=.30$, $\eta_p^2=.165$	—	—	—	—
TMT x Application x Salience	$F(8,52)=1.47$, $p=.19$, $\eta_p^2=.190$	—	—	—	—
LAG	$F(4,54)=2.21$, $p=.80$, $\eta_p^2=.141$	—	—	—	—
LAG x Application	$F(4,54)=.01$, $p=1.00$, $\eta_p^2=.001$	—	—	—	—
LAG x Salience	$F(8,50)=.52$, $p=.84$, $\eta_p^2=.077$	—	—	—	—
LAG x Application x Salience	$F(8,50)=.480$, $p=.86$, $\eta_p^2=.071$	—	—	—	—
Gestalt	$F(4,54)=.74$, $p=.57$, $\eta_p^2=.052$	—	—	—	—
Gestalt x Application	$F(4,54)=.48$, $p=.75$, $\eta_p^2=.034$	—	—	—	—
Gestalt x Salience	$F(8,50)=.74$, $p=.65$, $\eta_p^2=.106$	—	—	—	—
Gestalt x Application x Salience	$F(8,50)=.34$, $p=.95$, $\eta_p^2=.051$	—	—	—	—

Summary of Results

The first study examined the effects of salience type and salience application on data extraction performance. The first hypothesis was in regards to data extraction accuracy and response time performance. Hypothesis 1a predicted that salience techniques mapped to low level data would improve response time and accuracy performance of extracting low-level data compared to the baseline display. Additionally, hypothesis 1b predicted that alphanumeric techniques would improve response time and accuracy performance of low-level data extraction compared to color techniques. These were not supported. No interaction effects were found for accuracy or response time performance of low-level data extraction. Significant main effects were found salience type in the accuracy performance measure, and trends indicate that participants were generally more accurate using alphanumeric salience techniques. A significant main effect was also found for salience type in response time performance for low-level data extraction. Participants had faster response times when using the baseline display. Although main effects were not hypothesize, this finding is opposite of what was expected. However, even though participants were faster with the baseline display, they were not as accurate for low-level data extraction.

Hypothesis 4c and d tested the impacts of the experimental factors on mid-level data extraction performance. It was hypothesize that both salience techniques mapped to mid-level data would result in faster response time and higher accuracy performance of extracting mid-level data compared to the baseline display. More specifically, it was hypothesized that color techniques would result in faster response time and better accuracy performance of mid-level data extraction compared to alphanumeric techniques. These were not supported. No significant effects were found for mid-level data extraction in terms of accuracy and response time performance. Additionally, no main effects for salience type or salience application were found for accuracy and response time performance.

The second hypothesis served to examine the effects of the manipulations on an instructor task of knowing when to provide remediation. This task requires participants to not only extract information, but also to use that extracted information. Hypothesis 2a stated that mapped salience techniques would improve remediation performance in response time and accuracy compared to the baseline display. Hypothesis 2b further hypothesized that color salience techniques would improve remediation performance in response time and accuracy compared to alphanumeric salience techniques. Both of these hypotheses were partially supported. Accuracy scores for alphanumeric and color displays were higher than the baseline display. Additionally, participants were significantly more accurate using the color display compared to the other displays. Response times were not significantly different for salience type or application, and therefore that portion of the hypotheses was not supported.

Although no hypotheses were made for high-level extraction, some significant effects were found. For accuracy performance, participants were generally more accurate when using color techniques than the baseline and alphanumeric techniques for high-level data extraction. No significant effects for found for response time performance.

Hypothesis 3 predicted that salience techniques mapped to the display will reduce subjective workload compared to the baseline display. This was supported. TLX dimensions of mental, performance, and effort showed reduced workload for displays with mapped salience techniques. Global TLX workload ratings were also significantly lower for displays with mapped salience techniques compared to the baseline. Displays with color techniques mapped to low-level data resulted in lower effort and global TLX workload ratings.

Individual differences on the TMT were found to have an effect on data extraction performance. Those who were higher on executive function tend to have higher accuracy scores for mid- and high-level data extraction when using alphanumeric techniques. Additionally, those who had higher executive functioning had faster response times for low level data extraction, high level data extraction, and the remediation task. This significance did not change the overall findings in the performance scores.

Study 1 Discussion

In general, a majority of the performance related hypotheses were not supported. Some significant impact of salience application was found. There was a significant difference between salience types for high-level data extraction and the remediation task. Specifically, color had the most impact in improving accuracy of those tasks. Surprisingly, although participants rated the display where color techniques were mapped to low-level data to have the least amount of workload, this perception was not reflected in the performance sources. In all the performance dependent measures, a significant difference in salience application was not found. These results imply that although the salience application manipulation was different enough for participants to perceive at a subjective level, it was not different enough to impact objective performance measures. One possible explanation is that the salience application manipulations were too similar to each other for a difference to be found. Salience techniques applied to low-level data was perhaps too similar to salience techniques applied to mid-level data, and therefore did not have a significant impact on data extraction performance. When looking the results for low-level data extraction and mid-level data extraction, salience techniques applied to low-level data did improve accuracy performance for low level data extraction, compared to when they were applied to mid-level data. Additionally, salience techniques applied to mid-level data did improve accuracy performance for mid-level data extraction, compared to when salience techniques were applied to low-level data. However, these differences were not strong enough to show statistical significance. This implies that although the manipulations were different, they were not different enough from each other. Comparing these results (performance with one salience technique) to that of other salience application manipulations (performance with combination salience techniques) may show differences in performance. Application of one salience technique may be better or worse than application of combination salience techniques, but experimentation is needed to discovery this. The second study aims to study the impact of these combination salience techniques on performance.

CHAPTER 4: STUDY 2

The objective of study two is to expand upon the findings in study 1 to examine how salience type and salience application impact performance of data extraction. Significant main effects for salience type were found in Study 1, but no significant effects for salience application. This implies that the manipulation of salience application as not strong enough in Study 1. Study 2 investigates additional manipulations of salience application in which either the same salience techniques are applied to the display or different salience techniques area applied to the display.

Study 2 Method

Study Design

A 2x3 mixed factorial design was used for study2. The between-subjects variable was salience application (same and different) and the within-subjects variable was salience technique (baseline, color, and digital value).

Study Conditions

To determine the effects of salience techniques on multi-level data extraction when applied to different levels of data, the following conditions shown in Table 13 were used in Study 2.

Table 13. Conditions for Study 2.

		Salience Technique Type		
		Baseline	Color	Alphanumeric Values
Salience Application	Same	None	Color at low and Color at mid	Alphanumeric at low and Alphanumeric at mid
	Different	None	Color at low and Alphanumeric at mid	Alphanumeric at low and Color at mid

Participants

From previous studies that implemented the same salience techniques (such as Bennett et al., 2000; Bennett & Walters, 2001), the effect of these techniques is expected to be medium in size. For a medium effect size (f = .25), alpha at 0.05, power of .90, and 6 conditions in the experiment, the required sample size for study 2 (2 between x 3 within mixed model design) is 54 participants (Faul, Erdfreld, Buhner & Lang, 2009).

85 participants were recruited through the University of Central Florida SONA system and were given course credit for their participation in study 2. Seven participants were removed due to experimenter and system error. 13 participants were removed because they did not obtained 80% accuracy or higher on the knowledge test. 65 participants were included in the data analyses. Participants were randomly assigned to the same or different salience application condition. 43 of the participants were female, and 22 participants were male. Ages range from 18-23 years old (*M*=18.4, *SD*=.93). All participants had normal or correct-to-normal vision and color vision. None of the participants had any military or SBT experience. All participants were treated in adherence to the American Psychological Association (APA) guidelines.

Experimental Tasks and Materials

Testbed

The same stimulus-response software (Open Sesame; Mathôt, Schreij, & Theeuwes, 2012) used in study 1 was used to present the polar coordinate configural display to the participant in study 2. The software also collected responses and recorded response time. The displays shown to participants included different salience techniques applied to data levels as specified by the experimental conditions stated in Table 5. The second study employed a combination of the colors and alphanumeric values salience techniques. For each condition, salience techniques were applied to both low and mid-levels of data on the configural displays. Figures 23 through 26 depict how the colors and alphanumeric values techniques were applied to the low and mid-levels of data under each of the experimental conditions in Study 2.

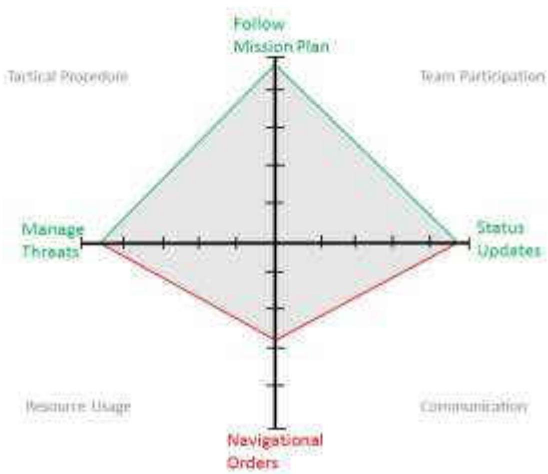


Figure 23. Color techniques applied to both low and mid-level data

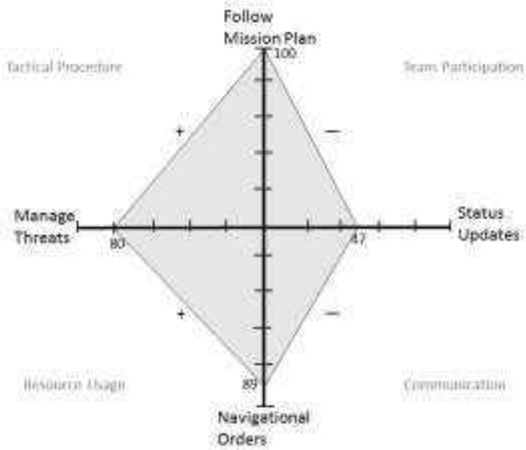


Figure 24. Alphanumeric values techniques applied to both low and mid-level data

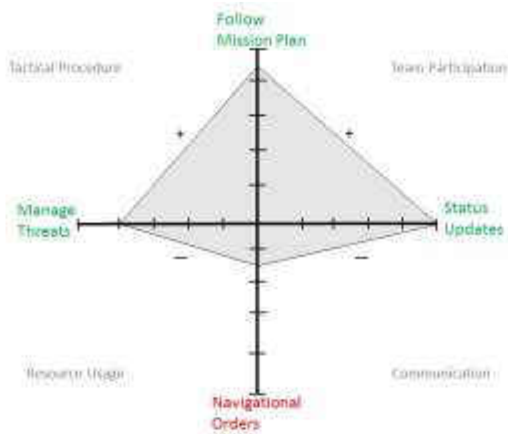


Figure 25. Color technique applied to low-level data and alphanumeric values technique applied to mid-level data

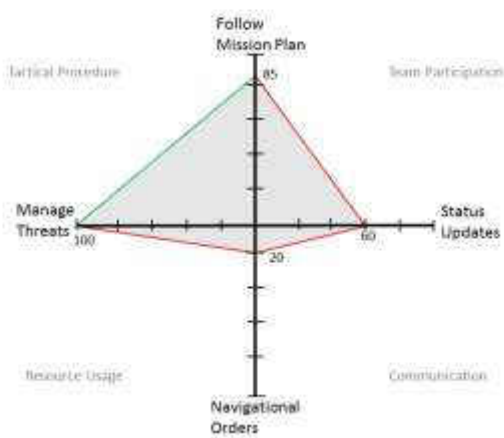


Figure 26. Alphanumeric values technique applied to low-level data and color technique applied to mid-level data

Experimental Tasks

The same experimental tasks from study 1 were also used in study 2. Table 4 lists the experimental tasks.

Subjective Measures

The same subjective measures from study 1 were used in study 2.

Performance Measures:

The same performance measures from study 1 were used in study 2.

Individual Differences Measures:

The same individual differences measures from study 1 were used in study 2.

Procedure

The same procedure used in study 1 was also used in study 2. The only exception was that participants were randomly assigned to the same or different condition (rather than low or mid condition).

Study 2 Results and Discussion

Random Assignment Checks

Checks for random assignment were also conducted on study 2 demographics. To check the effectiveness of random assignment, a series of t-tests were conducted using demographics information, individual differences, and accuracy and response time performance from the practice session and baseline as the dependent measure. No significant differences between groups were found in any of the dependent measures in Table 14. None of the participants had any military or SBT experience and therefore t-tests could not be completed for those measures.

Table 14. Random assignment checks for various dependent measures (study 2).

Variable	Mean Difference	Standard Error Difference	*t*	*p*
Gender	.07	.12	.60	..55
Age	.11	.23	.48	.64
Handedness	.004	.08	.05	..96
Year in School	-.04	.18	-.022	.83
Computer Experience	-.22	.15	-1.46	.15
Display Experience	.18	.11	1.53	.13
Configural Display Experience	-.14	.20	-.68	.50
Knowledge Test	-.46	.83	-.55	.58
Practice Accuracy	.03	.04	.77	.45
Practice Response Time	-462.51	292.51	-1.58	.12
Baseline Accuracy	.02	.02	.86	.39
Baseline Response Time	69.36	216.75	.32	.75
MRT	-2.04	1.41	-1.45	.15
LAG	-.01	.07	.84	.88
TMT	1.35	3.75	.36	.72
Gestalt	.81	1.88	.43	.67
Military Experience	0	0		
SBT Experience	0	0		

Data Extraction Performance

Approach to hypotheses testing

The same approach to hypotheses testing that was used in study 1 was used for study 2. The effects of salience type and data level salience application were examined using a 2x3 mixed model MANCOVA with the practice session as the covariates and the dependent variables of low-level data extraction, mid-level data extraction, high-level data extraction, and the remediation task. These dependent variables were shown to be moderately correlated, indicating that the use of the multivariate analysis is a good fit. Table 15 and Table 16 show the intercorrelations for accuracy and response time dependent measures.

The practice session accuracy and response time scores were significantly correlated to the dependent performance measures of low-level data extraction, mid-level data extraction, high-level data extraction, and the remediation task. The assumption of homogeneity of regression slopes was met for the two covariates. Additionally, there were no interactions between the covariates and the experimental manipulations.

Table 15. Intercorrelations for accuracy dependent measures by salience type and covariate

Accuracy Measures	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. Base Low	95.96	9.66	-											
2. Base Mid	89.30	11.76	.45**	-										
3. Base High	94.22	9.91	.65**	.54**	-									
4. Base Remediation	88.06	15.24	.38**	.52**	.57**	-								
5. Color Low	94.62	12.10	.44**	.35**	.43**	.28*	-							
6. Color Mid	95.00	11.22	.17	.34**	.35**	.20	.73**	-						
7. Color High	96.15	9.09	.27*	.30*	.51**	.30*	.54**	.67**	-					
8. Color Remediation	89.23	14.47	.16	.24	.41**	.30*	.36**	.54**	.59**	-				
9. Alpha Low	91.15	10.99	.19	.05	.13	.21	.32*	.25*	.19	.07	-			
10. Alpha Mid	97.12	6.54	.12	.17	.19	.04	.17	.37**	.22	.18	-.19	-		
11. Alpha High	98.46	5.19	.12	.36**	.25*	.26*	.22	.41**	.24	.23	.31*	.01	-	
12. Alpha Remediation	93.27	9.40	-.01	.18	.12	.22	-.07	-.02	.04	.03	-.26*	.04	.14	-

* $p<.05$. ** $p<.01$

Table 16. Intercorrelations for response time dependent measures by salience type and covariate

Response Time Measures	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. Base Low	4086.30	1028.22	-											
2. Base Mid	4629.98	1212.11	.44**	-										
3. Base High	3433.34	997.10	.46**	.64**	-									
4. Base Remediation	3173.77	1002.34	.54**	.64**	.67**	-								
5. Color Low	3919.77	1451.84	.61**	.37**	.39**	.42**	-							
6. Color Mid	4050.42	1599.31	.51**	.50**	.58**	.60**	.70**	-						
7. Color High	2424.40	1102.70	.43**	.44**	.54**	.49**	.53**	.74**	-					
8. Color Remediation	2360.14	1154.56	.34**	.32**	.55**	.39**	.50**	.63**	.69**	-				
9. Alpha Low	4196.85	1235.88	.47**	.59**	.51**	.61**	.51**	.42**	.42**	.36**	-			
10. Alpha Mid	4376.20	1292.99	.36**	.45**	.46**	.48**	.45**	.54**	.50**	.45**	.64**	-		
11. Alpha High	2801.48	1043.62	.33**	.47**	.49**	.45**	.39**	.38**	.25*	.12	.68**	.57**	-	
12. Alpha Remediation	2578.55	1071.27	.23	.49**	.36**	.46**	.37**	.25*	.22	.16	.71**	.60**	.76**	-

* $p<.05$. **$p<.01$.

Hypotheses Testing

The 2x3 mixed model MANCOVA examining salience type and salience application to data levels on the dependent variables of low-level data extraction accuracy, mid-level data extraction accuracy, high-data level extraction accuracy, and feedback accuracy used practice session accuracy as the covariate. The practice session accuracy covariate was found to be significant $F(4,59)=5.609$, $p=.001$, η_p^2 =0..276. There was no significant interaction of salience type and application (8,55)=1.196, $p=.318$, $\eta_p^2=0.146$. Additionally, there was no significant main effect for the between-subjects condition of application $F(4,59)=1.189$, $p=.325$, $\eta_p^2=0.075$. There was a significant main effect for the within-subject variable of salience $F(8,55)=7.012$, $p<.001$, $\eta_p^2=0.505$. The accuracy means and standard deviations for all four dependent measures are listed in Table 17. Specific significance for each dependent measure are listed below in their respective section.

Table 17. Percent Correct Means and Standard Deviations for Same and Different Conditions on Accuracy Performance Dependent Measures

Dependent Measures	Same Application Mean (SD) n=33	Different Application Mean (SD) n=32	Total Mean (SD) N=65
Low-level Data Extraction Accuracy			
Base	95.83 (10.67)	96.09 (8.66)	95.96 (9.66)
Color	92.05 (13.56)	97.27 (9.91)	94.62 (12.10)
Alpha	91.29 (12.30)	91.02 (9.65)	91.15 (10.99)
Mid-level Data Extraction Accuracy			
Base	89.66 (11.06)	88.93 (12.61)	89.30 (11.76)
Color	95.83 (9.20)	94.14 (13.08)	95.00 (11.22)
Alpha	97.35 (6.82)	96.88 (6.35)	97.11 (6.54)
High-level Data Extraction Accuracy			
Base	95.06 (10.81)	93.36 (8.97)	94.22 (9.91)
Color	96.59 (9.52)	95.70 (8.76)	96.15 (9.09)
Alpha	99.62 (2.18)	97.27 (6.91)	98.46 (5.19)
Remediation Task Accuracy			
Base	90.90 (15.08)	85.13 (15.08)	88.06 (15.24)
Color	90.53 (14.67)	87.89 (14.37)	89.23 (14.47)
Alpha	93.94 (9.44)	92.58 (9.45)	93.27 (9.40)

*Note: Higher numbers indicate better accuracy performance

The 2x3 mixed model MANCOVA was also used to examine salience type and application to data levels on the response times for the four dependent measures (low-level data extraction, mid-level data extraction, high-level data extraction, and the feedback task) with practice session response time as the covariate. The practice session response time covariate was significant $F(4,59)=11.391$, $p<.001$, $\eta_p^2=0.436$. There was a significant interaction of salience type and application $F(8,55)=6.119$, $p<.001$, $\eta_p^2=.0471$ There was also significant main effect for the between-subjects condition of application $F(4,59)=2.87$, $p=.031$, $\eta_p^2=0.163$. Finally, there was a significant main effect for the within-subject variable of salience $F(8,55)=2.405$, $p=.027$, $\eta_p^2=0.259$. The response time means and standard deviations for all four dependent measures are listed in Table 18. Specific significance for each dependent measure are listed below in their respective section.

Table 18. Means and Standard Deviations for Same and Different Conditions on Response Time Performance Dependent Measures (millisecond)

Dependent Measures (Response Time)	Same Application Mean (SD) n=33	Different Application Mean (SD) n=32	Total Mean (SD) N=65
Low-level Data Extraction			
Base	4181.89 (1123.69)	3987.72 (927.10)	4086.30 (1028.22)
Color	4230.88(1655.73)	3598.94 (1145.64)	3919.77 (1451.84)
Alpha	4487.41 (1362.16)	3897.22 (1027.36)	4196.85 (1235.88)
Mid-level Data Extraction			
Base	4726.81 (1156.97)	4530.12 (1277.18)	4629.98 (1212.11)
Color	4023.82 (1425.13)	4077.86 (1783.98)	4050.42 (1599.31)
Alpha	4654.61 (1155.77)	4089.10 (1380.22)	4376.20 (1292.99)
High-level Data Extraction			
Base	3376.61 (939.70)	3491.83 (1064.93)	3433.34 (997.10)
Color	2287.61 (832.75)	2565.46 (1324.40)	2424.40 (1102.70)
Alpha	3273.21 (1041.38)	2315.02 (804.93)	2801.48 (1043.62)
Remediation Task			
Base	3178.35 (998.67)	3169.05 (1022.07)	3173.77 (1002.34)
Color	2102.10 (760.65)	2626.24 (1418.02)	2360.14 (1154.56)
Alpha	3163.31 (1136.43)	1975.52 (544.76)	2578.55 (1071.27)

*Note: Lower numbers indicate better response time performance

Low-level Data Extraction Performance

Accuracy. There was no significant differences for the within-subjects of salience type [$F(2,124)=1.32$, $p=.271$, $\eta_p^2=.021$] or the between subjects variable of application [$F(1, 62) = 1.452$, $p = 0.233$, $\eta_p^2=0.023$] on low-level data extraction accuracy. Additionally, a significant interaction effect was not found for salience type and application on low-level data extraction accuracy performance $F(2, 124)=1.988$, $p=0.141$, $\eta_p^2=0.031$.

Response time. No significant interaction effect of salience type and application was found for low-level data extraction response time, $F(2,124)=1.84$, $p=.163$, $\eta_p^2=.029$. Additionally, there was no significant difference for the within-subjects variable of salience type for low-level data extraction response time, $F(2,124)=1.706$, $p=.186$, $\eta_p^2=.027$. Univariate tests of the between-subjects variable of application revealed significance for low-level data extraction response time $F(1, 62)=11.644$, $p=.001$, $\eta_p^2=.158$. Participants were significantly faster when using displays with difference salience application (M=3708.52, SD=1033.37) compared to displays with the same salience application (M=4415.88, SD= 1380.53) to low and mid data levels. See Figure 27 for a visual representation.

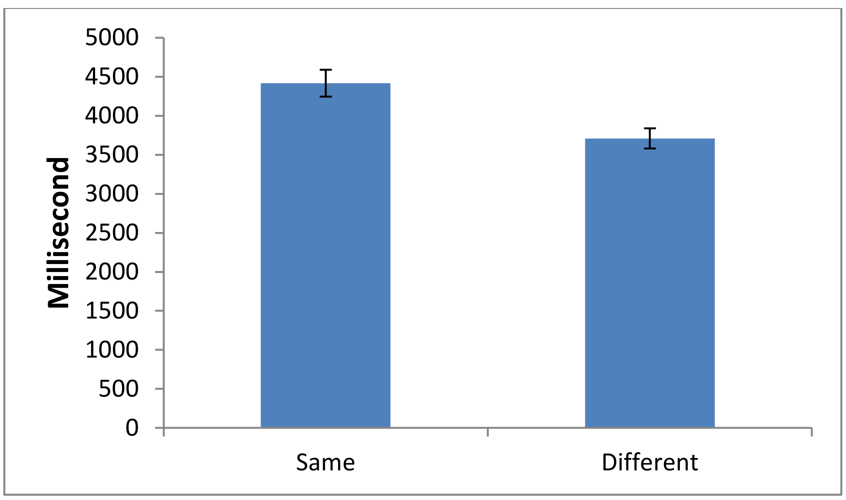


Figure 27. Low-level data extraction response time means for salience application condition

Mid-level Data Extraction Performance

Accuracy. Univariate tests of the within-subjects variable of salience revealed significant differences for mid-level data extraction accuracy $F(2,124)=6.817$, $p=.002$, $\eta_p^2=0.099$. Significant contrasts were found between the baseline display compared to the color and alphanumeric displays (p=.001). Participants were significantly more accurate in extracting mid-level data when using the color

(M=95.0%, SD=11.22) and alphanumeric display (97.11%, SD=6.54) than the baseline display (M=89.30%, SD=11.76). See Figure 28. Significant differences for the between subjects variable of application on mid-level data extraction accuracy was not found, $F(1,62)=.094$, $p=.761$, $\eta_p^2=0.002$. Additionally, there was not a significant interaction effect of salience type and application on mid-level data extraction accuracy performance $F(2,124)=.121$, $p=.887$, $\eta_p^2=0.019$.

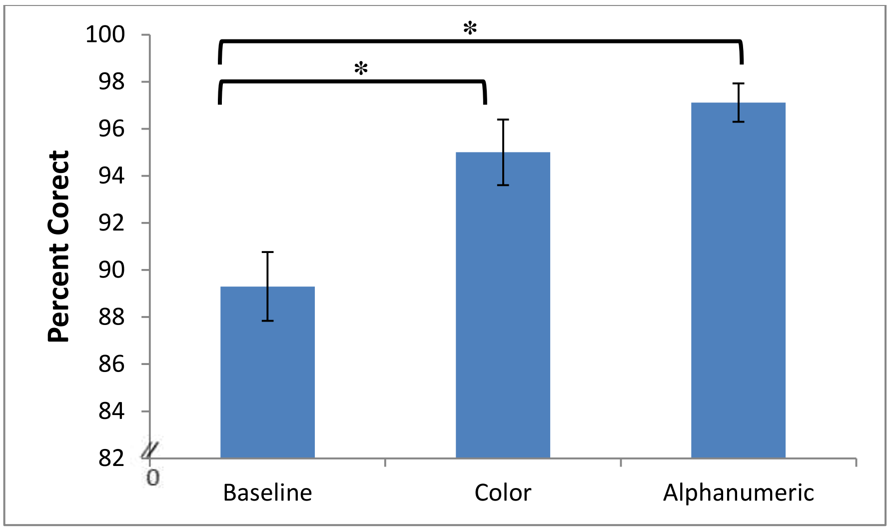


Figure 28. Mid-level data extraction accuracy performance by salience type.

Response time. No significant interaction effect of salience type and application was found for mid-level data extraction response time, $F(2,124)=1.087$, $p=.341$, $\eta_p^2=.017$. Additionally, there was no significant difference for the within-subjects variable of salience type for mid-level data extraction response time, $F(2,124)=2.55$ $p=.082$, $\eta_p^2=.040$. Univariate tests of the between-subjects variable of application revealed significance for mid-level data extraction response time $F(1,62)=4.834$, $p=.032$, $\eta_p^2=.072$. Participants were significantly faster when using displays with difference salience application (M=4097.24, SD=1480.46) compared to displays with the same salience application (M=4599.44, SD= 1246.96) to low and mid data levels, $p=.032$. See Figure 29 for a visual representation.

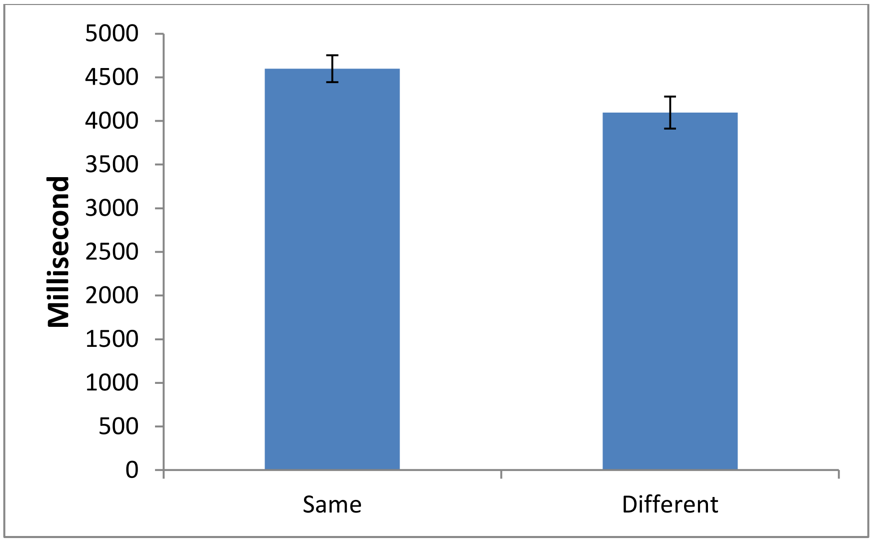


Figure 29. Response time means for mid-level data extraction of the application condition.

High-level Data Extraction Performance

Accuracy. Univariate tests of the within-subjects variable of salience revealed significant differences for high-level data extraction accuracy $F(2,124)=14.945$, $p<.001$, $\eta_p^2=.194$. Contrasts revealed significant differences between the baseline and alphanumeric displays, $p<.001$. Participants were significantly more accurate in high-level data extraction when using the alphanumeric display (M=98.46%, SD=9.09) compared to the baseline display (M=94.22%, SD=9.91). Figure 30 shows a graphical representation of the accuracy performance by salience type. Significant differences for the between subjects variable of application on high-level data extraction accuracy was not found, $F(1,62)=.606$, $p=.439$, $\eta_p^2=0.010$. Additionally, there was not a significant interaction effect of salience type and application on high-level data extraction accuracy performance $F(2,124)=.328$, $p=.721$, $\eta_p^2=0.005$. See Figure 30 for a visual representation.

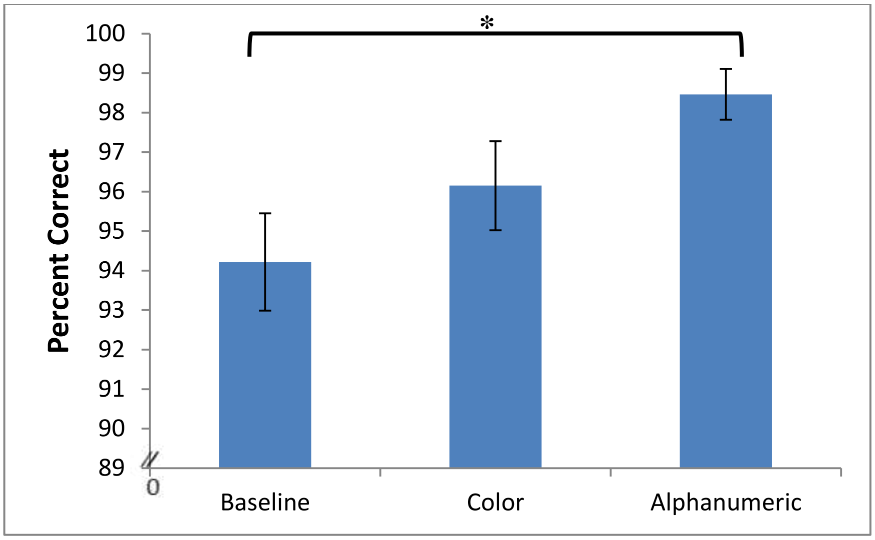


Figure 30. High-level data extraction accuracy performance by salience type.

Response time. Univariate tests of the within-subjects variable of salience revealed significant differences high-level data extraction response time $F(2,124)=6.371$, $p=.002$, $\eta_p^2=0.093$. Contrasts for high-level data extraction response times revealed significant differences between the baseline compared to the alphanumeric displays ($t(62)=8.01$, $p<.001$) and the color display, $t(62)=579$, $p<.001$. Additionally, the response times using the color display were significantly different from response times on the alphanumeric display. Participants had significantly faster response times using the color display ($M=2424.40$, $SD=1102.70$), followed by the alphanumeric display ($M=2801.48$, $SD=1043.62$), and the slowest response times using the baseline display ($M=3433.34$, $SD=997.10$) when extracting high-level data. See Figure 31 for a visual representation.

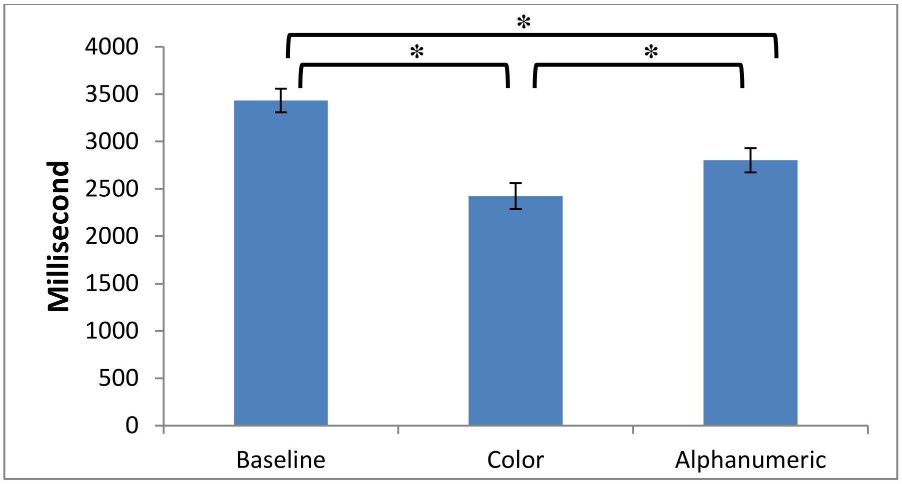


Figure 31. Response time means for high-level data extraction by salience type.

Univariate tests of the between-subjects variable of application revealed significance for high-level data extraction response time, $F(1,62)=5.421$, $p=.023$, $\eta_p^2=.080$. Participants were significantly faster at extracting high-level data with displays using different salience applications (M=2690.53, SD=1064.75) compared to displays using the same salience application (M=3076.35, SD=937.94). See Figure 32 for a visual representation.

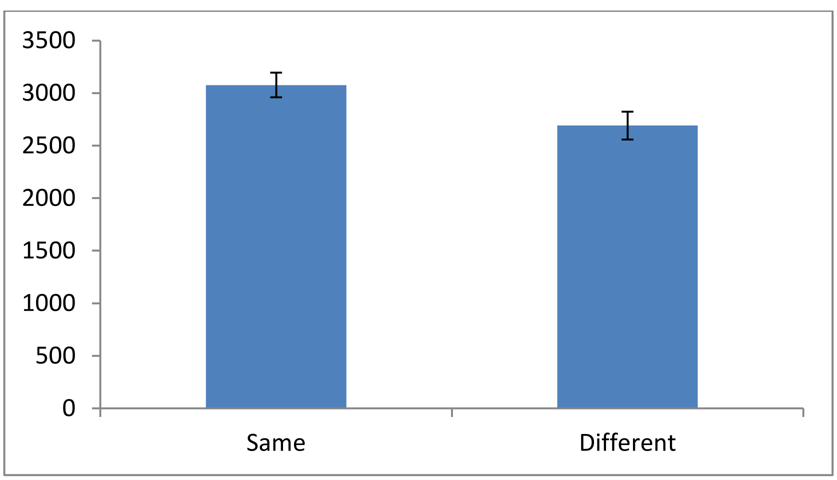


Figure 32. Response time means for high-level data extraction by salience application

There was a significant interaction of salience type and application for high-level data extraction response time, $F(2,124)=11.654$, $p<.001$, $\eta_p^2=.158$. Participants were significant faster using alphanumeric techniques applied in the different condition compared to alphanumeric techniques applied in the same condition [$t(62)=5.05$, $p<.001$], and the baseline [$t(62)= 7.25$, $p<.001$]. When same salience technique were applied to the display, participants had significantly faster response times for high-level data extraction using the color techniques compared to the baseline ($t(62)=-5.93$, $p<.001$) and alphanumeric techniques ($t(62)= -4.64$, $p<.001$). Figure 33 graphically represents these interaction effects.

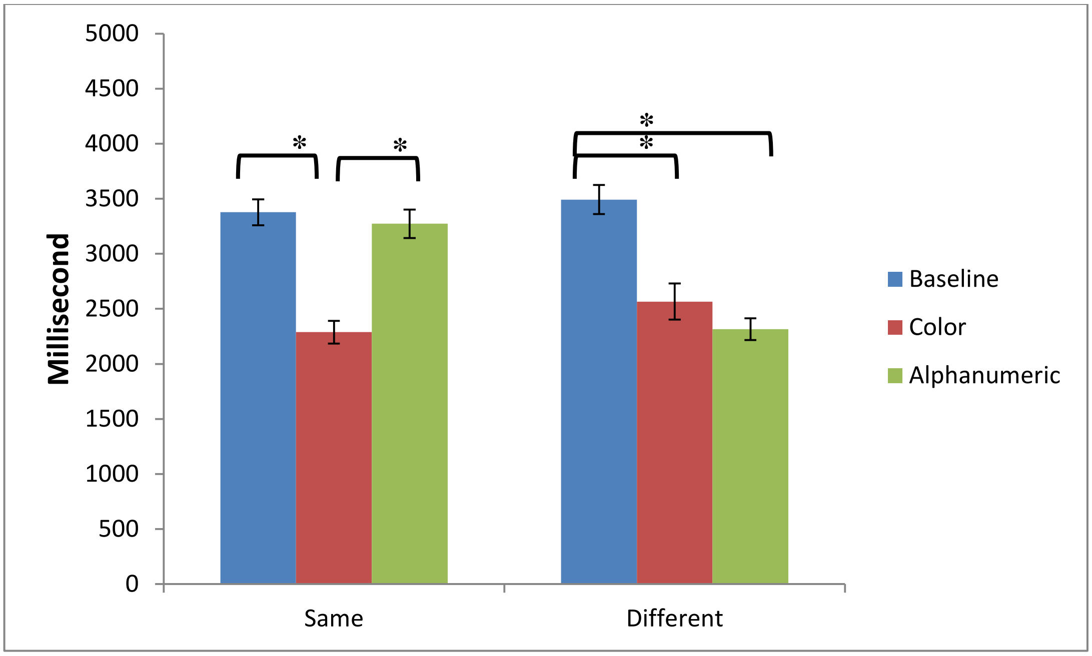


Figure 33. Response time means for high-level data extraction

Remediation Data Extraction Performance

Accuracy. Univariate tests of the within-subjects variable of salience revealed significant differences for the remediation task accuracy [$F(2,124)=7.416$, $p=.001$, $\eta_p^2=.107$]. Contrasts revealed significant differences between the baseline and alphanumeric display, $p=.007$. Participants were much more accurate in the remediation task when using the alphanumeric display (M=93.27%, SD=9.40) compared to the baseline display (M=88.06%, SD=15.24). See figure # for a graphical representation. Significant differences for the between subjects variable of application on the remediation task accuracy was not found, $F(1,62)=1.583$, $p=.213$, $\eta_p^2=0.025$. Additionally, there was not a significant interaction effect of salience type and application on the remediation task accuracy performance $F(2,124)=.340$, $p=.712$, $\eta_p^2=0.005$. See Figure 34 for a visual representation.

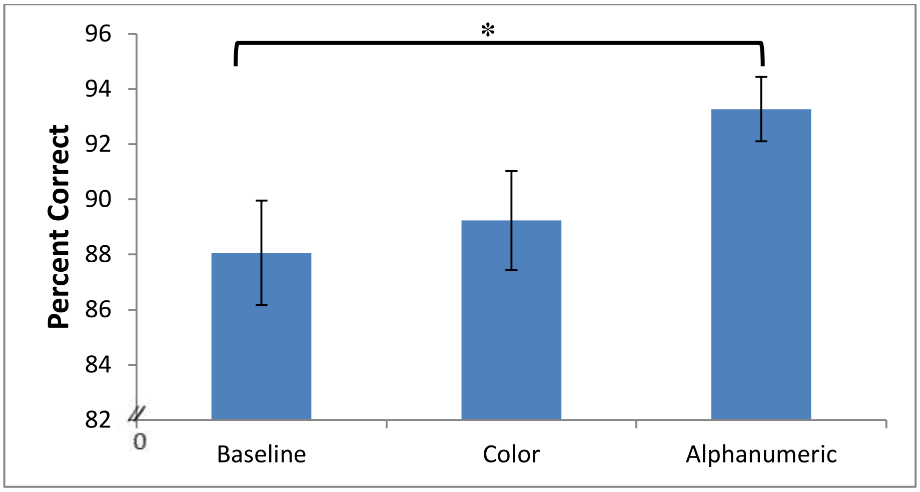


Figure 34. Remediation task accuracy by salience type.

Response time. . There was no significant differences for the within-subjects of salience type $F(2,124)=1.714$, $p=.184$, $\eta_p^2=.027$. Univariate tests of the between-subjects variable of application revealed significance for the remediation task response time $F(1,62)=5.577$, $p=.021$, $\eta_p^2=.083$. Participants were significantly faster using displays with different salience application (M=2501.21, SD=994.95) than display with the same salience application (M=2900.95, SD=961.92) for the remediation task. See Figure 35 for a visual representation.

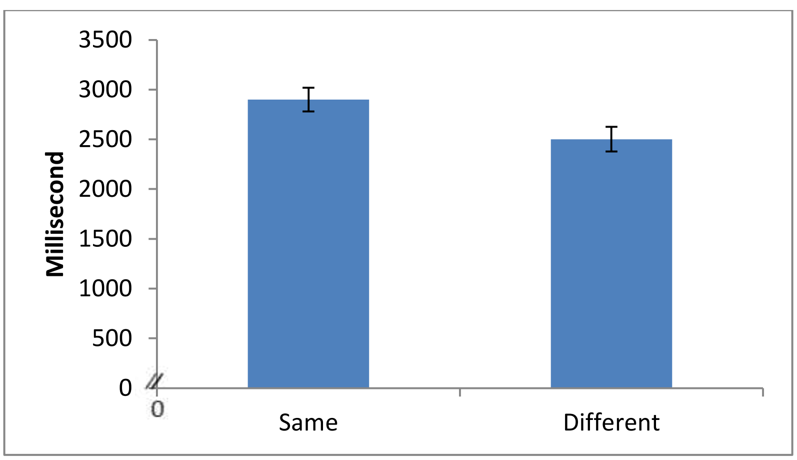


Figure 35. Remediation task response time means for same and different conditions

Significant interactions were found for feedback task response time, $F(2,124)=18.81$, $p<.001$, $\eta_p^2=.233$. Participants were significant faster using alphanumeric techniques applied in the different condition compared to alphanumeric techniques applied in the same condition [$t(62)=6.38$, $p<.001$], color techniques applied in the different condition [$t(62)=3.00$, $p=.004$], and the baseline [$t(62)= 7.22$, $p<.001$]. When same salience technique were applied to the display, participants had significantly faster response times for high-level data extraction using the color techniques compared to the baseline ($t(62)=-5.15$, $p<.001$) and alphanumeric techniques ($t(62)= -5.05$, $p<.001$). Participants had the fastest response time with the alphanumeric displays (M=1975.52, SD=544.76), followed by the color display (M=2626.24, SD=1418.02), and slowest response time with the baseline display (M=3169.05, SD=1022.07). Figure 36 graphically represents these interaction effects.

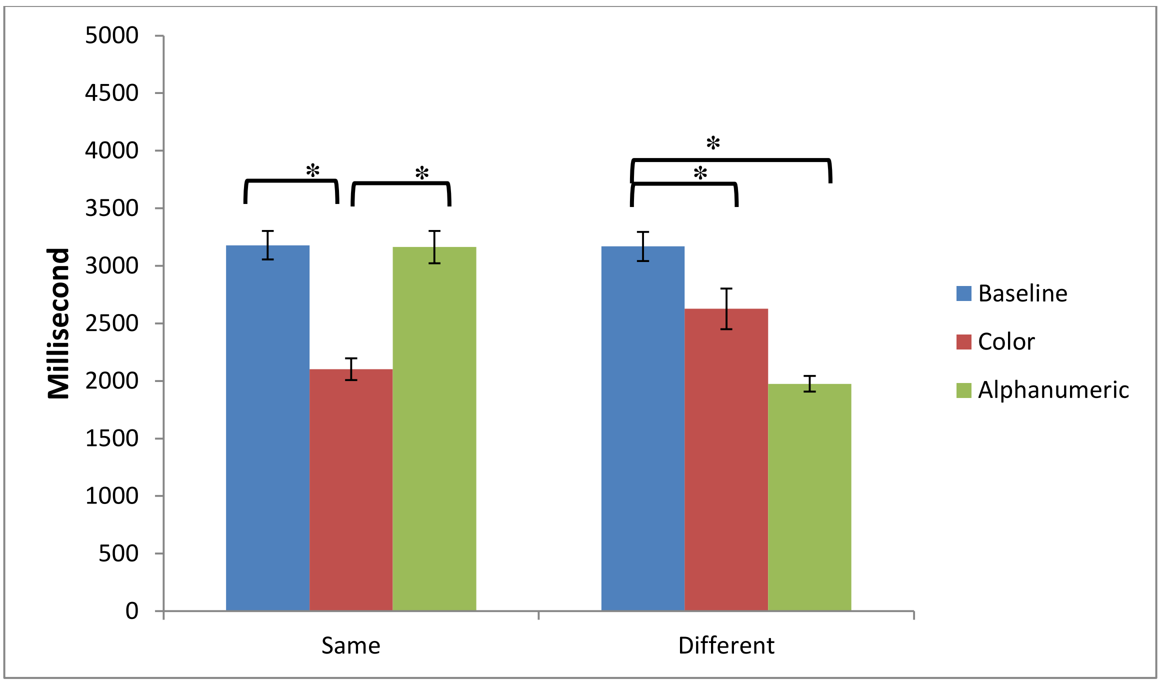


Figure 36. Remediation task response time means by salience type and application

Subjective Workload

A 2 (salience application) x3 (salience type) mixed model MANOVA conducted with the dependent measures of the six dimensions of the NASA TLX (mental, physical, frustration, temporal performance, and effort). A separate 2x3 mixed model ANOVA was used to examine the global workload measure. There was no significant interaction effect of salience application and salience type on the workload dimension measures [$F(12,52)=1.00$, $p=.462$, $\eta_p^2=.188$] or for global workload [$F(2,62)=.28$, $p=.76$, $\eta_p^2=.009$]. There was a significant effect of the within-subjects variable of salience type on the workload dimension measures [$F(12,52)=2.45$, $p=.013$, $\eta_p^2=.361$] for global workload, $F(2,62)=4.85$, $p=.011$, $\eta_p^2=.135$.

Mauchley's test of sphericity was statistically significant for the dependent measure of mental ($\chi2(2)=7.125$, $p=0.028$), frustration ($\chi2(2)=16.624$, $p=0.000$), physical ($\chi2(2)=17.41$, $p=0.000$), and performance ($\chi2(2)=6.593$, $p=0.037$). These significant results indicate that the assumption of sphericity had been violated. As such, the degrees of freedom of were adjusted using the Huynh-Feldt estimates of sphericity.

Univariate ANOVA contrasts of the within-subjects variable revealed significant differences for mental [$F(1.884,118.67)=4.87$, $p=.010$, $\eta_p^2=.072$], frustration [$F(1.68,105.97)=6.614$, $p=.003$, $\eta_p^2=.095$], performance [$F(1.897,119.53)=5.273$, $p=.007$, $\eta_p^2=.077$], and effort [$F(2,126)=4.686$, $p=.011$, $\eta_p^2=.069$].

Participants reported higher subjective mental workload for the baseline display compared to the color display [$t(63)=2.92$. $p=.005$] (Figure 37). Participants also reported high frustration workload when using the baseline display compared to the color [$t(63)=2.39$. $p=.020$] and alphanumeric [$t(63)=3.13$. $p=.003$] displays (Figure 38). For the performance workload dimension, participants reported significantly lower performance when using the baseline display compared to the color display [$t(63)$=-3.21, p=.002] (Figure 39). For the effort workload dimension, participants reported significantly higher effort when using the baseline display compared to the color [$t(63)$=2.33, p=.023] and alphanumeric [$t(63)$=2.81, p=.007] displays (Figure 40). Additionally, participants reported significantly higher global workload when using the baseline displayed compared to the color [$t(63)$=2.99, p=.004] and baseline [$t(63)$=2.88, p=.005] displays (Figure 41). Table 19 displays the means and standard deviations for each TLX dimension and the global workload by salience type.

Table 19. Means and (standard deviations) for TLX measures by salience type (study 2).

Salience Type	Mental*	Physical	Temporal	Performance*	Effort*	Frustration*	Global*
Baseline	28.31 (20.51)	13.15 (13.88)	21.92 (20.82)	80.54 (23.74)	33.00 (24.27)	23.00 (23.33)	23.22 (15.64)
Color	21.62 (18.84)	10.85 (10.59)	19.92 (19.95)	88.23 (17.86)	26.38 (22.46)	16.15(17.11)	17.94 (13.41)
Alphanumeric	24.00 (19.45)	9.92 (7.98)	20.77 (20.37)	84.92 (21.22)	25.85 (20.93)	14.08 (13.66)	18.33 (11.57)

*Denotes significance, $p<.05$

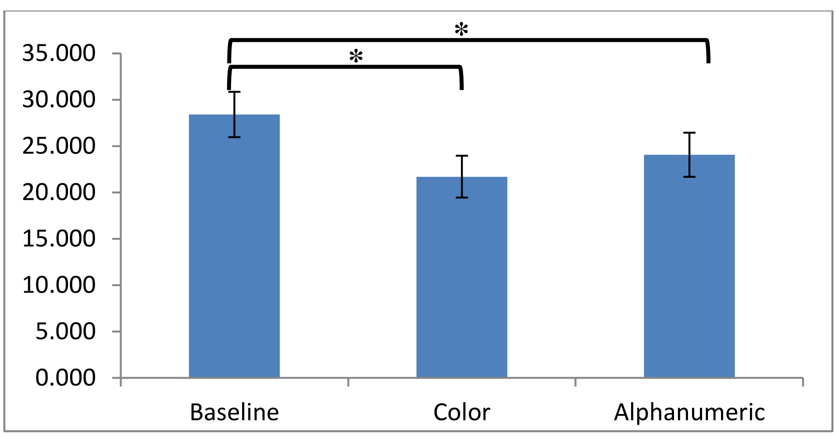


Figure 37. TLX mental workload means for salience type.

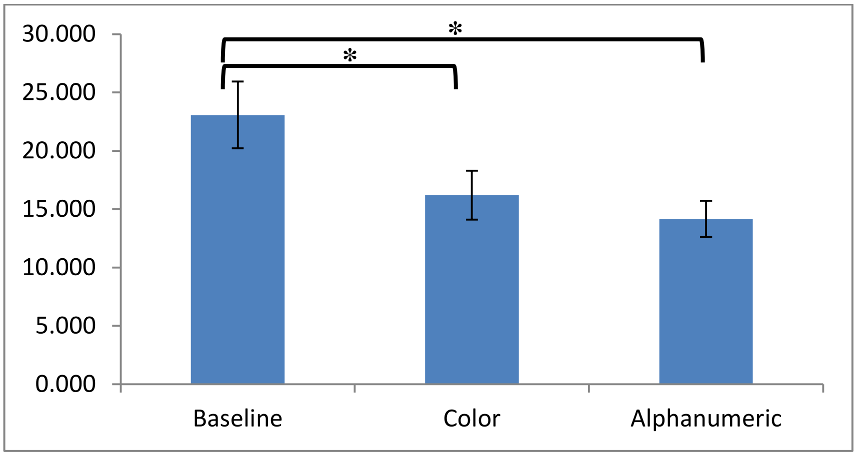


Figure 38. TLX frustration workload for salience type.

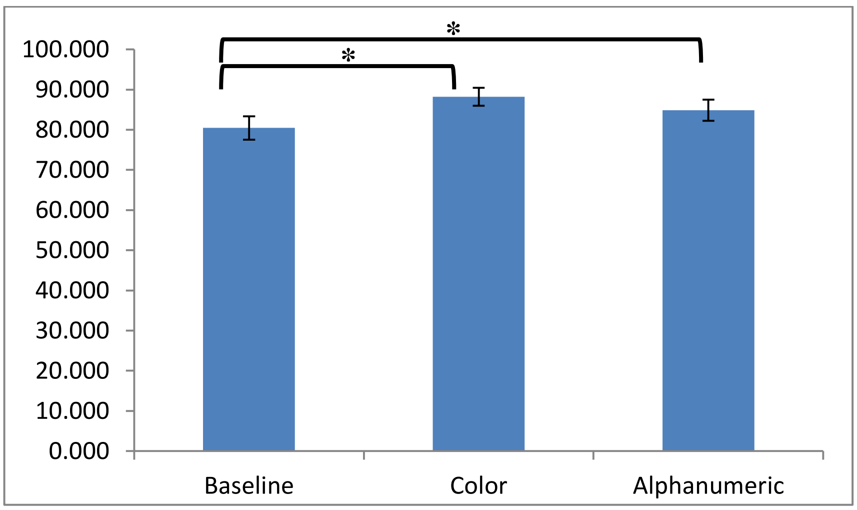


Figure 39. TLX performance for salience type.

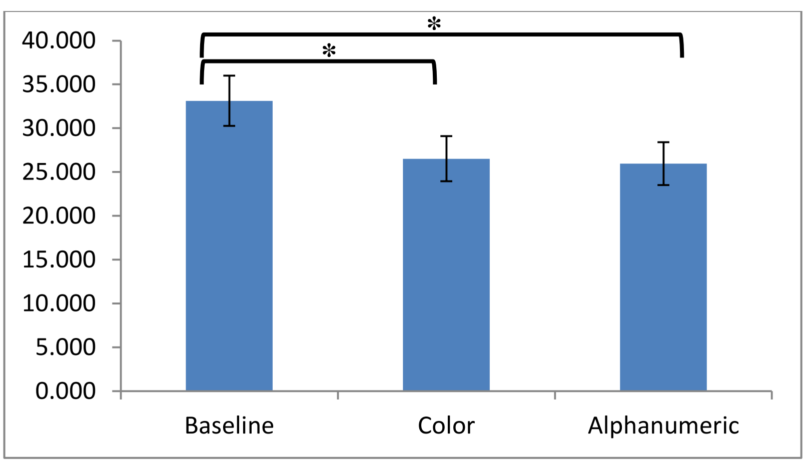


Figure 40. TLX effort workload for salience type.

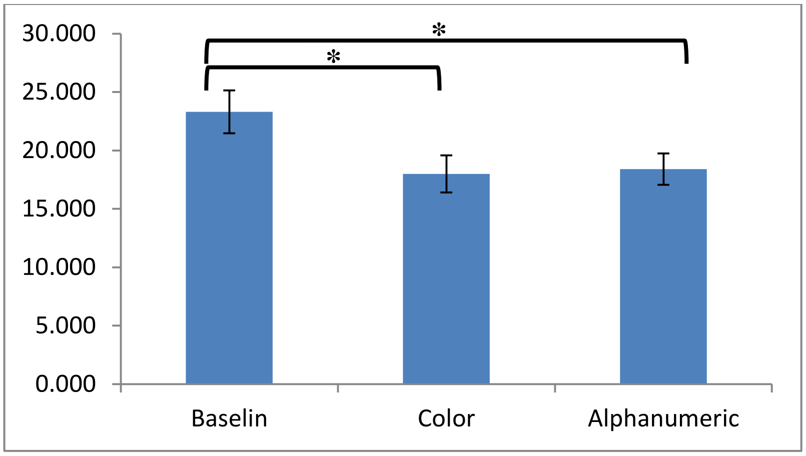


Figure 41. TLX global workload for salience type.

There was a significant effect of the between-subjects variable of salience application on the workload dimension measures [$F(6,58)=2.78$, $p=.019$, $\eta_p^2=.224$] and for the global workload [$F(1,63)=12.834$, $p=.001$, $\eta_p^2=.169$]. . Univariate ANOVA contrasts of the within-subjects variable revealed significant differences for mental [$F(1,63)=7.309$, $p=.009$, $\eta_p^2=.104$], frustration [$F(1,63)=6.953$, $p=.011$, $\eta_p^2=.099$], temporal [$F(1,63)=7.651$, $p=.007$, $\eta_p^2=.108$], effort [$F(1,63)=13.739$, $p<.001$ $\eta_p^2=.179$]. In general, participants reported higher ratings for the global workload and sub dimensions of mental, frustration, temporal, and effort dimensions when using different salience techniques mapped to the display. See Table 20 for the means and standard deviations for the TLX workload dependent measures.

Table 20. Means and standard deviations for TLX measures

TLX Measure	Same Application Mean (SD) n=33	Different Application Mean (SD) n=32
Mental*	19.34 (16.10)	30.10 (21.46)
Frustration*	13.38 (14.71)	22.24 (19.46)
Physical	10.30 (10.69)	12.34 (10.79)
Temporal*	14.65 (15.81)	27.29 (22.69)
Performance	87.22 (21.88)	81.82 (19.70)
Effort*	20.51 (17.17)	36.56 (24.68)
Global*	15.21 (11.47)	24.59 (13.96)

*Denotes significance, $p<.05$.

Usability

A 2x3 mixed model MANOVA was conducted to examine the effects of salience application and salience type on the MUQ sub-dimensions (simplicity, usefulness, functionality, consistency, proficiency, satisfaction, behavior, improvement, and mental model). A separate 2x3 mixed modal ANOVA was used to examine the global usability measure. There was no significant interaction effect of salience application and salience type on the MUQ dimensions [F(18,44)=.66, p=.834, η_p^2 =.211] or for global usability [F(2,60)=.36, p=.702, η_p^2 =.012]. Additionally, there was no significant effect of the between-subjects variable of salience application on MUQ dimensions [F(9,53)=.79, p=.613, η_p^2 =.118] . There was a significant main effect of application for global usability [F(1,61)=6.15, p=.016, η_p^2 =.092] in which same application (M=4.10, SD=.51) was rated higher than different application (M=3.85, SD=.58).

There was significant omnibus effect of the within-subjects variable of salience type on the MUQ measures, F(18,44)=2.95, p=.002, η_p^2 =.547. Univariate tests of the within-subjects variable of salience type revealed significant effects on Simplicity [F(10,52)=12.58, p<.001, η_p^2 =.171], Functionality [F(2,122)=2.13, p<.001, η_p^2 =.166], Usefulness [F(2,122)=14.53, p<.001, η_p^2 =.192], Proficiency F(2,122)=7.74, p<.001, η_p^2 =.113], Satisfaction [F(2,122)=5.40, p=.006, η_p^2 =.081], Behavior [F(2,122)=5.45, p=.005, η_p^2 =.082], Improvement [F(2,122)=8.77, p<.001, η_p^2 =.126], and Mental Model [F(2,122)=5.16, p=.007, η_p^2 =.078]. There was a significant main effect for the global usability score F(2,60)=10.94, $p < .001$, η_p^2 =.267.

Contrasts for the significant univariate tests for each dependent measure revealed that color and alphanumeric techniques received higher usability ratings compared to the baseline for the sub-dimensions of simplicity, functionality, usefulness, proficiency, satisfaction, behavior, improvement, mental model and the global usability score, $p < .05$ (see Figures 42-50 for visual representations).

Table 21 shows the means and standard deviations for the usability measures for each within-subjects manipulation.

Table 21. Means and standard deviations for usability dependent measures by salience type.

Usability Measure	Baseline Mean (SD)	Color Mean (SD)	Alphanumeric Mean (SD)
Simplicity*	3.91 (0.81)	4.41 (0.73)	4.33 (0.57)
Functionality*	3.88 (0.60)	4.29 (0.60)	4.20 (0.53)
Usefulness*	3.57 (0.85)	4.11 (0.79)	4.07 (0.65)
Consistency	4.13 (0.54)	4.28 (0.74)	4.24 (0.53)
Proficiency*	3.95 (0.66)	4.25 (0.62)	4.13 (0.58)
Satisfaction*	3.39 (0.91)	3.74 (0.81)	3.71 (0.63)
Behavior*	3.86 (0.82)	4.17 (0.77)	4.19 (0.57)
Improvement*	3.24 (1.05)	3.84 (0.96)	3.68 (0.81)
Mental Model*	3.79 (0.73)	4.08 (0.70)	3.75 (0.61)
Global Usability*	3.75 (0.61)	4.13 (0.62)	4.06 (0.45)

*Denotes significance at $p<.05$.

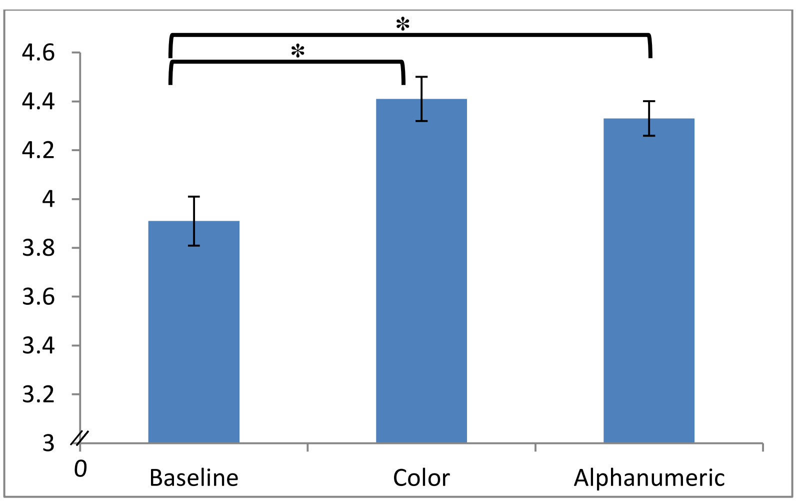


Figure 42. Simplicity by salience type.

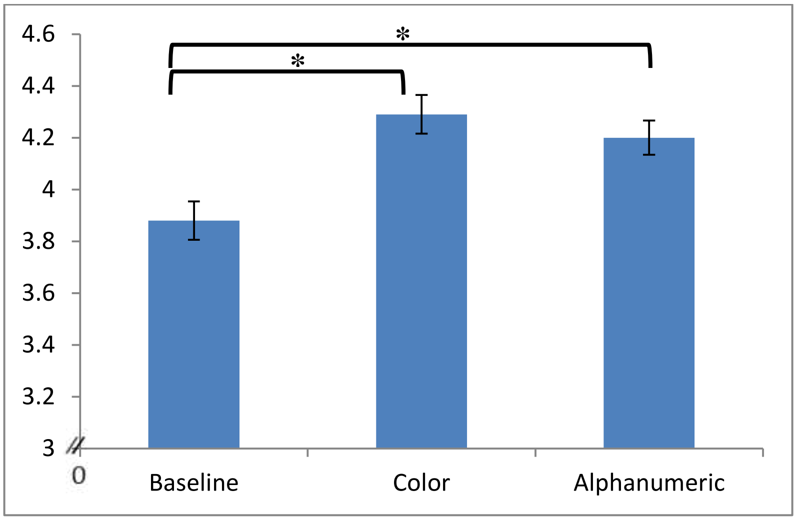


Figure 43. Functionality by salience type

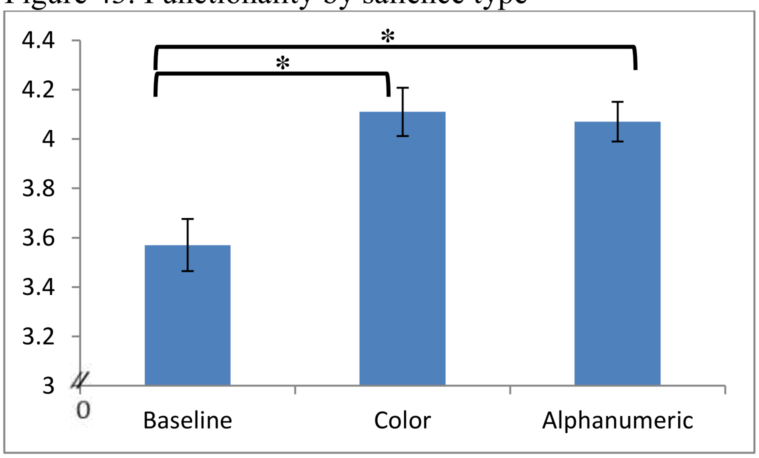


Figure 44. Usefulness by salience type.

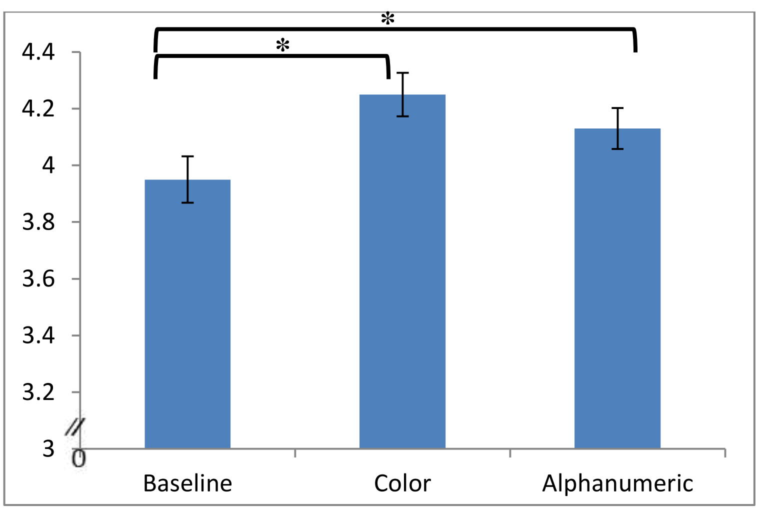


Figure 45. Proficiency by salience type.

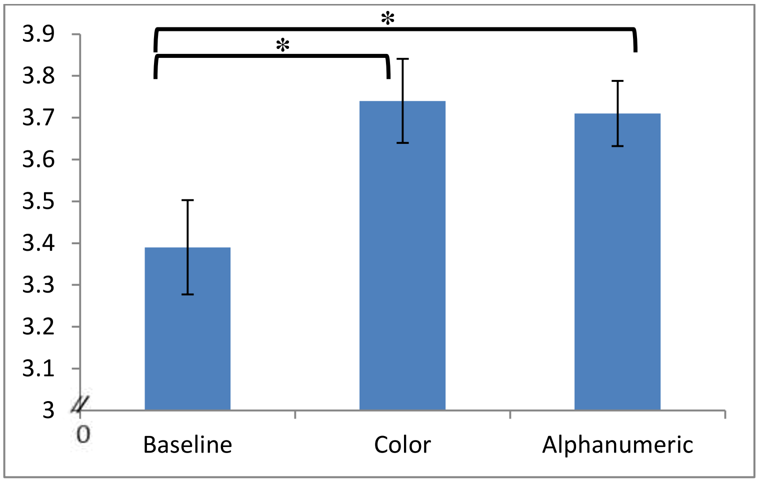


Figure 46. Satisfaction by salience type.

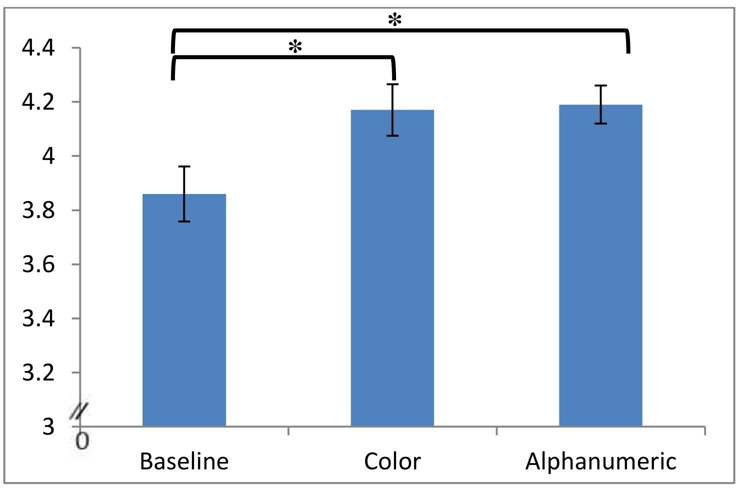


Figure 47. Behavior by salience type.

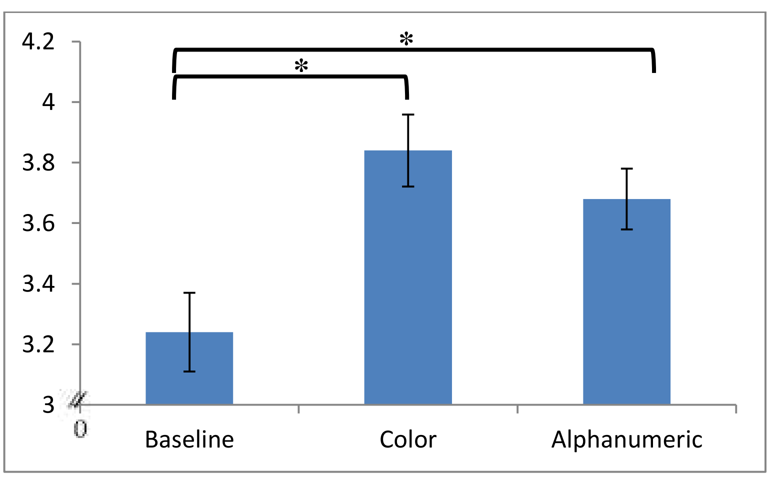


Figure 48. Improvement by salience type.

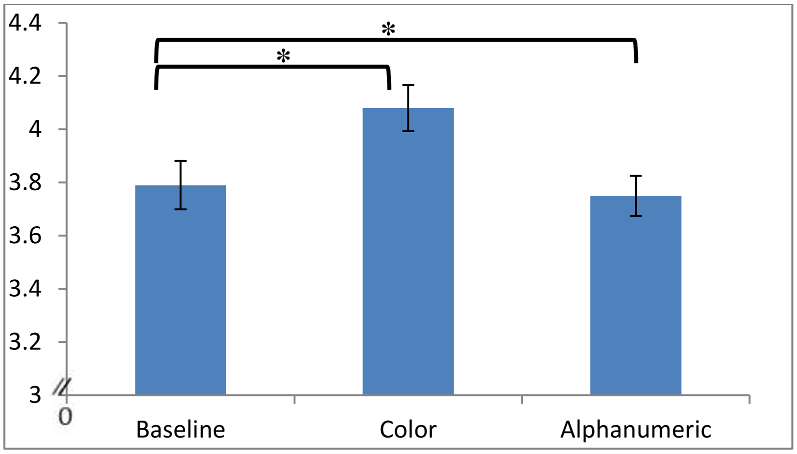


Figure 49. Mental model by salience type.

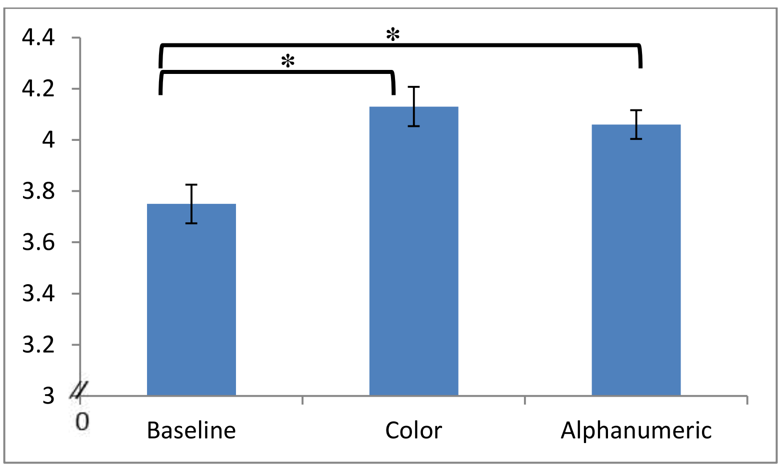


Figure 50. Global usability by salience type.

Individual Differences

Individual differences measures of MRT, TMT, LAG, and Gestalt were examined to determine if they had any relationship to the performance of data extraction above and beyond the experimental factors effects reported in the data extraction performance section. Each measure was evaluated using the General Linear Model separately. A 2(salience application) X 3(salience type) mixed MANCOVA with the practice session and individual differences measures as the covariates was computed for accuracy and response time performance for the dependent measures of low-, mid-, high-level data extraction, and the remediation task. Interactions were computed with overall performance scores. Results of the analyses are listed in Table 22 and Table 23. If the omnibus multivariate result was not significant, the univariate result was not report or interpreted.

For accuracy performance, none of the individual differences measures had significant impact on the performance measures. For response time performance, there was a significant main effect for mental rotation. However, subsequent analyses on each of the dependent measure did not reveal any significant impact on response time performance. There was also a LAG by salience application effect on response time performance. Further analysis revealed significant impact of LAG by salience application for mid-level data extraction response time. However, none of the subsequent correlations were significant.

Table 22. Study 2 Effects of individual differences on accuracy performance

Individual Differences Measure	Omnibus Multivariate Test
MRT	$F(4,56)$=1.91, p=.12, η_p^2=.12
MRT x Application	$F(4,56)$=.1.28, p=.29, η_p^2=.084
MRT x Salience	$F(8,52)$=.80, p=.60, η_p^2=.110
MRT x Application x Salience	$F(8,52)$=.96, p=.48, η_p^2=.129
TMT	$F(4,56)$=.85, p=.50, η_p^2=.057
TMT x Application	$F(4,56)$=1.14, p=.35, η_p^2=.075
TMT x Salience	$F(8,52)$=1.07, p=.40, η_p^2=.141
TMT x Application x Salience	$F(8,52)$=1.07, p=.40, η_p^2=.141
LAG	$F(4,56)$=.47, p=.76, η_p^2=.032
LAG x Application	$F(4,56)$=2.22, p=.08, η_p^2=.137
LAG x Salience	$F(8,52)$=1.51, p=.18, η_p^2=.188
LAG x Application x Salience	$F(8,52)$=.60, p=.78, η_p^2=.084
Gestalt	$F(4,56)$=.63, p=.65, η_p^2=.043
Gestalt x Application	$F(4,56)$=.73, p=.58, η_p^2=.049
Gestalt x Salience	$F(8,52)$=.79, p=.61, η_p^2=.109
Gestalt x Application x Salience	$F(8,52)$=1.69, p=.13, η_p^2=.206

Table 23. Study 2 Effects of individual differences on response time performance

Individual Differences Measure	Omnibus Multivariate Test	Performance Dependent Variables			
		Low	Mid	High	Remediation
MRT	$F(4,56)=2.82$, $p=.033$ $\eta_p^2=.168$	$F(1,59)=.05$, $p=.83$ $\eta_p^2=.007$	$F(1,59)=3.69$, $p=.06$ $\eta_p^2=.059$	$F(1,59)=.39$, $p=.53$ $\eta_p^2=.007$	$F(1,59)=.13$, $p=.72$ $\eta_p^2=.002$
MRT x Application	$F(4,56)=.67$, $p=.12$, $\eta_p^2=.046$	—	—	—	—
MRT x Salience	$F(8,52)=.62$, $p=.75$, $\eta_p^2=.087$	—	—	—	—
MRT x Application x Salience	$F(8,52)=.67$, $p=.72$, $\eta_p^2=.093$	—	—	—	—
TMT	$F(4,56)=.51$, $p=.73$ $\eta_p^2=.035$	—	—	—	—
TMT x Application	$F(4,56)=.82$, $p=.52$, $\eta_p^2=.055$	—	—	—	—
TMT x Salience	$F(8,52)=.55$, $p=.82$, $\eta_p^2=.078$	—	—	—	—
TMT x Application x Salience	$F(8,52)=1.46$, $p=.20$, $\eta_p^2=.183$	—	—	—	—
LAG	$F(4,56)=1.10$, $p=.37$ $\eta_p^2=.073$	—	—	—	—
LAG x Application	$F(4,56)=3.23$, $p=.02$, $\eta_p^2=.188$	$F(1,59)=.03$, $p=.87$, $\eta_p^2=.026$	$F(1,59)=5.31$, $p=.025$, $\eta_p^2=.083$ Same: r=.08, $p=.66$ Different: r=.20, $p=.28$	$F(1,59)=.17$, $p=.69$, $\eta_p^2=.003$	$F(1,59)=.40$, $p=.53$, $\eta_p^2=.007$
LAG x Salience	$F(8,52)=.93$, $p=.50$, $\eta_p^2=.125$	—	—	—	—
LAG x Application x Salience	$F(8,52)=.98$, $p=.46$, $\eta_p^2=.131$	—	—	—	—
Gestalt	$F(4,56)=.79$, $p=.54$ $\eta_p^2=.053$	—	—	—	—
Gestalt x Application	$F(4,56)=.50$, $p=.74$, $\eta_p^2=.034$	—	—	—	—
Gestalt x Salience	$F(8,52)=.84$, $p=.54$, $\eta_p^2=.114$	—	—	—	—
Gestalt x Application x Salience	$F(8,52)=.98$, $p=.47$, $\eta_p^2=.130$	—	—	—	—

Summary of results

Study 2 also examined the effects of salience application and salience type on data extraction. The one difference between the first and second study was the manipulation of salience application. Where in study 1, only one salience technique was applied to either low or mid-level data on the display. The second study applied two salience techniques to the display. Application in the second study was either the same salience technique was applied to low- and mid-data levels or different salience techniques were applied.

For data extraction performance, hypothesis 4 was divided into 3 subparts such that 4a predicted that both salience techniques mapped to low and mid-levels data would improve response time and accuracy performance of multi-level data extraction compared to the baseline display. Hypothesis 4a was partially supported. For mid- and high-level data extraction, accuracy scores were better with salience techniques compared to the baseline. General trends show that participants were most accurate using alphanumeric techniques, followed by color techniques, and least accurate with the baseline display for mid- and high-level data extraction. Response time performance was only significantly improved with high-level data extraction. Color and alphanumeric techniques resulted in significantly faster response times compared to the baseline for high-level data extraction.

4b predicted that different salience techniques mapped to low and mid-level data would improve multi-level data extraction performance compared to same salience techniques. This hypothesis was partially supported. No significant differences were found between same and different application conditions for accuracy performance on the low-, mid-, and high-level data extraction. However, significant impact of salience application was found for response time performance. Participants were significantly faster using displays with different salience techniques for low-, mid-, and high-level data extraction.

Finally, 4c predicted that alphanumeric techniques mapped to low level data and colors techniques mapped to mid-level data would improve response time and accuracy performance of multi-level data extraction compared to other same and different combinations of salience techniques. Hypothesis 4c was not supported. No interaction effects were found for the accuracy performance scores in any of the dependent measures. For response time performance, only high-level data extraction response times showed a significant interaction effect. Participants were faster using alphanumeric displays is the different condition and color displays in the same condition. However, displays where alphanumeric techniques were mapped low level data and colors techniques to mid-level data did not result in significantly faster than response times compared to when color techniques were mapped to low and mid-data levels on the display. These two techniques had one thing in common in which color techniques were mapped to mid-level data. Perhaps this is the reason why the two techniques resulted in better performance but were not significantly different from each other.

Hypothesis 5a and 5b tested the effects of salience application and salience type on the remediation task. It was predicted that a) mapped salience techniques would improve remediation performance in response time and accuracy compared to the baseline display, and b) remediation performance in response time and accuracy would improve when color techniques are mapped to low-level data, regardless of which salience technique is applied to mid-level data. Hypothesis 5a was partially supported. For accuracy performance, alphanumeric techniques resulted in higher accuracy compared to the baseline display for the remediation task. No significant differences were found for salience type in response times for the remediation task. .

Hypothesis 5b was not supported. No interaction effects were found for accuracy performance measures. There was a significant interaction effect where participants were faster using alphanumeric displays is the different condition and color displays in the same condition. However, the same color condition was not significantly different from the different alphanumeric condition, $t(62)=.77$, $p=.445$.

Hypothesis 6a and 6b examined how workload was impacted from the experimental factors. Hypothesis 6a predicted that different salience techniques applied to the display will reduce subjective workload compared to when two of the same salience techniques are applied to the display. This hypothesis was not supported. Different application of salience techniques resulted in higher workload for the mental, frustration, temporal, effort, and global workload. This is surprising as participants were significantly faster with different salience techniques compared to the same salience techniques on all dependent measures (low, mid, high-level extraction and the remediation task). Hypothesis 6b predicted that alphanumeric techniques are mapped to low level data and colors techniques to mid-level data, subjective workload would be significantly lower compared to other same and different salience combinations. This hypothesis was not supported. Although significant effects for salience application and salience types were found each individual, significant interaction effects were not found

Some usability effects were found with salience type where addition of salience techniques (color or alphanumeric values) increased perceived usability for data extraction performance. Differences in application and interaction of salience type and application were not found.

Study 2 discussion

Study 2 held a mixture of partially supported and not supported hypotheses. Results from this study indicate that applying different salience techniques combinations are better than the same salience technique combinations for response time. For both high-level extraction and the remediation task, response time performance was generally best with different application of alphanumeric values mapped to low-level data and color mapped to mid-level data or color mapped to both low and mid-levels of data on the display. One reason why these two combinations were both successful in reducing response times is that they both had color techniques mapped to mid-level data. Comparisons with results of Study 1 would need to be made to determine if there are differences in these two combinations and just applying color at mid-data levels. Analyses comparing salience combinations across all levels of application and salience types would reveal which salience combination, if any, was superior.

Additionally, different subjective workload and usability results were found with study 2 than study 1 when salience application was changed. Analysis combining the two studies is needed in order to determine which salience combination can achieve the greatest reduction in workload and highest usability. The following chapter will combine study 1 and study 2 to analyze them collectively and test the across studies hypotheses.

CHAPTER 5: ACROSS STUDIES

Two better than one?

To test if two salience techniques are better than one salience techniques for data extraction performance, the between subjects variable of salience application was grouped by number of salience techniques applied. The between subjects variable of salience application for each study was collapsed so that a new between subjects variable for this analysis was number of salience techniques and it included two levels: one salience technique applied and two salience techniques applied Therefore, a 2(number of salience techniques) X 3(salience technique) mixed model MANCOVA was conducted with the dependent measures of low-, mid-, high-level data extraction, and the remediation task with practice session as the covariate for each accuracy and response time performance measures.

Accuracy. In the MANCOVA conducted for accuracy performance, the covariate of practice session accuracy scores was significant, $F(4,122)=14.10$, $p<.001$, $\eta_p^2=0.321$. There was a significant interaction of salience type and number of salience techniques $F(8,118)=4.97$, $p<.001$, $\eta_p^2=.252$. There was no significant main effect for the between-subjects condition of number of salience techniques $F(4,112)=.23$ $p=.919$, $\eta_p^2=0.008$. Finally, there was a significant main effect for the within-subject variable of salience type, $F(8,118)=12.62$, $p<.001$, $\eta_p^2=0.461$. The significant main effect of the within-subjects variable of salience type will not be reported as the interest of this analysis test is to see if number of salience techniques had significant impact to performance. Additionally, the significant main effect of salience type was also observed for testing of specific salience combinations and will be reported in the next section.

For the interaction effect of number of salience techniques and salience application, significant impacts were found for mid-level data extraction [$F(2,250)=5.03$, $p=.007$, $\eta_p^2=0.039$] and high-level data extraction [$F(2,250)=7.54$, $p=.001$, $\eta_p^2=0.057$]. For mid-level data extraction accuracy, two salience techniques were better with alphanumeric techniques compared to the baseline [$t(122)=5.06$, $p<.001$] and when one alphanumeric salience technique was applied [$t(122)=2.11$, $p=.036$]. Two color salience techniques also resulted in better accuracy scores compared to the baseline $t(122)=3.5$, $p<.001$. Figure 51 shows a visual representation.

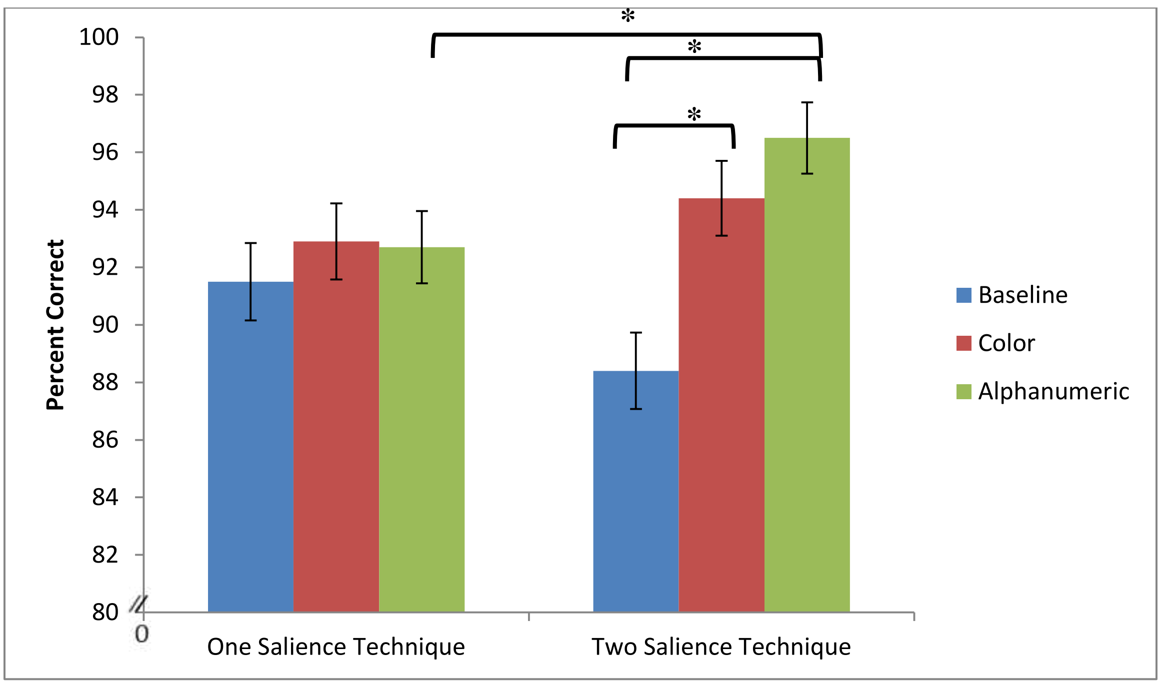


Figure 51. Mid-level data extraction accuracy by number of salience techniques and salience type.

In high-level data extraction, one color salience technique resulted in higher accuracy scores compared to the baseline [$t(122)=5.45$, $p<.001$],one alphanumeric technique [$t(122)=3.92$, $p<.001$], and two color salience techniques [$t(122)=2.15$, $p=.036$]. Two alphanumeric salience techniques resulted in better accuracy scores compared to the baseline [$t(122)=2.64$, $p<.001$] and one salience technique [$t(122)=2.86$,$p=.006$]. See Figure 52 for a visual representation.

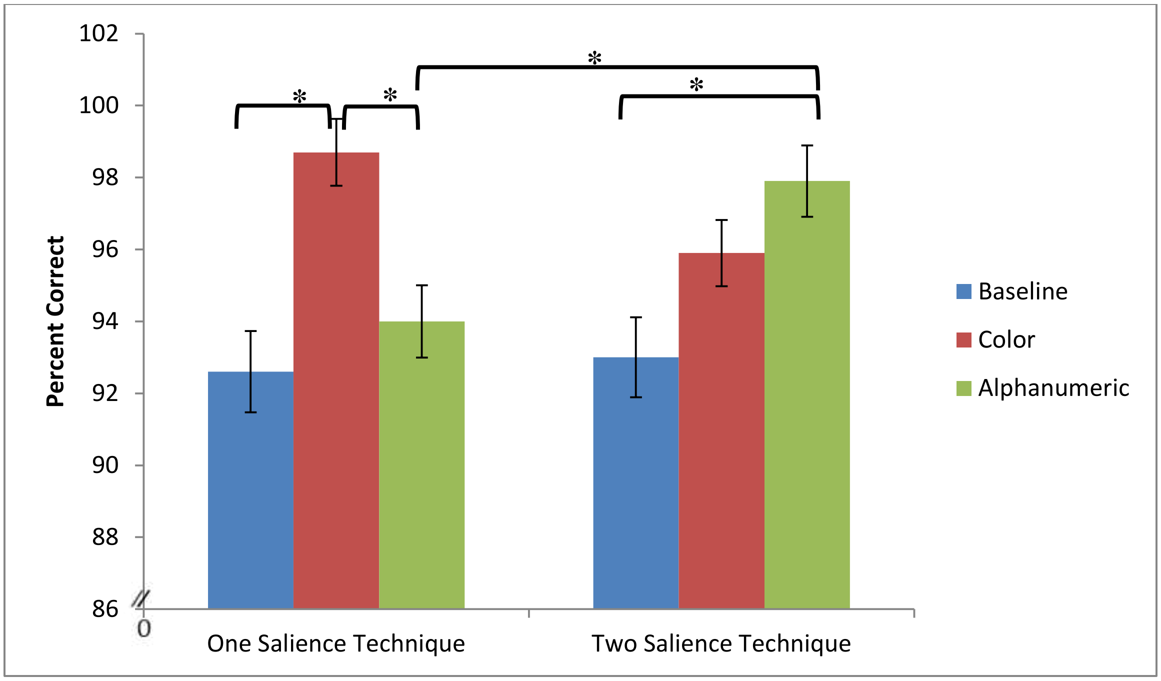


Figure 52. High-level data extraction accuracy by number of salience techniques and salience type.

Response time. The 2x3 mixed model MANCOVA revealed significant impact of the practices session response time covariate $F(4,122)=24.25$, $p<.001$, $\eta_p^2=.443$. No significant interaction of number of salience technique by salience type was found, $F(8,118)=1.61$, $p=.13$, $\eta_p^2=.098$. Additionally, no significant main effect for the between subjects variable of number of salience techniques was found, $F(4,122)=1.27$, $p=.285$, $\eta_p^2=.040$. There was a significant main effect for the within subjects variable of salience type, $F(8,118)=4.97$, $p=.00$, $\eta_p^2=.252$. Again, the significant main effect of the within-subjects variable of salience type will not be reported as the interest of this analysis test is to see if number of salience techniques had significant impact to performance, and the main effect of salience type for specific salience technique combinations will be described in the section below.

Testing by specific combinations

Accuracy. A 4 (salience application) x 3 (salience type) mixed model MANCOVA was conducted with the dependent measures of low-, mid-, high, and remediation task accuracy scores with practice sessions accuracy as the covariate. The practice session response time covariate was significant $F(4,120)=13.185$, $p<.001$, $\eta_p^2=0.350$. There was a significant interaction of salience type and application $F(24,337.04)=2.30$, $p=.001$, $\eta_p^2=.136$. There was no significant main effect for the between-subjects condition of application $F(12,317.78)=1.117$, $p=.345$, $\eta_p^2=0.036$. Finally, there was a significant main effect for the within-subject variable of salience $F(8,116)=11.788$, $p<.001$, $\eta_p^2=0.448$.

Mauchley's test of sphericity was statistically significant for the dependent measure of high-level data extraction response ($\chi2(2)=9.526$, $p=0.009$). These significant results indicate that the assumption of sphericity had been violated. As such, the degrees of freedom of were adjusted using the Huynh-Feldt estimates of sphericity.

Univariate tests revealed significant effect of the within-subjects variable of salience type on all four dependent measures of low-level data extraction accuracy [$F(2,246)=4.678$, $p=.010$, $\eta_p^2=.037$], high-level data extraction accuracy [$F(1.949,229.74)=26.128$, $p<.001$, $\eta_p^2=.175$], and the remediation task accuracy [$F(2,246)=16.914$, $p<.001$, $\eta_p^2=.121$.

Although the univariate test for low-level data extraction accuracy was significant, pairwise comparisons did not reveal any significant differences between the salience types (see Figure 53). For high-level data extraction, salience type had a significant effect on participants' accuracy scores when color [$t(123)=5.625$, $p<.001$] and alphanumeric techniques [$t(123)=3.1$, $p=.010$] were applied compared to the baseline (see Figure 54). For the remediation task, salience type had a significant effect on participants' accuracy scores when color [$t(123)=3.07$, $p=.014$] and alphanumeric techniques [$t(123)=3.31$, $p=.016$] were applied compared to the baseline (see Figure 55). Means and standard deviations for each level of extraction by salience type are listed in Table 24.

Table 24. Accuracy performance means and standard divinations by salience type.

Dependent Measure (Accuracy)	Mean N=128	Standard Deviation
Low-level Data Extraction*		
Base	94.15	11.98
Color	93.75	11.09
Alpha	92.09	12.24
Mid-level Data Extraction		
Base	89.94	11.61
Color	93.65	10.98
Alpha	94.63	10.72
High-level Data Extraction*		
Base	92.81	11.23
Color	97.27	7.51
Alpha	96.00	8.72
Remediation Task*		
Base	86.31	16.59
Color	90.53	14.78
Alpha	91.60	11.75

*Note: Higher numbers indicate better accuracy performance

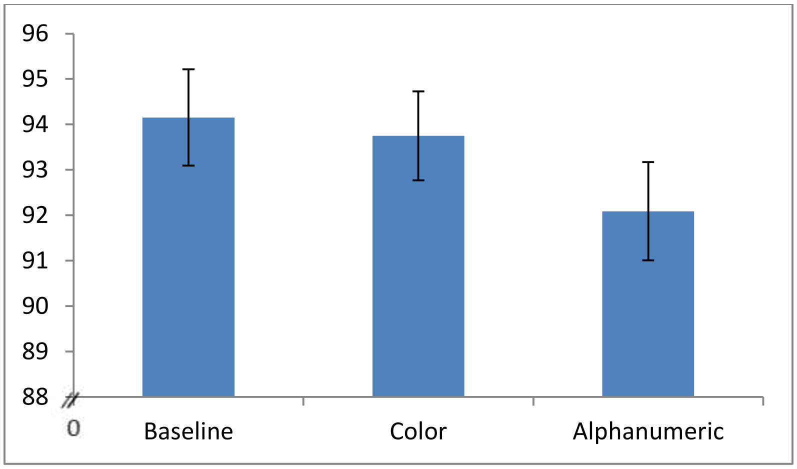


Figure 53. Accuracy means for low-level data extraction by salience type.

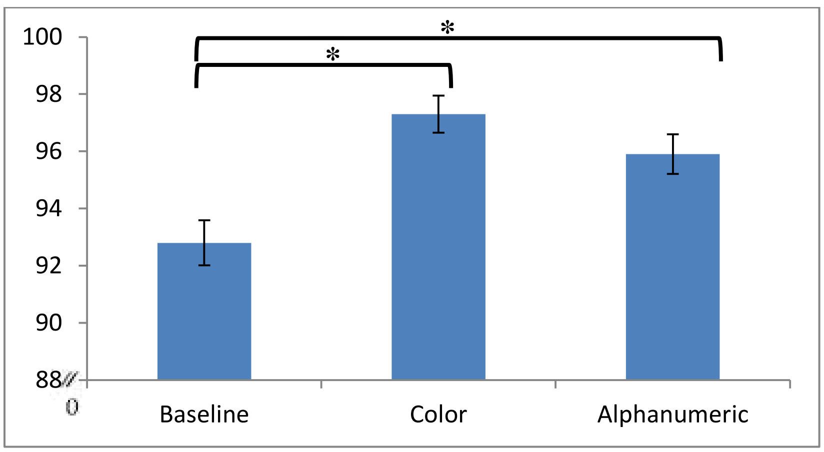


Figure 54. Accuracy means for high-level data extraction by salience type.

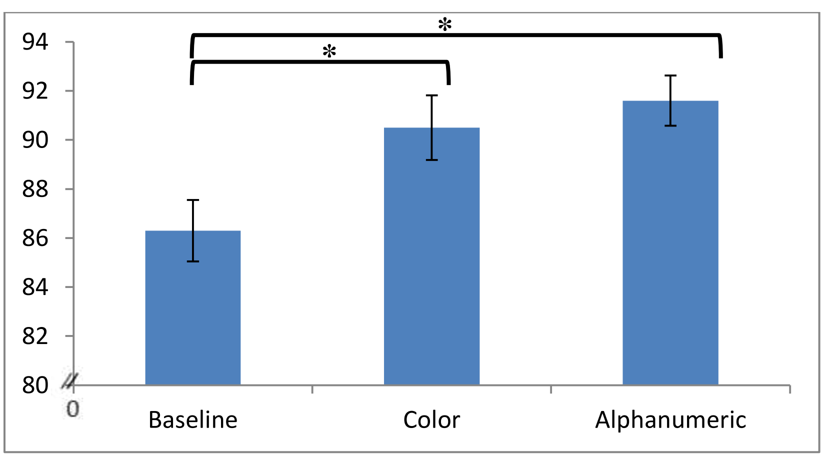


Figure 55. Accuracy means for remediation task by salience type.

Further analysis of the interaction effect revealed significant effect of salience application and salience type on high-level data extraction accuracy $F(5.85, 229.74)=2.826$, $p=.012$, $\eta_p^2 =.064$. When alphanumeric techniques were applied in the low condition, participant were significantly less accurate at high-level data extraction compared to when alphanumeric values were applied in the mid condition [$t(123)=-2.05$, $p=.041$], same condition [$t(123)=-3.55$, $p=.001$], different condition[$t(123)=-2.6$, $p=.011$], and when color was applied to the low condition [$t(123)=-3.47$, $p=.001$]. Color applied to the mid condition resulted in higher accuracy scores compared to when color was applied to the different condition [$t(123)=2.22$, $p=.034$], alphanumeric techniques at the mid condition [$t(123)=2.12$, $p=.032$], and the baseline [$t(123)=4.0$, $p<.001$]. When techniques were applied in the low condition, participants using color were significantly more accurate at high-level data extraction than when using the baseline [$t(123)=3.56$, $p=.001$]. When techniques were applied in the same condition, participants using alphanumeric techniques were significantly more accurate at high-level data extraction than when using the baseline [$t(123)=3.0$, $p=.004$]. When one salience technique was applied to the display, color was superior (low: 5.7%; mid: 6.3%, increases from the baseline). However, when two salience techniques were applied to the display, alphanumeric techniques were better (same: 5.7%; different: 4.3%, increases from the baseline). See Figure 56 for a visual representation.

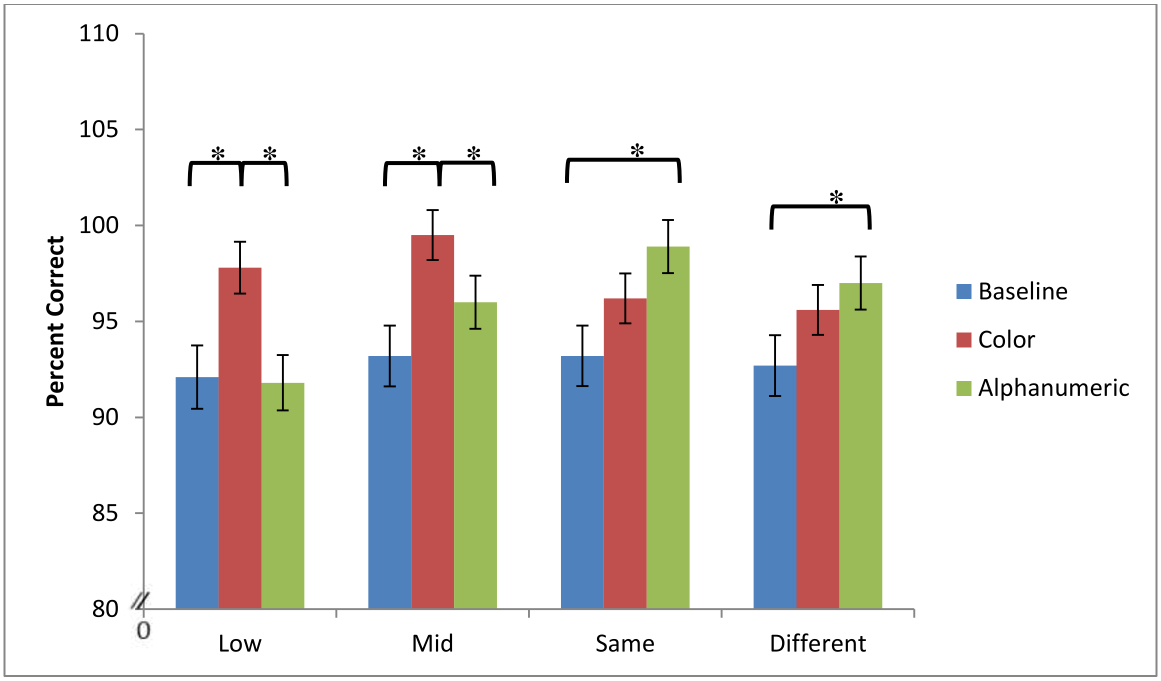


Figure 56. Accuracy means for high-level data extraction by salience type and salience application.

Response time. A 4 (salience application) x 3 (salience type) mixed model MANCOVA was conducted with the dependent measures of low-, mid-, high, and remediation task response times with practice sessions response time as the covariate. The practice session response time covariate was significant $F(4,120)=26.581$, $p<.001$, $\eta_p^2=0.470$. There was a significant interaction of salience type and application $F(24,337.04)=3.334$, $p<.001$, $\eta_p^2=.186$. There was also significant main effect for the between-subjects condition of application $F(12,317.78)=1.854$, $p=.039$, $\eta_p^2=0.058$. Finally, there was a significant main effect for the within-subject variable of salience $F(8,116)=4.57$, $p<.001$, $\eta_p^2=0.240$.

Mauchley's test of sphericity was statistically significant for the dependent measure of high-level data extraction response ($\chi2(2)=6.263$, $p=0.044$) and remediation task response time ($\chi2(2)=11.297$, $p=0.004$). These significant results indicate that the assumption of sphericity had been violated. As such, the degrees of freedom of were adjusted using the Huynh-Feldt estimates of sphericity.

Univariate tests revealed significant effect of the within-subjects variable of salience type on all four dependent measures of low-level data extraction response time [$F(2,246)=5.168$, $p=.006$, $\eta_p^2=.040$], mid-level data extraction response time [$F(2,246)=4.154$, $p=.017$, $\eta_p^2=.033$], high-level data extraction response time [$F(1.997,245.59)=5.112$, $p=.007$, $\eta_p^2=.040$], and the remediation task response time [$F(1.925,236.74)=3.22$, $p=.044$, $\eta_p^2=.026$.

Participants had significantly faster response time for low-level data extraction when using alphanumeric techniques compared to the baseline [$t(123)=3.40$, $p=.001$] and color techniques [$t(123)=2.50$, $p=.014$] (see Figure 57). In mid-level data extraction, participants were significantly faster using the color techniques compared to the baseline [$t(123)=-4.96$, $p<.001$] and alphanumeric techniques [$t(123)=-3.12$, $p=.002$] (see Figure 58). For high-level data extraction, participants had significantly faster response times using color techniques compared to the baseline [$t(123)=-9.96$, $p<.001$] and alphanumeric techniques [$t(123)=-4.28$, $p<.001$]. Additionally, participants were significantly faster at high-level data extraction when using the alphanumeric techniques compared to the baseline display, $t(123)=6.82$, $p<.001$ (see Figure 59). In the remediation task, participants had significantly faster response times with the color techniques compared to the baseline [$t(123)=-7.87$, $p<.001$] and alphanumeric techniques [$t(123)=-3.83$, $p<.001$]. Additionally, participants were significantly faster at the remediation task when using alphanumeric techniques compared to the baseline display, $t(123)=-5.41$, $p<.001$ (see Figure 60). Means and standard deviations for data level extraction response times for salience type are listed in table 24.

Table 25. Response means (millisecond) and standard deviations for low-, mid-, high-, and remediation task by salience type.

Dependent Measure (Response Time)	Mean n=128	Standard Deviation
Low-level Data Extraction		
Base	3966.56	1038.21
Color	4019.09	1444.96
Alpha	4313.49	1381.46
Mid-level Data Extraction		
Base	4631.28	1388.36
Color	4071.44	1541.81
Alpha	4425.98	1314.49
High-level Data Extraction		
Base	3464.85	1157.02
Color	2506.95	1086.03
Alpha	2920.52	1038.55
Remediation Task		
Base	3160.80	964.89
Color	2379.71	1093.45
Alpha	2748.29	998.93

*Note: Lower numbers indicate better response time performance

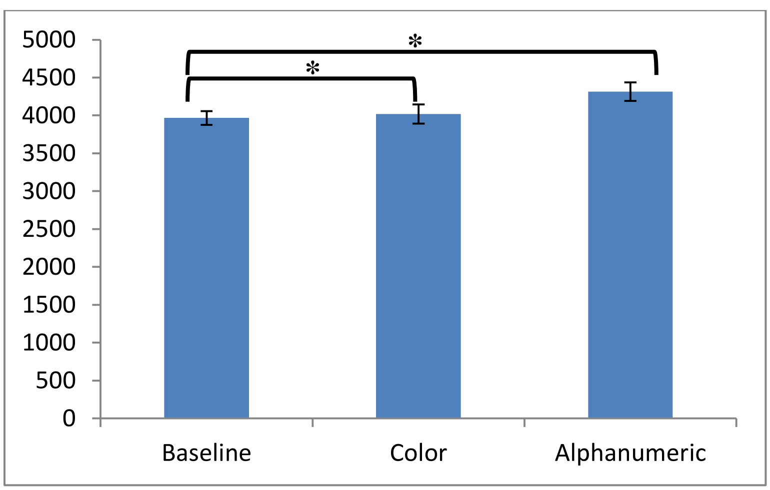


Figure 57. Response time for low-level by salience type.

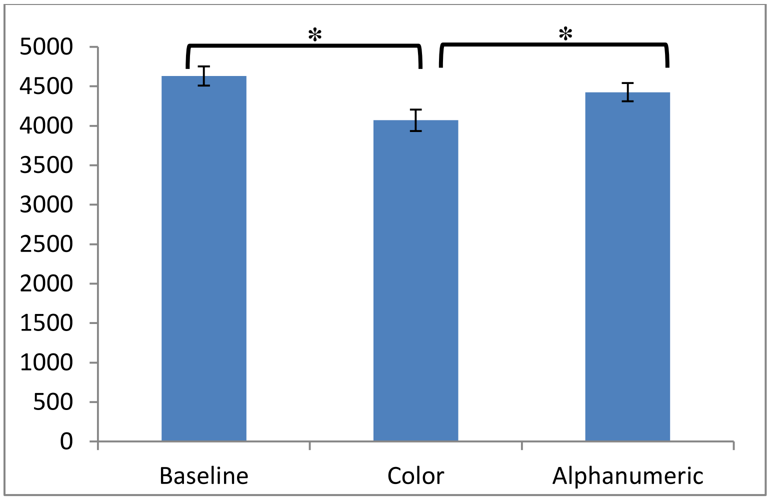


Figure 58. Response time for mid-level extraction by salience type.

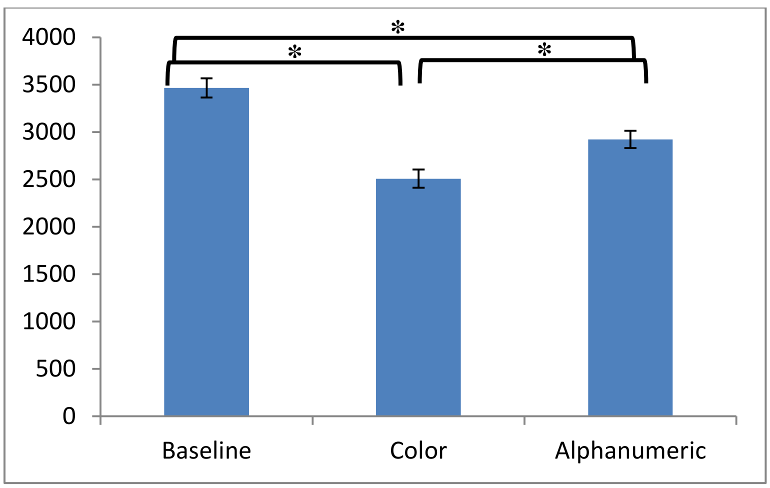


Figure 59. Response time for high-level extraction by salience type.

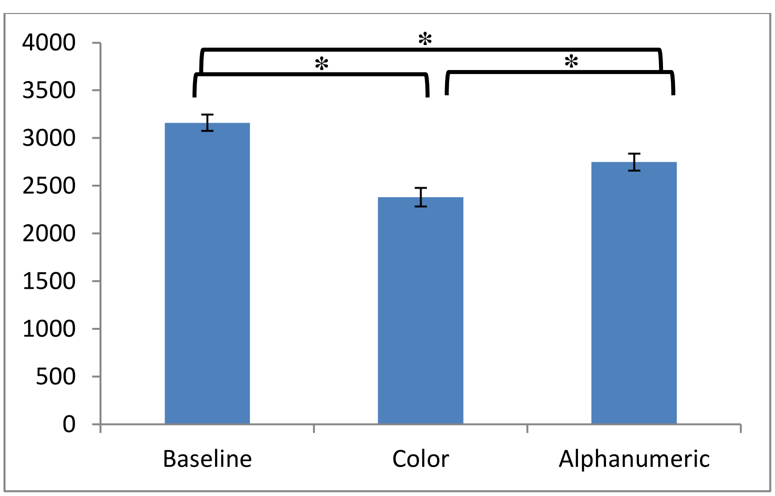


Figure 60. Response time for remediation task by salience type.

Univariate test examining effects of the between-subjects variable of salience application on each of the response time dependent measures showed significant effects on low-level data extraction response time [$F(3,123)=5.716$, $p=.001$, $\eta_p^2=.122$], mid-level data extraction response time [$F(3,123)=3.102$, $p=.029$, $\eta_p^2=.070$], high-level data extraction response time [$F(3,123)=3.335$, $p=.002$, $\eta_p^2=.075$], and remediation task response time [$F(3,123)=4.731$, $p=.004$, $\eta_p^2=.103$].

Participants were significantly faster at low-level data extraction when different salience techniques were mapped to the display compared to when the same salience techniques were mapped to the display [$t(123)=-3.34$, $p=.001$] and when only one salience technique was mapped to the mid-level data [$t(123)=-3.70$, $p<.001$] (see Figure 61). For mid-level data extraction, participants were again significantly faster when different salience techniques were mapped to the display compared to when the same salience techniques were mapped to the display [$t(123)=-2.27$, $p=.025$] and when only one salience technique was mapped to the mid-level data [$t(123)=-2.90$, $p=.005$] (see Figure 62). For high-level data extraction, participants were significant faster when different salience techniques were mapped to the display compared to when the same salience techniques were mapped to the display [$t(123)=-2.08$, $p=.034$] and when only one salience technique was mapped to the mid-level data [$t(123)=-3.706$ $p=.003$] or low-level data [$t(123)=-2.15$, $p=.034$] (see Figure 63). For the remediation task, participant were when different salience techniques were mapped to the display compared to when the same salience techniques were mapped to the display [$t(123)=-2.41$, $p=.017$] and when only one salience technique was mapped to the mid-level data [$t(123)=-3.67$, $p<.001$] (see Figure 64). Table 26 lists the response time means and standard deviations for each dependent measure by salience application condition.

Table 26. Response means (millisecond) for low-, mid-, high-, and remediation task by salience application

Dependent Measure (Response Time)	Low Application Mean (SD) n=31	Mid Application Mean (SD) n=32	Same Application Mean (SD) n=33	Different Application Mean (SD) n=32
Low-level Data Extraction	4005.63 (1299.50)	4412.88 (1302.84)	4334.15 (1380.53)	3635.93 (1033.37)
Mid-level Data Extraction	4339.33 (1263.10)	4656.13 (1612.61)	4510.86 (1245.95)	3993.26 (1480.46)
High-level Data Extraction	3030.41 (1069.63)	3189.08 (1251.92)	3007.69 (937.94)	2629.95 (1064.75)
Remediation Task	2712.80 (793.48)	3041.76 (1041.23)	2838.68 (965.25)	2454.55 (994.95)

*Note: Lower numbers indicate better response time performance

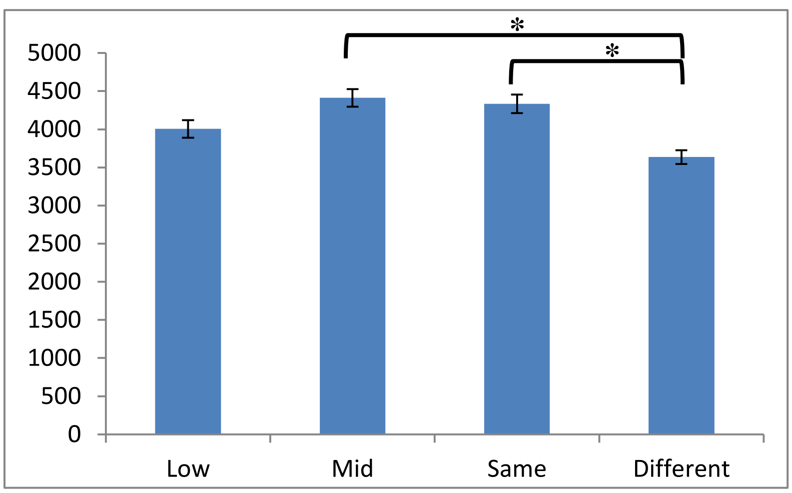


Figure 61. Response time for low-level extraction by salience application

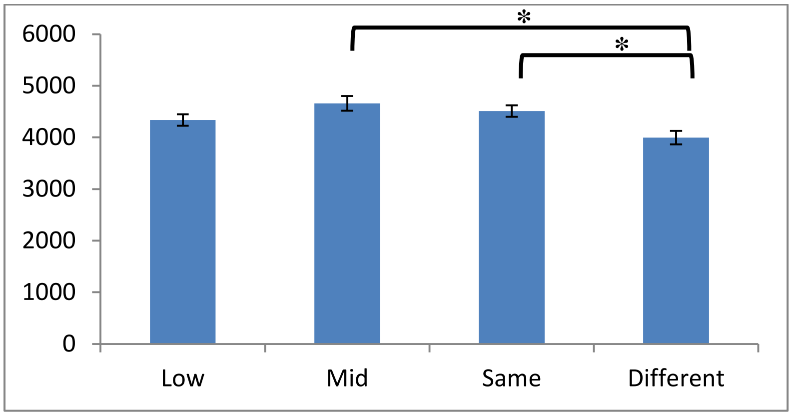


Figure 62. Response time for mid-level extraction by salience application.

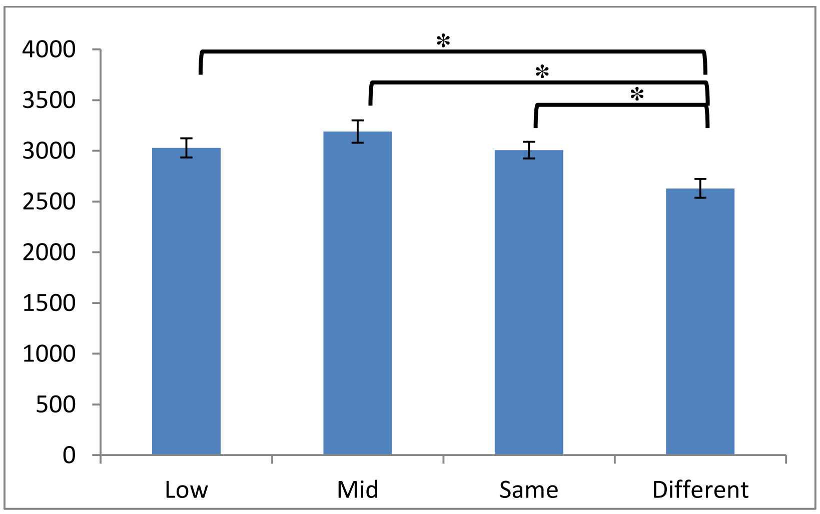


Figure 63. Response time for high-level extraction by salience application.

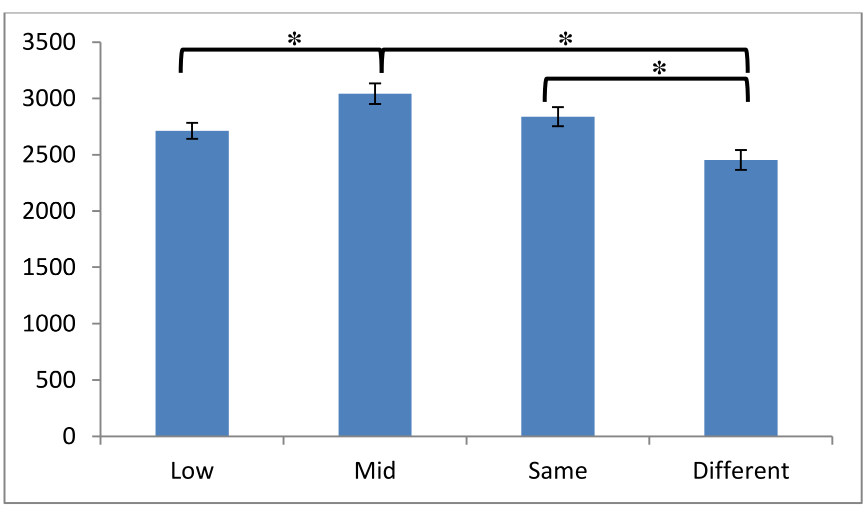


Figure 64. Response time for remediation task by salience application.

Further analysis of the interaction effect revealed significant effect of salience application and salience type on low-level data extraction response time [$F(6,246)=2.704$, $p=.015$, $\eta_p^2 = .062$], high-level data extraction response time [$F(5.99, 245.59)=4.3$, $p<.001$, $\eta_p^2 = .095$], and remediation task response time [$F(5.77, 236.74)=8.195$, $p<.001$, $\eta_p^2 = .167$].

In low-level data extraction, alphanumeric techniques applied in the different condition resulted in significantly faster response times compared to when alphanumeric techniques were applied to the low condition [$t(123)=-2.89$, $p=.005$], the mid condition [$t(123)=-4.07$, $p<.001$], and the same condition [$t(123)=-3.50$, $p=.001$]. Additionally, application of color techniques in the different condition results significantly faster response times compared to color application in the same condition,[$t(123)=-2.66$, $p=.009$], color applied to the mid condition [$t(123)=3.59$, $p<.001$], and the baseline display [$t(123)=2.14$, $p=.035$]. When one salience technique was applied to the display, both color and alphanumeric techniques did not perform better than the baseline. When two salience techniques were applied to the display, only different techniques reduced response time compared to the baseline (color: 11.5% reduction; alphanumeric: 6.8% reduction) for low-level extraction. Figure 65 visually represents the differences between salience type and salience application just described.

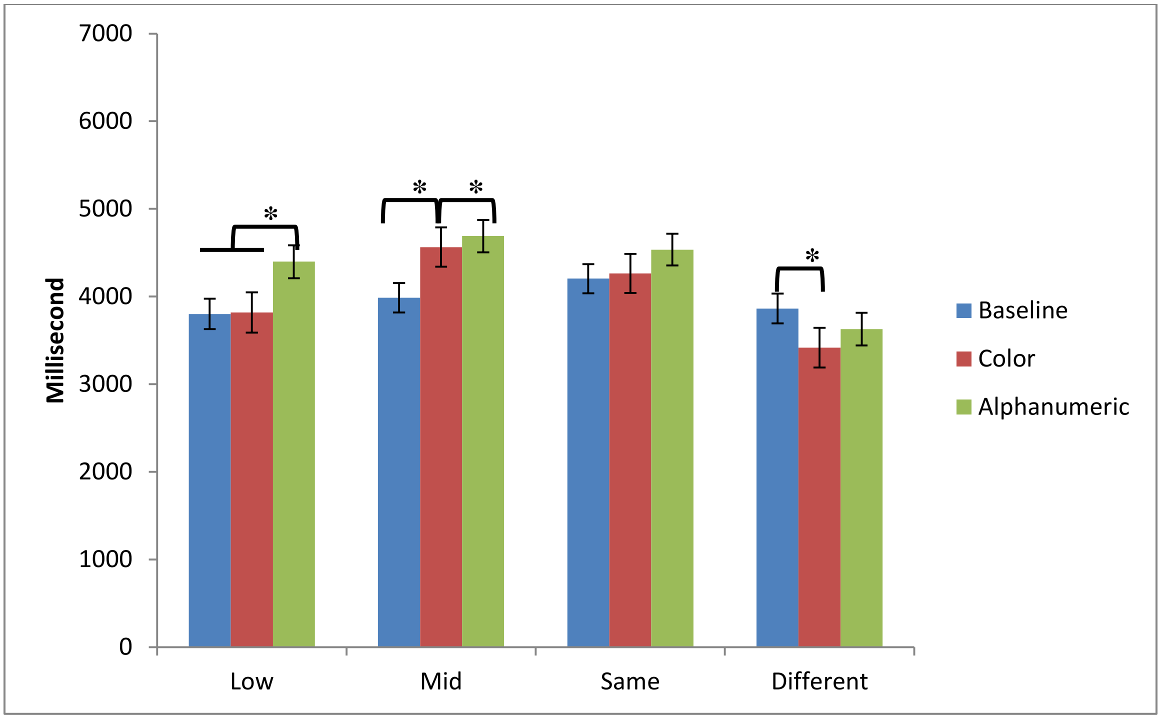


Figure 65. Response time means for low-level data extraction by salience type and salience application.

In high-level data extraction response time performance, application of alphanumeric values showed the most significant difference. Participants had significantly faster response times using salience techniques in the different alphanumeric condition compared to alphanumeric techniques applied in the low condition [$t(123)=-3.90$, $p<.001$], mid condition [$t(123)=-4.19$, $p<.001$], same condition[$t(123)=-5.05$, $p<.001$], the baseline [$t(123)= 6.95$, $p<.001$]. Color salience techniques did not reveal statistical significant differences when compared by application, but as can be seen by Figure 66, color applied in the different condition and in the same condition had the lowest response times. Therefore, these were examined to determine if those color application were significantly different from other salience techniques of the same application. When color was applied in the same condition, participants had significantly faster response times using color techniques compare to the baseline [$t(123)=-5.77$,

$p < .001$] and alphanumeric techniques in the same condition [$t(123)=-5.19$, $p < .001$]. Additionally, color techniques in the different condition were resulted in significantly faster response times compared to the baseline, $t(123)=4.71$, $p < .001$. When one salience technique was applied to the display, color was superior compared to the baseline (low: 25.8% reduction; mid: 25.1% reduction). When two salience techniques were applied to the display, color was superior for same application (32.0% reduction), and alphanumeric was superior for different application (47.0% reduction) compared to the baseline.

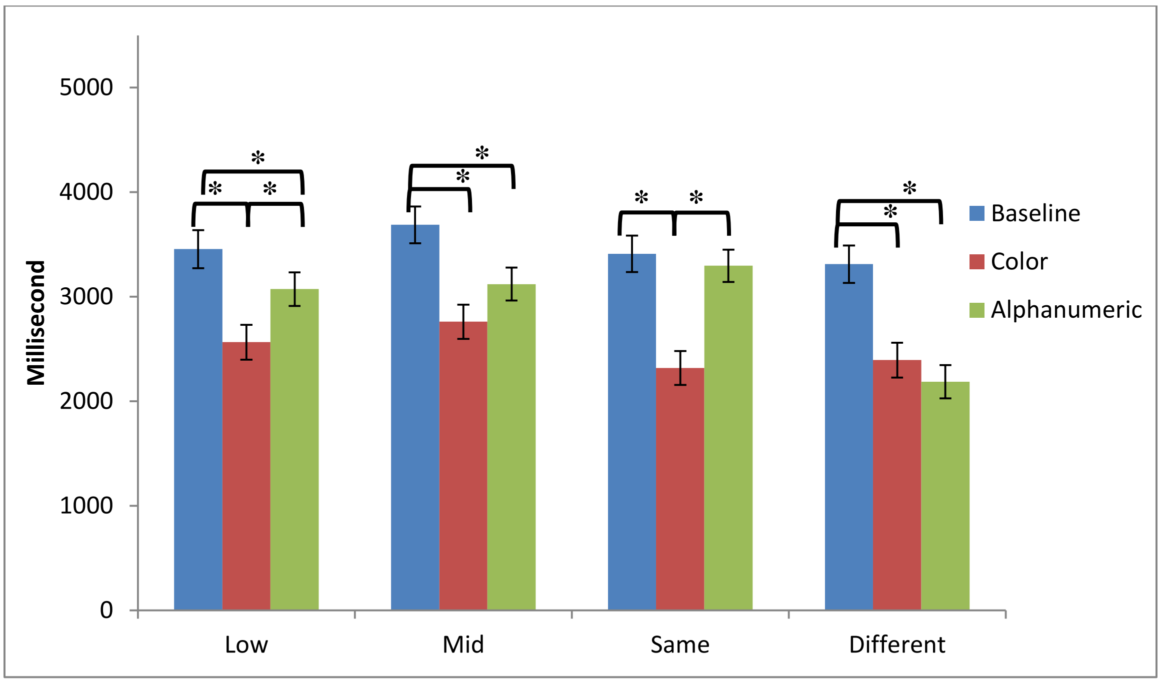


Figure 66. Response time means for high-level data extraction by salience type and salience application.

In remediation task response time performance, application of alphanumeric values again showed the most significant difference. Participants had significantly faster remediation task response times using salience techniques in the different alphanumeric condition compared to alphanumeric techniques applied in the low condition [$t(123)=-4.82$, $p < .001$], mid condition [$t(123)=-6.12$, $p < .001$], same condition[$t(123)=-6.73$, $p < .001$], color applied in the different condition [$t(123)=-325$, $p=.002$], and the baseline [$t(123)=-7.59$, $p=.001$]. Same color techniques applied revealed significant differences. Participants had significantly faster remediation task response times when color was applied in the same condition compare to when color was applied in the mid condition [$t(123)=-2.28$, $p=.025$], when same alphanumeric techniques were applied [$t(123)=-5.58$, $p<.001$], the baseline [$t(123)=-5.52$, $p<.001$]. Figure 67 shows that same color application response time is similar to low color application response times. When salience techniques were applied in the low condition, participants had significantly faster remediation task response times using color techniques compare to the baseline [$t(123)=-4.09$, $p<.001$] and alphanumeric techniques [$t(123)=-3.03$, $p=.003$]. When one salience technique was applied to the display, color was superior compared to the baseline (low: 27.0% reduction; mid: 20.0% reduction). When two salience techniques were applied to the display, color was superior for same application (33.6% reduction), and alphanumeric was superior for different application (47.3% reduction).

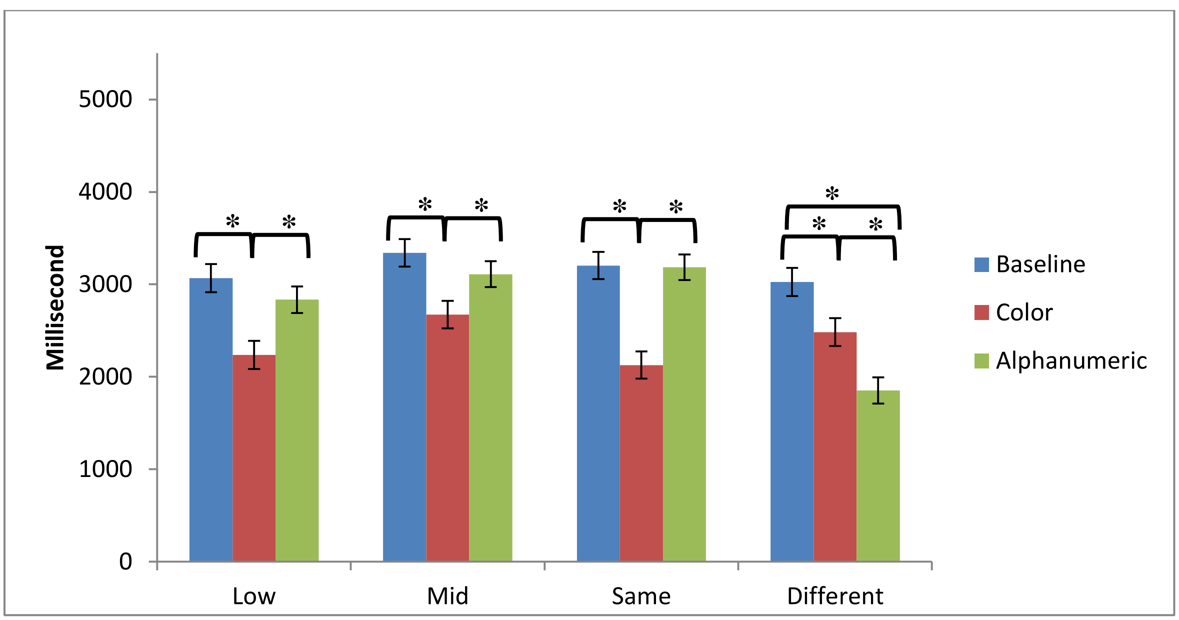


Figure 67. Response time means for the remediation task by salience type and salience application.

Subjective Workload

A 4 (salience application) x 3 (salience type) mixed model MANOVA was conducted with the dependent measures the six dimensions of the NASA TLX (mental, physical, frustration, temporal performance, and effort). A separate 4x3 mixed modal ANOVA was used to examine the global TLX score. There was no significant interaction of application and salience type on workload dimensions [$F(36,334)=1.09$, $p=.35$, $\eta_p^2=.103$] and for global workload. [$F(6,248)=1.56$, $p=.159$, $\eta_p^2=.037$]. Additionally there was not a significant main effect for application, $F(18,337.07)=1.30$, $p=.19$, $\eta_p^2=.061$. There was significant main effect of salience application on global workload, $F(3,124)=4.26$, $p=.007$, $\eta_p^2=.093$, in which the same (M=15.21, SD=11.47) application had significantly lower workload compared to low (M=24.32, SD=15.72), mid (M=22.86, SD=14.70), and different (M=24.59, SD=13.96) applications.

There was a significant main effect for salience type, $F(12,113)=2.80$, $p=.002$, $\eta_p^2=.229$. Univariate analyses revealed that significant differences in workload between salience types were found for Mental demand [$F(2,248)=8.77$, $p<.001$, $\eta_p^2=.066$], Physical demand [$F(2,248)=4.17$, $p=.017$, $\eta_p^2=.033$], Performance [$F(2,248)=12.58$, $p<.001$, $\eta_p^2=.092$], Effort [$F(2,248)=8.52$, $p<.001$, $\eta_p^2=.064$], Frustration [$F(2,248)=7.74$, $p=.001$, $\eta_p^2=.059$]. A significate main effect was found for Global workload [$F(2,248)=14.35$, $p<.001$, $\eta_p^2=.104$]. Higher workload was generally associated with the baseline display. Means and standard deviation are presented in Table 27. See figures 68-73 for a visual representation.

Table 27. Means and (standard deviations) for workload dimensions by salience type.

Salience Type	Mental*	Physical*	Temporal	Performance*	Effort*	Frustration*	Global*
Baseline	31.57 (23.65)	14.05 (15.29)	24.13 (21.67)	79.70 (23.05)	34.96 (24.65)	24.02 (24.89)	24.97 (15.71)
Color	24.81 (20.87)	12.23 (13.78)	21.32 (20.60)	87.55 (17.71)	27.62 (22.44)	17.56 (18.91)	19.46 (13.53)
Alphanumeric	27.28 (22.08)	11.69 (11.49)	22.21 (19.81)	84.40 (20.78)	30.15 (24.05)	17.64 (17.04)	20.80 (12.91)

*Denotes significance at *p*<.05.

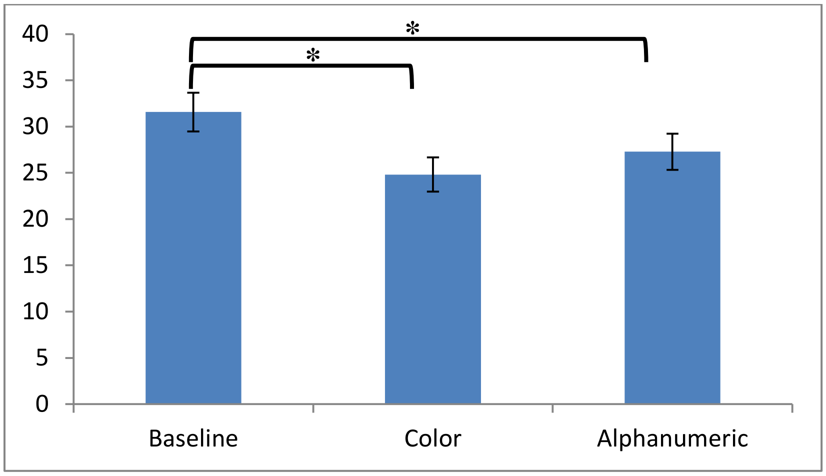


Figure 68. Mental workload by salience type.

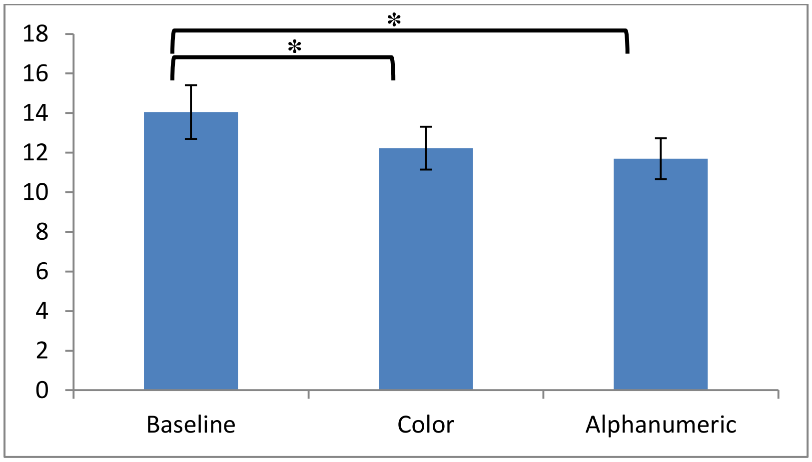


Figure 69. Physical workload by salience type.

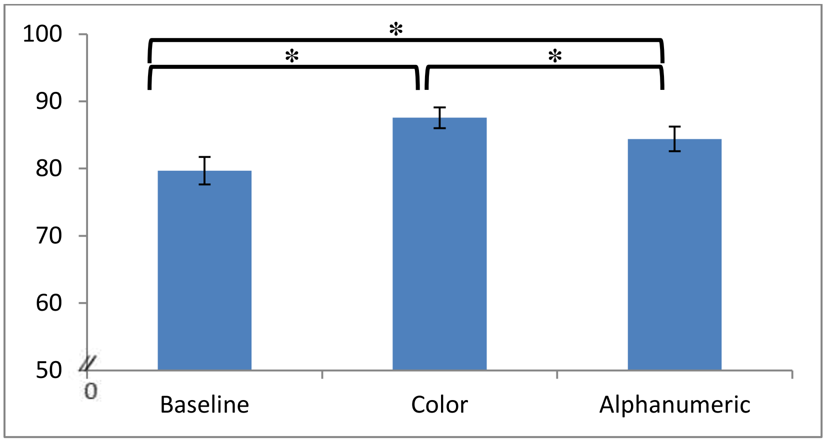


Figure 70. Performance by salience type

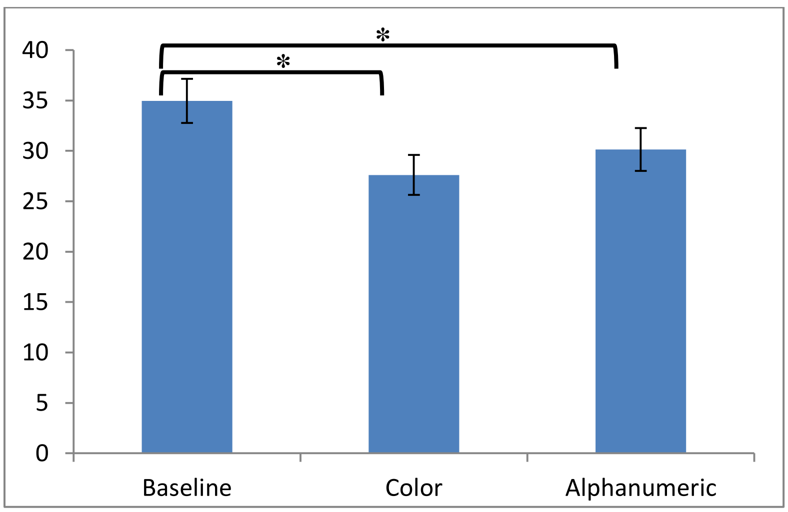


Figure 71. Effort by salience type.

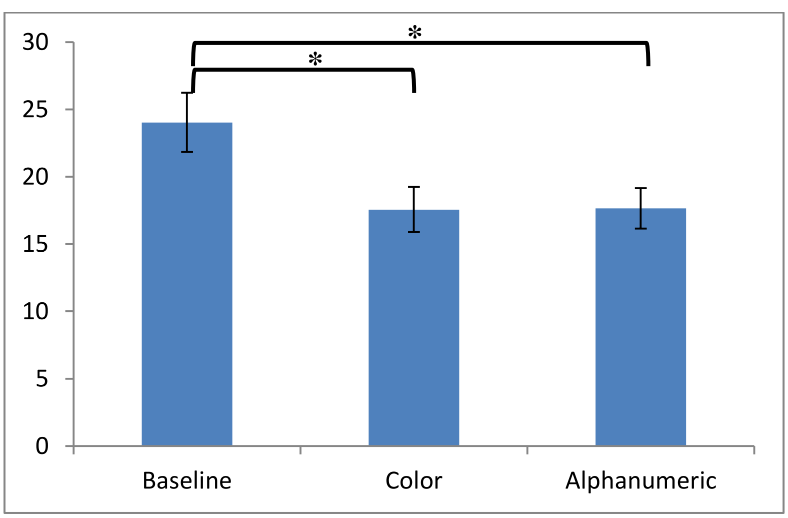


Figure 72. Frustration by salience type.

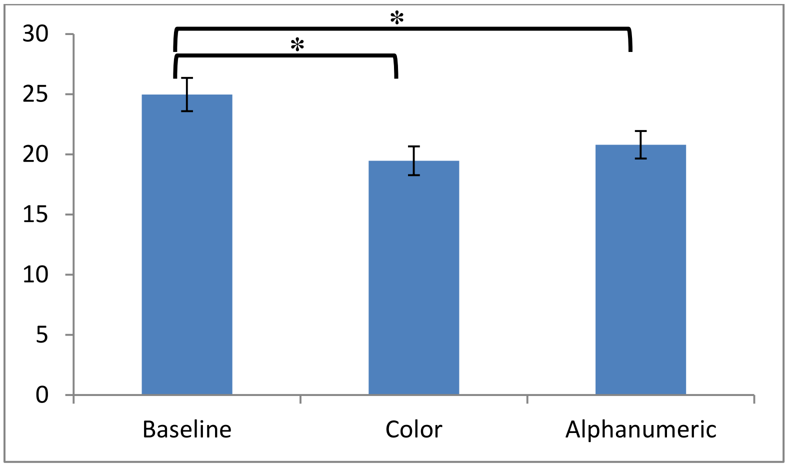


Figure 73. Global workload by salience type.

Usability

A 4 (salience application) x 3 (salience type) mixed model MANCOVA was conducted with the MUQ sub dimensions (simplicity, usefulness, functionality, consistency, proficiency, satisfaction, behavior, improvement, and mental model). A separate 2x3 mixed modal ANOVA was used to examine the global usability measure. There was no significant interaction effect of salience application and salience type on the MUQ dimensions [$F(54,318)=1.06$, $p=.37$, $\eta_p^2=.153$] or for global usability [$F(6,244)=1.87$, $p=.09$, $\eta_p^2=.044$]. Additionally, there was no significant effect of the between-subjects variable of salience application on MUQ measures, [$F(27,345)=.64$, $p=.93$, $\eta_p^2=.048$] or for global usability [$F(3,122)=2.07$, $p=.108$, $\eta_p^2=.048$]. There was significant omnibus effect of the within-subjects variable of salience type on the MUQ measures, $F(18,104)=2.08$, $p=.012$, $\eta_p^2=.264$. Univariate tests of the within-subjects variable of salience type revealed significant effects on Simplicity [$F(2,42)=13.141$, $p<.001$, $\eta_p^2=.098$], Functionality [$F(2,242)=9.42$, $p<.001$, $\eta_p^2=.072$], Usefulness [$F(2,242)=12.54$, $p<.001$, $\eta_p^2=.094$], Proficiency $F(2,242)=4.75$, $=.009$, $\eta_p^2=.038$], Satisfaction [$F(2,242)=6.56$, $p=.002$,

η_p^2 =.051], Behavior [F(2,242)=6.66, p=.002, η_p^2 =.052], Improvement [F(2,242)=9.36, p<.001, η_p^2 =.072], and Mental Model [F(2,242)=5.59, p=.004, η_p^2 =.044]. A significant main effect for salience type was found for the global usability score $F(2,121)=10.04$, $p < .001$, η_p^2 =.142.

Contrasts for the significant univariate tests for each dependent measure revealed that color and alphanumeric techniques received higher usability ratings compared to the baseline for the sub-dimensions of simplicity, functionality, usefulness, proficiency, satisfaction, behavior, improvement, mental model and the global usability score, $p < .05$. However, color and alphanumeric values were not significant different from each other in the usability measures. Table 28 the means and standard deviations for the usability measures for each within-subjects manipulation. See Figures 74-82 for visual representations.

Table 28. Means and (standard deviation) for usability dimensions by salience type.

Usability Measure	Baseline Mean (SD)	Color Mean (SD)	Alphanumeric Mean (SD)
Simplicity*	3.98 (0.79)	4.35 (0.71)	4.25 (0.67)
Functionality*	3.92 (0.62)	4.18 (0.63)	4.11 (0.64)
Usefulness*	3.65 (0.87)	4.01 (0.83)	3.98 (0.72)
Consistency	4.10 (0.60)	4.19 (0.74)	4.20 (0.59)
Proficiency*	4.03 (0.67)	4.21 (0.64)	4.14 (0.61)
Satisfaction*	3.44 (0.89)	3.74 (0.83)	3.65 (0.75)
Behavior*	3.90 (0.86)	4.17 (0.75)	4.12 (0.71)
Improvement*	3.31 (1.03)	3.74 (0.96)	3.66 (0.88)
Mental Model*	3.79 (0.78)	4.00 (0.73)	3.97 (0.73)
Global Usability*	3.79 (0.63)	4.06 (0.61)	4.01 (0.55)

*Denotes significance at $p < .05$

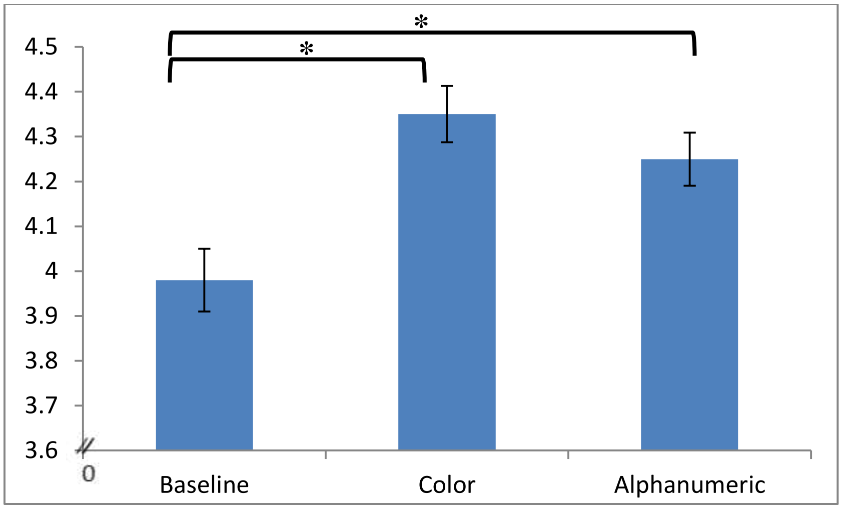


Figure 74. Simplicity by salience type.

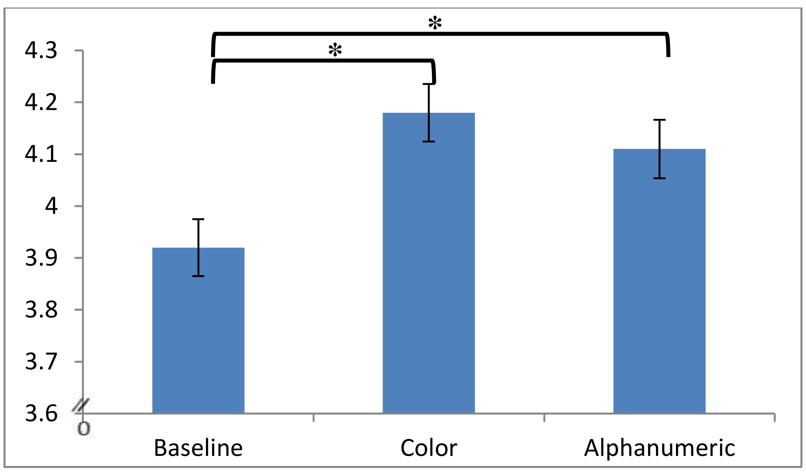


Figure 75. Functionality by salience type.

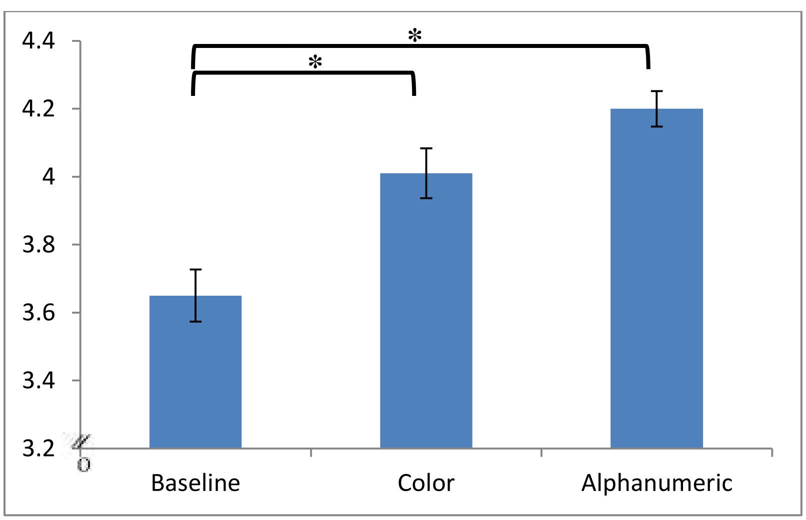


Figure 76. Usefulness by salience type.

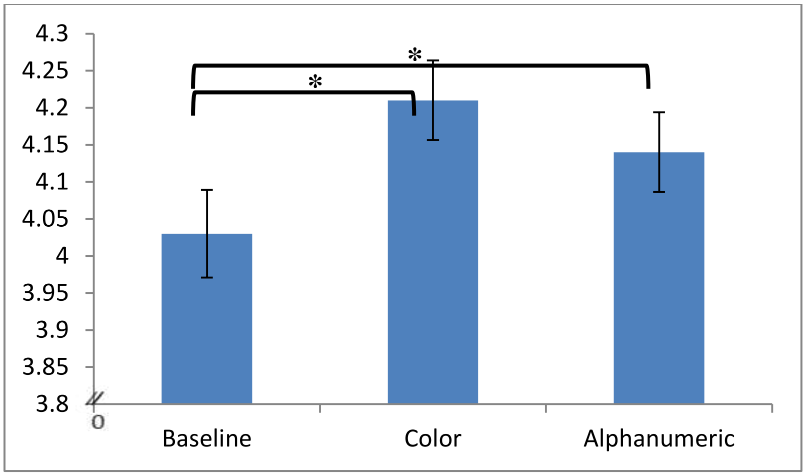


Figure 77. Proficiency by salience type.

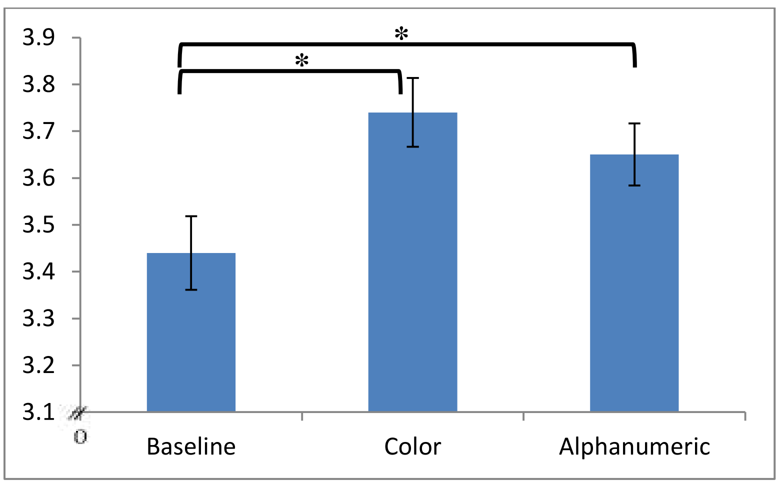


Figure 78. Satisfaction by salience type.

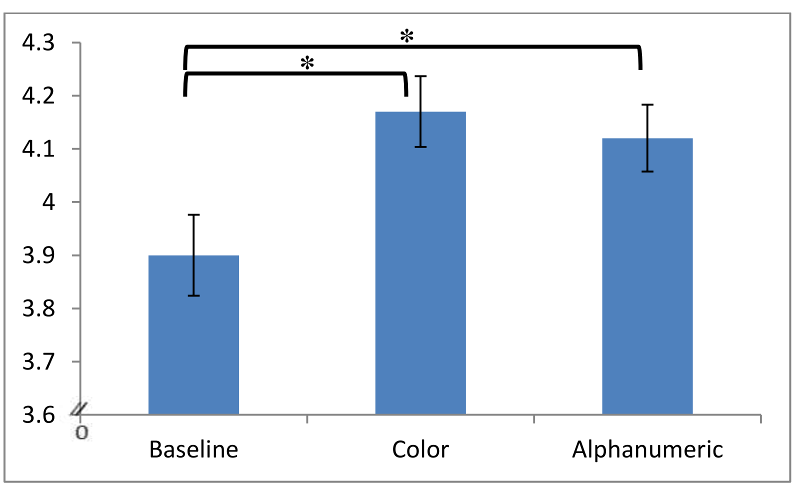


Figure 79. Behavior by salience type.

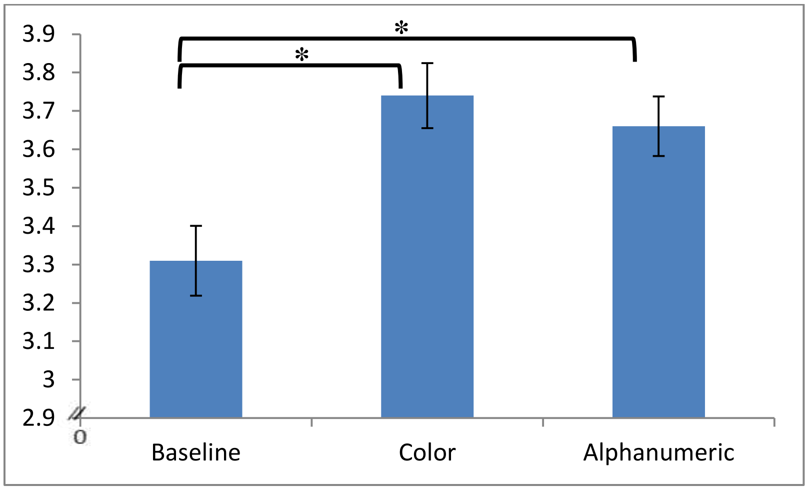


Figure 80. Improvement by salience type.

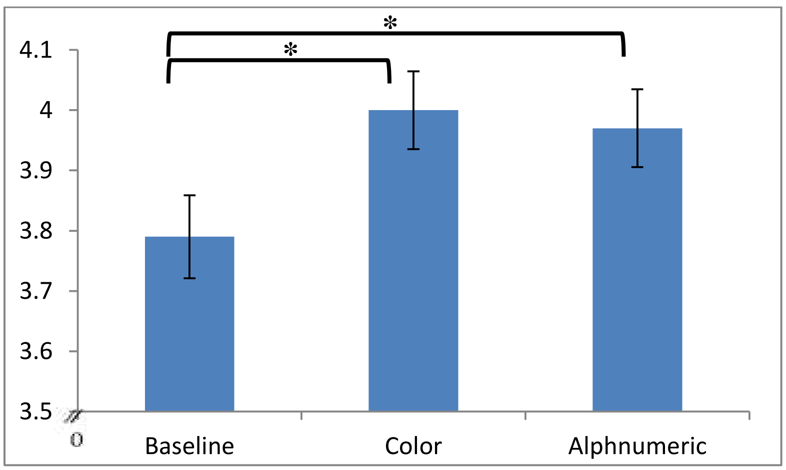


Figure 81. Mental model by salience type.

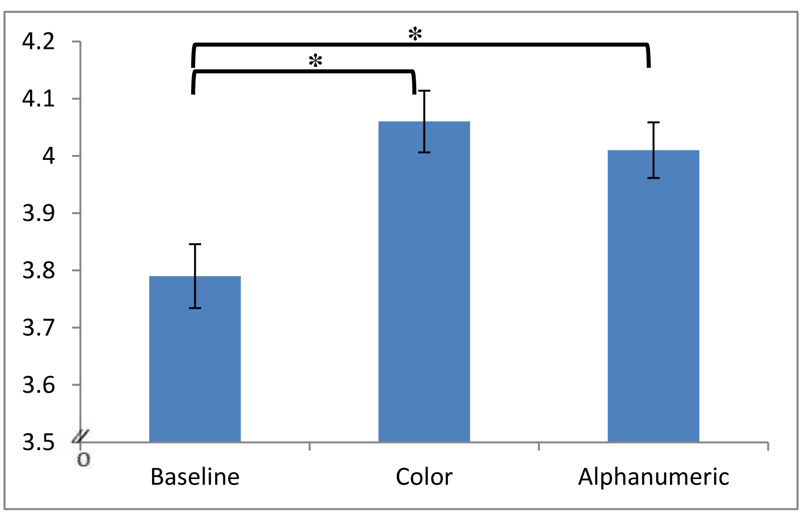


Figure 82. Global usability by salience type.

Individual Differences

Individual differences measures of MRT, TMT, LAG, and Gestalt were examined to determine if they had any relationship to the performance of data extraction above and beyond the experimental factors and effects reported in the data extraction performance section. Each measure was evaluated using the General Linear Model separately. A 4(salience application) X 3(salience type) mixed MANCOVA with the practice session and individual differences measures as the covariates was computed for accuracy and response time performance for the dependent measures of low-, mid-, high-level data extraction, and the remediation task. Interactions were computed with overall performance scores. Results of the analyses are listed in Table 29 and 30. If the omnibus multivariate result was not significant, the univariate result was not report or interpreted.

For accuracy performance, spatial ability had a significant impact on performance for high-level extraction and the remediation task. Subsequent correlations revealed that participants who were high on spatial ability also had higher accuracy scores for high-level extraction and the remediation task. TMT by salience type also had significant impact on low-level extraction. Subsequent correlations revealed that participants who were faster on the TMT were more accurate on low-level extraction, but only when using alphanumeric techniques. For response time performance, only working memory had significant omnibus effect, but subsequent univariate analysis did not reveal significant impact at any of the four dependent measures.

Table 29. Across study-Effects of individual differences on accuracy performance

Individual Differences Measure	Omnibus Multivariate Test	Performance Dependent Variables			
		Low	Mid	High	Remediation
MRT	$F(4,112)=1.40$, $p=.03$ $\eta_p^2=.089$	$F(1,115)=.63$, $p=.43$, $\eta_p^2=.005$	$F(1,115)=1.39$, $p=.24$, $\eta_p^2=.012$	$F(1,115)=6.70$, $p=.01$, $\eta_p^2=.055$ r=.35, $p<.001$	$F(1,115)=8.53$, $p=.004$, $\eta_p^2=.069$ r=.34, $p<.001$
MRT x Application	$F(12,342)=.67$, $p=.78$, $\eta_p^2=.023$	—	—	—	—
MRT x Salience	$F(8,108)=.90$, $p=.52$, $\eta_p^2=.062$	—	—	—	—
MRT x Application x Salience	$F(24,330)=1.28$, $p=.18$, $\eta_p^2=.085$	—	—	—	—
TMT	$F(4,113)=1.51$, $p=.20$, $\eta_p^2=.051$	—	—	—	—
TMT x Application	$F(12,345)=.71$, $p=.74$, $\eta_p^2=.024$	—	—	—	—
TMT x Salience	$F(8,109)=2.66$, $p=.01$, $\eta_p^2=.163$	$F(2,232)=3.63$, $p=.03$, $\eta_p^2=.030$ Base: r=-.02, $p=.84$ Color: r=-.10, $p=.28$ Alpha: r=-.21, $p=.02$	$F(2,232)=2.06$, $p=.13$, $\eta_p^2=.017$	$F(2,232)=3.02$, $p=.051$, $\eta_p^2=.025$	$F(2,232)=2.10$, $p=.13$, $\eta_p^2=.018$
TMT x Application x Salience	$F(24,333)=1.18$, $p=.26$, $\eta_p^2=.078$	—	—	—	—
LAG	$F(4,113)=.68$, $p=.61$, $\eta_p^2=.024$	—	—	—	—
LAG x Application	$F(12,345)=1.06$, $p=.39$, $\eta_p^2=.036$	—	—	—	—
LAG x Salience	$F(8,109)=.1.39$, $p=.21$, $\eta_p^2=.092$	—	—	—	—
LAG x Application x Salience	$F(24,333)=1.14$, $p=.30$, $\eta_p^2=.076$	—	—	—	—
Gestalt	$F(4,113)=.72$, $p=.58$ $\eta_p^2=.025$	—	—	—	—
Gestalt x Application	$F(12,345)=.72$, $p=.74$, $\eta_p^2=.024$	—	—	—	—
Gestalt x Salience	$F(8,109)=.75$, $p=.65$, $\eta_p^2=.052$	—	—	—	—
Gestalt x Application x Salience	$F(24,333)=1.26$, $p=.20$, $\eta_p^2=.083$	—	—	—	—

Table 30. Across study-Effects of individual differences on response time performance

Individual Differences Measure	Omnibus Multivariate Test	Performance Dependent Variables			
		Low	Mid	High	Remediation
MRT	$F(4,112)=2.00$, $p=.10$ $\eta_p^2=.067$	—	—	—	—
MRT x Application	$F(12,342)=.97$, $p=.48$, $\eta_p^2=.033$	—	—	—	—
MRT x Salience	$F(8,108)=.43$, $p=.90$, $\eta_p^2=.031$	—	—	—	—
MRT x Application x Salience	$F(24,342)=.62$, $p=.92$, $\eta_p^2=.043$	—	—	—	—
TMT	$F(4,113)=2.19$, $p=.08$ $\eta_p^2=.072$	—	—	—	—
TMT x Application	$F(12,345)=1.10$, $p=.36$, $\eta_p^2=.037$	—	—	—	—
TMT x Salience	$F(8,109)=.70$, $p=.68$, $\eta_p^2=.049$	—	—	—	—
TMT x Application x Salience	$F(24,333)=1.44$, $p=.09$, $\eta_p^2=.094$	—	—	—	—
LAG	$F(4,113)=2.80$, $p=.03$ $\eta_p^2=.090$	$F(1,116)=2.10$, $p=.15$, $\eta_p^2=.018$	$F(1,116)=.06$, $p=.80$, $\eta_p^2=.001$	$F(1,116)=2.93$, $p=.09$, $\eta_p^2=.025$	$F(1,116)=.00$, $p=.99$, $\eta_p^2<.001$
LAG x Application	$F(12,345)=.98$, $p=..47$, $\eta_p^2=.033$	—	—	—	—
LAG x Salience	$F(8,109)=.40$, $p=.92$, $\eta_p^2=.029$	—	—	—	—
LAG x Application x Salience	$F(24,333)=.66$, $p=.89$, $\eta_p^2=.045$	—	—	—	—
Gestalt	$F(4,113)=1.19$, $p=.32$ $\eta_p^2=.040$	—	—	—	—
Gestalt x Application	$F(12,345)=.39$, $p=.97$, $\eta_p^2=.013$	—	—	—	—
Gestalt x Salience	$F(8,109)=1.14$, $p=.34$, $\eta_p^2=.077$	—	—	—	—
Gestalt x Application x Salience	$F(24,333)=.70$, $p=.85$, $\eta_p^2=.048$	—	—	—	—

Exploratory analyses

Exploratory analyses were conducted to determine if the experimental manipulations affected each task differently. A 4(salience application) X 4(task) X 3(salience type) mixed model ANCOVA was conducted for accuracy performance scores with the practices session as the covariate. The practice session covariate was significant, $F(1,123)=49.35$, $p<.001$, $\eta_p^2 =.286$. There was a significant main effect for salience type, $F(2,122)=25.435$, $p<.001$, $\eta_p^2 =.294$. There was also a significant two way interaction of task and salience type on accuracy performance, $F(6,118)=4.04$, $p=.001$, $\eta_p^2 =.170$. Finally there was a significant three way interaction of task, salience type, and salience application on accuracy performance, $F(18,360)=2.68$, $p<.001$, $\eta_p^2 =.118$. Only the three way interaction will be reported and interpreted as it provides the results of interest.

In low application, salience techniques had significant impact on high-level data extraction [$F(1,122)=9.56$, $p<.001$, $\eta_p^2 =.135$] and the remediation task [$F(1,122)=3.96$, $p=.021$, $\eta_p^2 =.061$]. See Figure 83 for a visual representation. Trends indicate that although salience techniques had impact on tasks, the two salience techniques did not impact task significantly different, with the exception of color, which was superior for high-level data extraction.

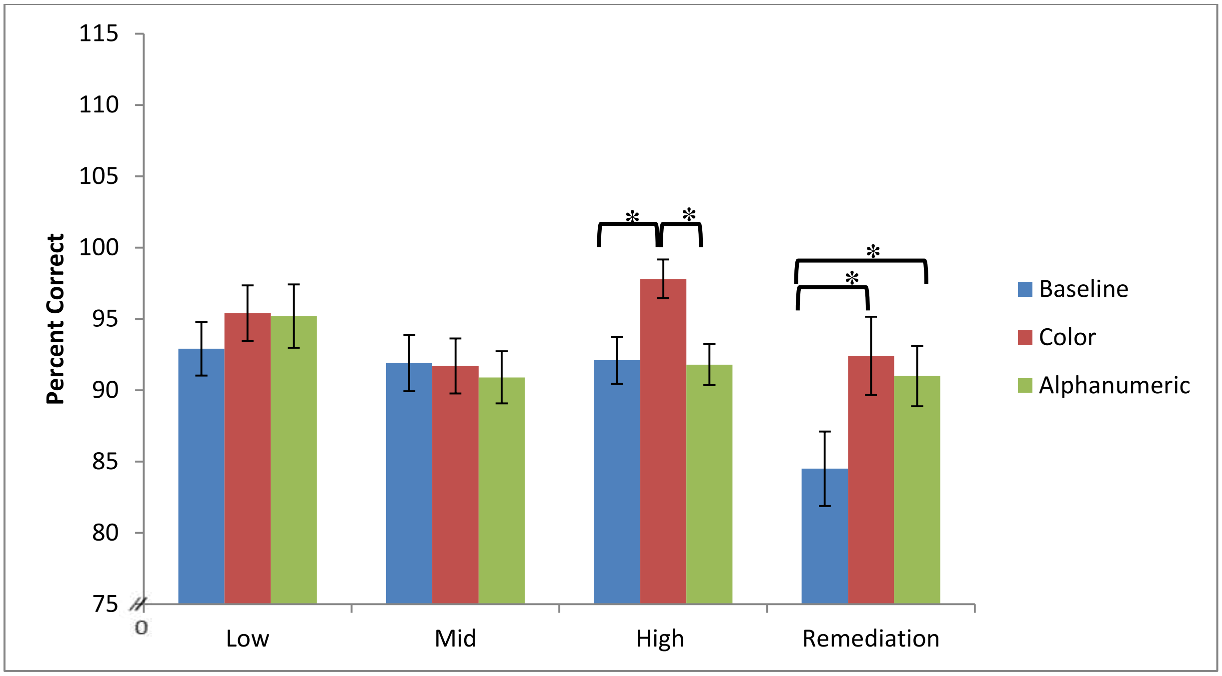


Figure 83. Accuracy performance on low application by salience type and task.

In mid-level application, salience techniques had significant impact on high-level data extraction, $F(1,122)=8.91$, $p<.001$, $\eta_p^2=.127$. Trends indicate that, like low application, although salience techniques had impact on tasks, the two salience techniques did not impact task significantly different, with the exception of color, which was superior for high-level data extraction. See Figure 84 for a visual representation.

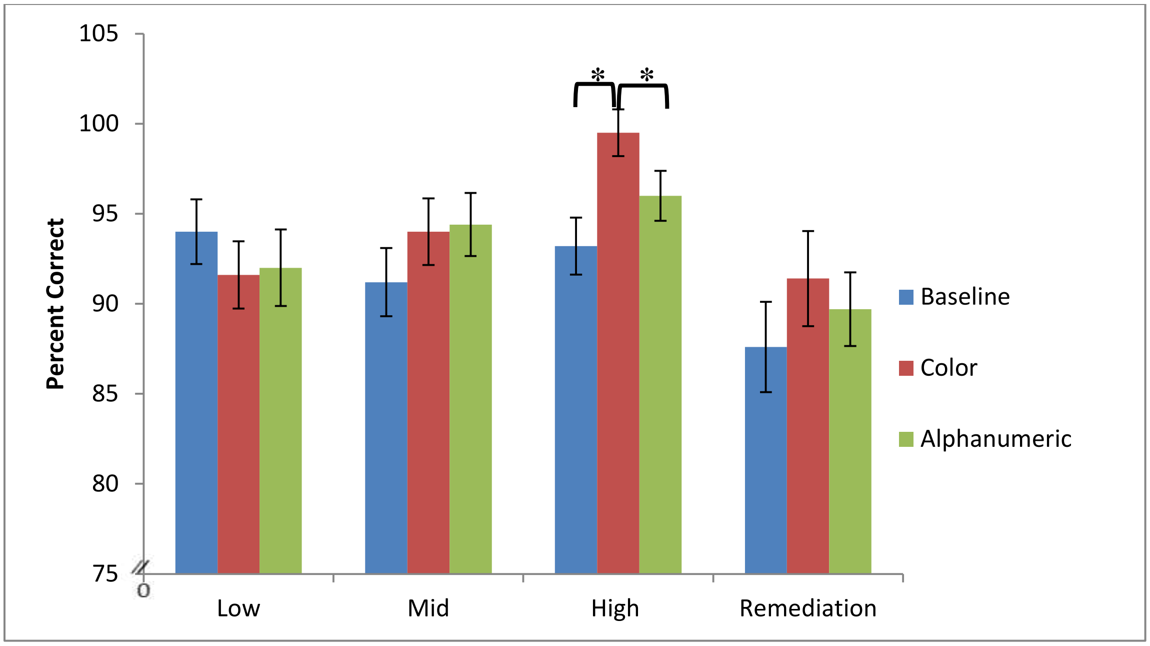


Figure 84. Accuracy performance on mid application by salience type and task.

In same application, salience techniques had significant impact on mid-level data extraction [$F(1,122)=6.90$, $p=.001$, $\eta_p^2=.102$] and high-level data extraction [$F(1,122)=4.39$, $p=.014$, $\eta_p^2=.067$]. Trends indicate that although salience techniques had impact on tasks, the two salience techniques did not impact task significantly different. Salience techniques increased accuracy scores, with the exception for the low-level data extraction task. Salience techniques had the opposite effect on accuracy scores for low-level data extraction, however, this impact was not statically significant, $p>.05$. See Figure 85 for a visual representation.

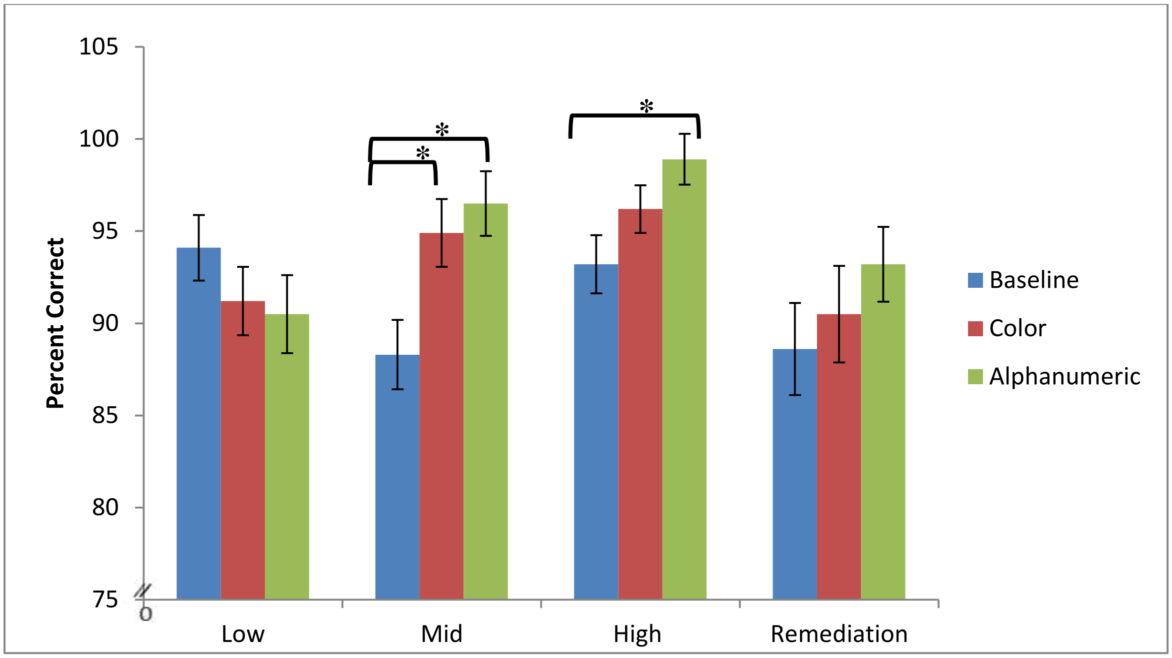


Figure 85. Accuracy performance on same application by salience type and task

In different application, salience techniques had significant impact on low-level extraction [$F(1,122)=4.15$, $p=.018$, $\eta_p^2=.064$], mid-level extraction [$F(1,122)=6.42$, $p=.002$, $\eta_p^2=.095$], and the remediation task [$F(1,122)=3.17$, $p=.045$, $\eta_p^2=.049$]. Salience did not significantly impact high-level data extraction, but was reaching significance [$F(1,122)=2.67$, $p=.074$, $\eta_p^2=.042$]. Trends indicate that the different alphanumeric salience technique combination improved accuracy performance the most, with the exception of the low-level data extraction task where the opposite was found. See Figure 86 for a visual representation.

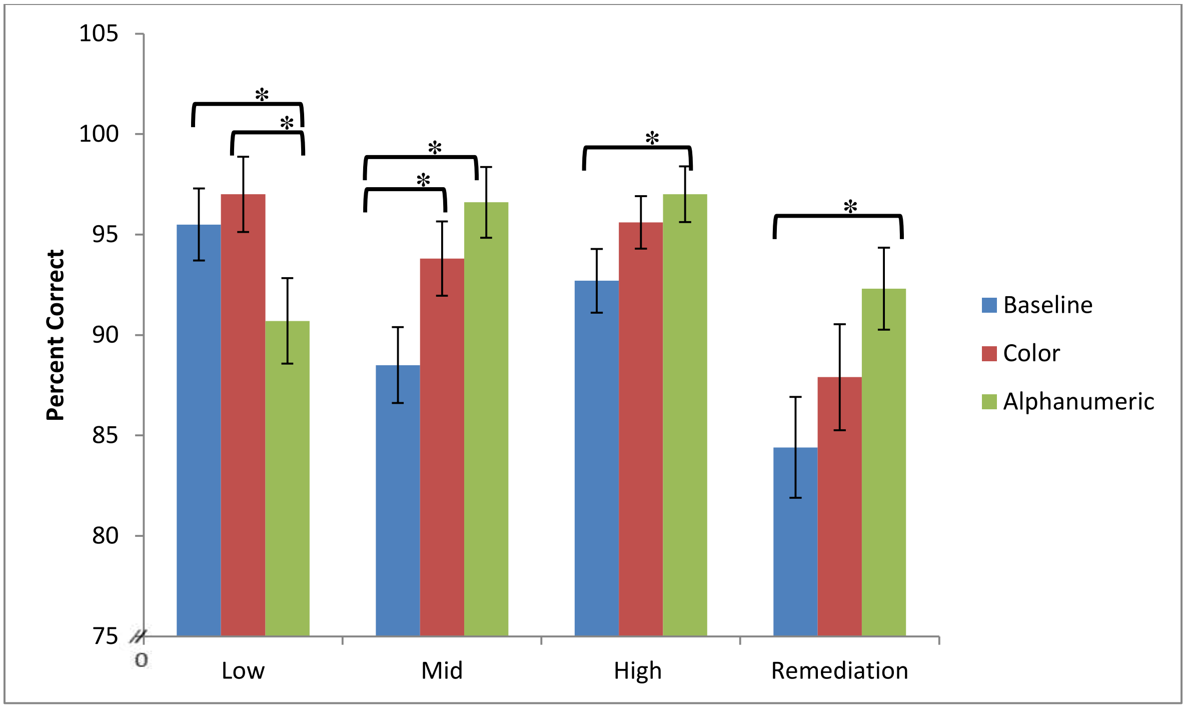


Figure 86. Accuracy performance for different application by salience type and task.

A 4(salience application) X 4(task) X 3(salience type) mixed model ANCOVA was also conducted for response time performance scores with the practices session as the covariate. The practice session response time covariate was significant, $F(1,123)=105.526$, $p<.001$, $\eta_p^2=.462$. There was a significant main effect for salience type [$F(2,122)=25.435$, $p<.001$, $\eta_p^2=.294$], salience application [$F(3,123)=5.65$, $p=.001$, $\eta_p^2=.121$], and task [$F(3,121)=5.22$, $p=.002$, $\eta_p^2=.115$]. There was also a significant two way interaction of task and salience type [$F(6,118)=3.63$, $p=.002$, $\eta_p^2=.156$] and salience type by salience application [$F(6,246)=4.28$, $p<.001$, $\eta_p^2=.095$] on response time performance. Finally there was a significant three way interaction of task, salience type, and salience application on response time performance, $F(18,360)=2.30$, $p=.002$, $\eta_p^2=.103$. Only the three way interaction will be reported and interpreted as it provides the results of interest.

In low application, salience techniques had significant impact on low-level data extraction [$F(2,122)=4.41$, $p=.014$, $\eta_p^2=.135$], high-level data extraction [$F(2,122)=10.27$, $p<.001$, $\eta_p^2=.144$], and the remediation task [$F(2,122)=7.36$, $p<.001$, $\eta_p^2=.120$] for response times. See Figure 87 for a visual representation. Trends indicate that color techniques mapped to low-level data tend to be superior in reducing response times for high-level data extraction and the remediation task but not significantly better for low and mid-level data extraction.

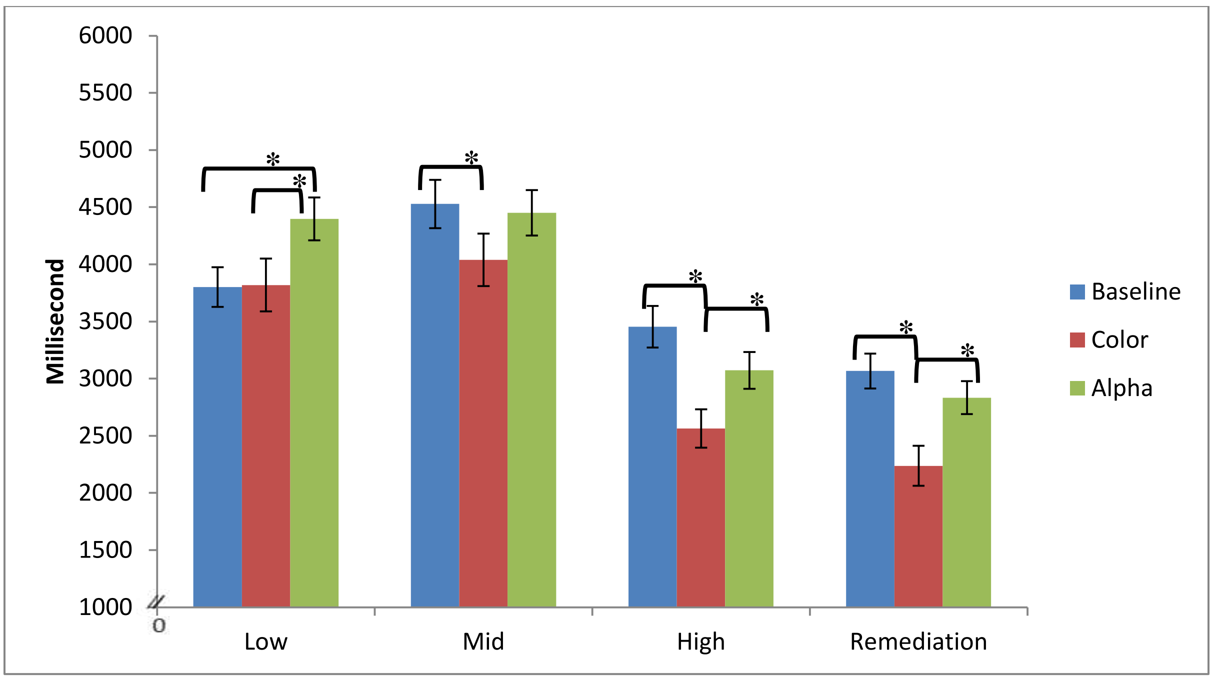


Figure 87. Response time performance on low application by salience type and task.

In mid application, salience techniques also had significant impact on low-level data extraction [$F(2,122)=7.36$, $p=.001$, $\eta_p^2=.120$], high-level data extraction [$F(2,122)=12.92$, $p<.001$, $\eta_p^2=.175$], and the remediation task [$F(2,122)=5.67$, $p=.004$, $\eta_p^2=.085$] for response times. See Figure 88 for a visual representation. Trends indicate that, again, color techniques mapped to mid-level data tend to be superior in reducing response times for high-level data extraction and the remediation task but not significantly better for low and mid-level data extraction when compared to alphanumeric techniques mapped to mid-level data.

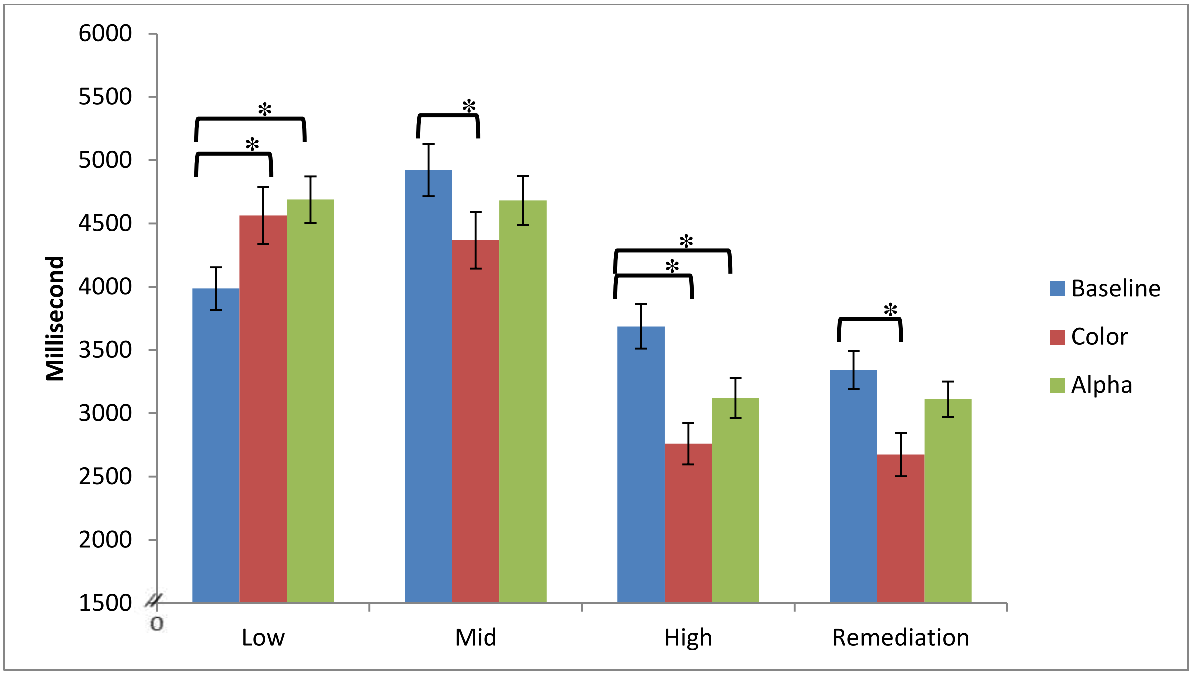


Figure 88. Response time performance on mid application by salience type and task.

In same application, salience techniques also had significant impact on mid-level data extraction [$F(2,122)=5.94$, $p=.003$, $\eta_p^2=.089$], high-level data extraction [$F(2,122)=18.27$, $p<.001$, $\eta_p^2=.230$], and the remediation task [$F(2,122)=18.06$, $p<.001$, $\eta_p^2=.228$] for response times. See Figure 89 for a visual representation. Trends indicate that color techniques mapped to low-level data tend to be superior in reducing response times for mid-level data extraction, high-level data extraction, and the remediation task but not significantly better for low data extraction.

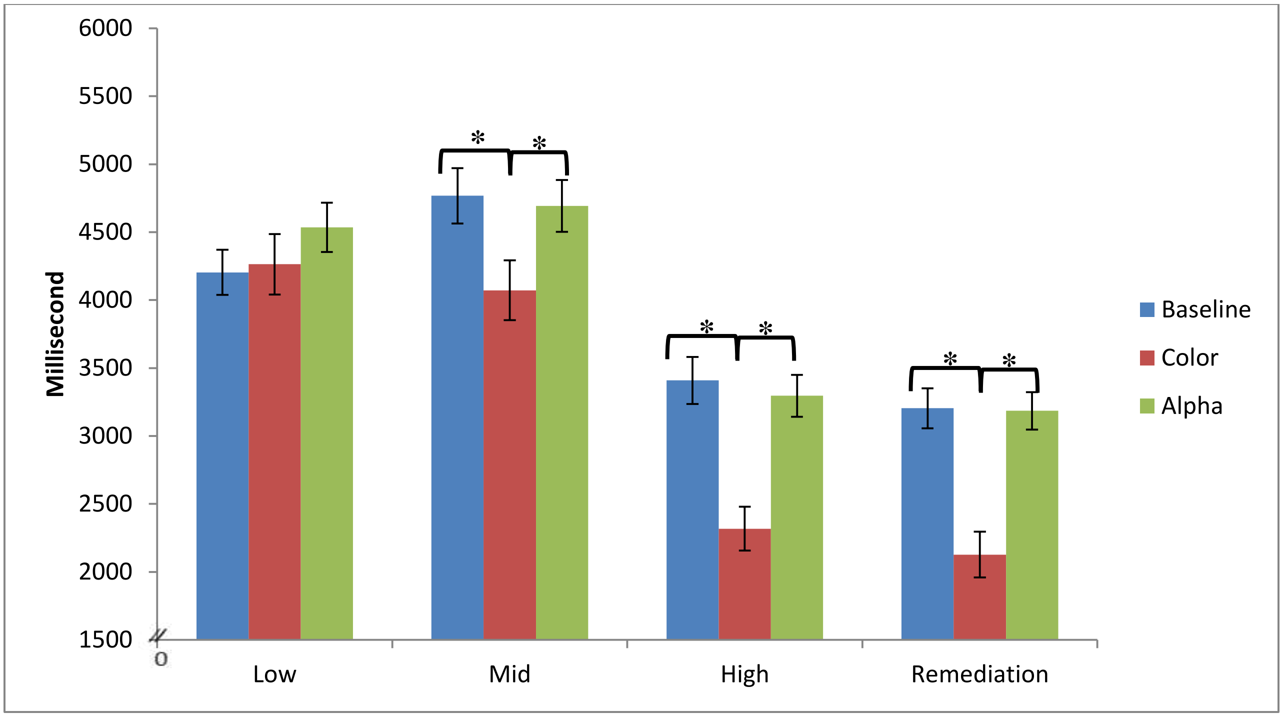


Figure 89. Response time performance on same application by salience type and task.

In different application, salience techniques also had significant impact on high-level data extraction [$F(2,122)=25.92$, $p<.001$, $\eta_p^2=.298$] and the remediation task [$F(2,122)=28.69$, $p<.001$, $\eta_p^2=.320$] for response times. See Figure 90 for a visual representation. Trends indicate that salience techniques at the different application generally reduced response times for the four tasks. Specifically, alphanumeric techniques in the different application were superior for the remediation task.

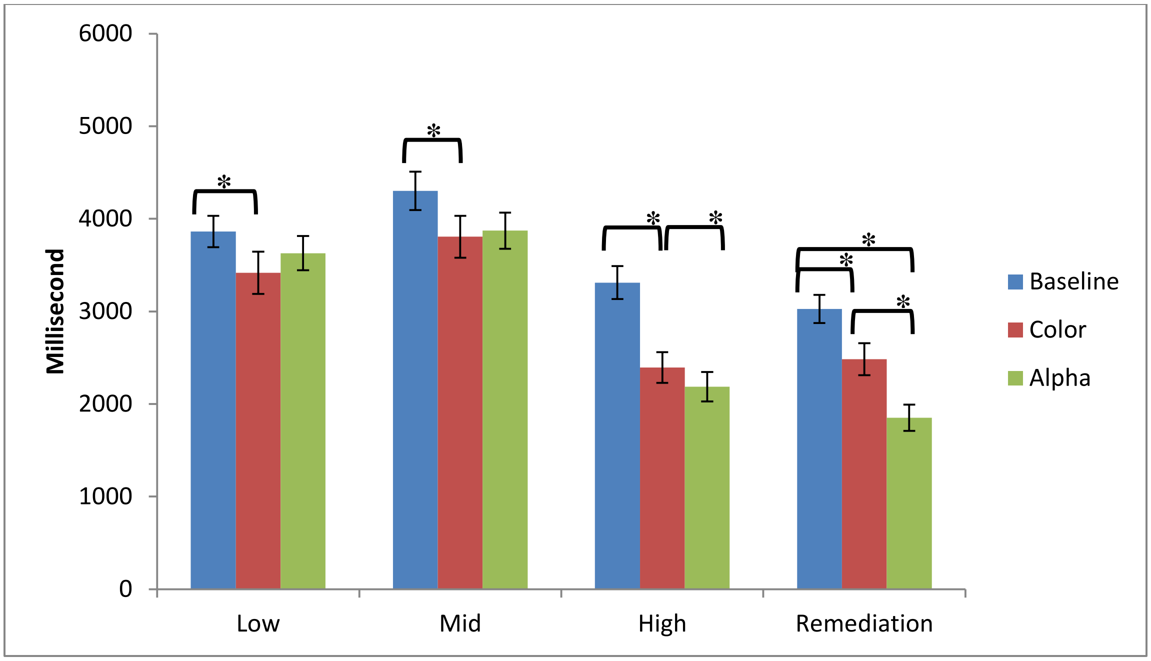


Figure 90. Response time performance on different application by salience type and task.

Summary of Across Study Results

Hypotheses were made for across the two studies to compare salience application and salience type on data extraction performance. Hypothesis 7a predicted that salience techniques to two data levels will improve response time and accuracy performance of multi-level data extraction compared to mapping one salience techniques to the display. This was not supported. No significant difference was found between having two salience techniques verses one salience technique for accuracy scores and response time. Number of salience technique did interact with salience type for mid and high-levels of data extraction on accuracy performance scores. For mid-level data extraction, trends show that two salience techniques with alphanumeric values generally accuracy scores the most compared to one salience technique. In high-level data extraction, both one color technique and two alphanumeric techniques resulted in higher accuracy performance. Although two salience techniques did increase accuracy scores for some levels of data extraction, this was not unanimous for all the experimental tasks.

Hypothesis 7b predicted that mapping different salience techniques to two levels of data would improve response time and accuracy performance of multi-level data extraction compared to mapping one salience techniques to the display. This was partially supported. Although this effect was not found with accuracy measures, response time performance measures did reveal differences. The different application condition resulted in significantly faster response times for low-, mid-, high-level extraction, and the remediation task.

Hypothesis 7c predicted that mapping alphanumeric techniques to low-level data and colors techniques to mid-level data will improve response time and accuracy performance of multi-level data extraction compared mapping one salience techniques to the display. This was partially supported. This specific interaction effect was not found in the accuracy measures. An interaction effect for high-level data extraction accuracy was found, but color in the mid condition and alphanumeric values in the same condition resulted in the highest accuracy scores, not the hypothesized combination. For response time performance, alphanumeric techniques in the different condition did result in significantly faster response time for high-level data extraction and the remediation task. However, other combinations were also successful in reducing response time. For high-level data extraction, color in the same condition and color in the different condition had faster response times. For the remediation task, color in the same condition reduced response times.

In subjective workload measures, no significant interactions of application and salience type were found. However, a significant main effect for salience type was found. Although both color and alphanumeric techniques generally decreased subjective workload and increased subjective performance, the two salience techniques were not significant from each other. This suggests that adding salience was better than no added salience but the specific type and combinations of salience techniques did not significantly impact subjective workload. However, in general, the subjective workload ratings overall were on the low end, indicating that the tasks were not challenging enough. Additionally, only significant main effect for salience type was found for subjective usability ratings. Trends show that although color and alphanumeric techniques were more highly rated in terms of usability compared to the baseline, they were not significantly different from each other.

Across study discussion

The purpose of the across study analyses was to determine how mapped salience techniques to one data level or combination salience techniques mapped to two data levels differ, and which was superior. Results from the across study analyses indicate that adding more salience is not better than less salience, and salience does not automatically improved performance. Instead, it is salience that is correctly mapped to domain tasks which improves performance the most. This finding is consistent to that from other studies examining salience and mapping (i.e. Bennett et al., 2000; Bennett & Walters, 2001).

Table 31 shows the salience combinations that lead to improvements in performance (i.e. higher accuracy or faster response time). As can be seen from the table, there was not just one salience technique or combination that was superior compared to others. In fact, for all tasks where significant interactions of salience application and salience type were found, at least two combinations were found to significantly improve performance. This suggests that when mapping salience techniques to domain tasks, there is more than one salience technique or salience combinations that improve performance. The possibility of multiple well-mapped salience techniques is consistent with findings of previous studies (Bennett et al., 2000; Bennett

& Walters, 2001). However, previous research for focused attention tasks did find one salience technique that was superior to others. Perhaps the complexity of a hierarchal domain mapped to the configural displays allows for more salience combinations that would improve performance. Additional research is needed to investigate if hierarchical domains really can have multiple well-mapped salience combinations or if one superior combination can be found.

Table 31. Significant impact of salience combinations on accuracy and response time

	Task	**Color at Low**	**Alpha at Low**	**Color at Mid**	**Alpha at Mid**	**Color at Low; Color at mid**	**Alpha at Low; Alpha at Mid**	**Color at Low; Alpha at Mid**	**Alpha at Low, color at Mid**
Accuracy	Low								
	Mid								
	High	✓		✓			✓		✓
	Remediation								
Response Time	Low							✓	✓
	Mid								
	High	✓		✓		✓		✓	✓
	Remediation	✓				✓			✓

In low-level extraction response time, the different combination salience techniques were successful in reducing response times. However, it is surprising that the salience techniques mapped individually did not show this effect. This result suggests that addition of mapped salience to other data-levels on the displays provided some transfer effect. Indeed, no salience techniques were mapped to high-level data on the display, however, results tells us that mapping salience techniques to low- and mid-levels of data on the display supported high-level extraction. This finding is consistent to previous studies, where they also found transfer effects from salience techniques mapped to the display.

Although multiple salience combinations were found to improve performance (see Table 31), the alphanumeric values at low-level data and color at mid-level data salience combination does reappear for each task measurement where significant differences were found. This suggests that this combination would be best suited for supporting SBT instructors, when hierarchical domain tasks are collectively taken into account. To further add support for this specific salience combination, Figure 91 shows the interaction effect of salience application and salience for mid-level data extraction accuracy performance. Although this interaction effect was not statistically significant at a $p < .05$, it was close to significance at $p = .066$. For mid-level data extraction accuracy performance, again multiple salience combinations were superior in increasing accuracy scores, but among those is the combination of alphanumeric values mapped to low-level data and color mapped to mid-level data. This adds additional support for the use of this specific salience combination for the SBT instructor domain.

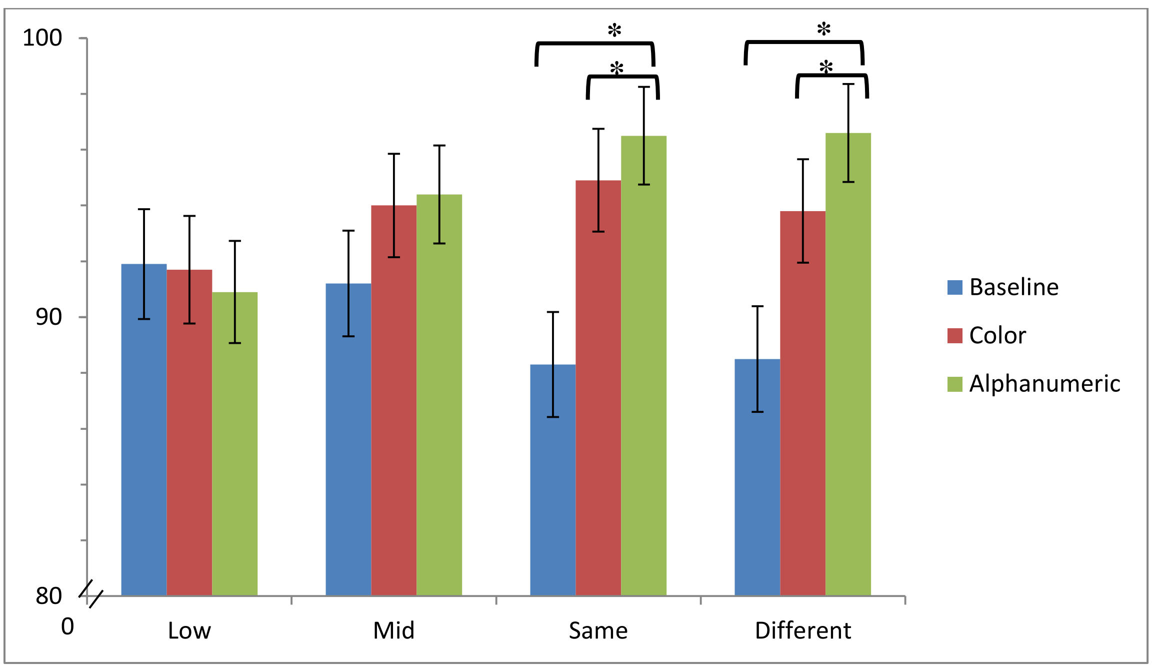


Figure 91. Mid-level data extraction accuracy performance

In general, significant differences were found for response times but not accuracy performance, which the exception of high-level data extraction. This may be due to the task not being challenging enough, as indicated by the low workload scores across all displays. However, it is not because the salience techniques and combinations didn't have any effect on accuracy scores. The exploratory analyses show us that salience techniques did impact task accuracy. In fact, the different salience combinations had the most impact on accuracy scores. However, this specific combination, although was significantly different from the baseline, was perhaps just not different enough from other salience combinations for some accuracy scores. In a more complex environment, differences in accuracy scores may have been found.

The less complex environment used in this research may have also been the reason why generally no significant impacts were found with the individual differences measures. Previous research has shown where working memory had significant impact on task performance but only in displays with complex tasks (Lohse, 2010). Therefore, individual differences are still important considerations, but may not be a significant impact in environments that are less challenging. In more complex environments, these individual differences may have a greater impact and be examined to see how these individual differences impact multi-level data extraction via configural displays.

Although the different alphanumeric salience technique combination was found to be a superior combination for most response time measures of each task, and some accuracy measures (e.g. high-level data extraction, and to some extent, mid-level data extraction), the lack of findings in other measures, and the decrease in accuracy performance for low-level data extraction, suggests that the mapping of these salience techniques is not perfect. Perhaps there are better salience combinations that will improve accuracy and response time measures for all tasks. Indeed, the lack of findings in subjective usability suggests that the mappings, although did show differences, were not different enough to impact perceived usability. Perhaps with the addition of a challenging environment and/or different mappings, greater results can be seen. The findings of this dissertation provide a start to understanding mappings for hierarchical data extraction, and more research is needed to examine the impact of different salience mappings on multi-level data extraction.

CHAPTER 6: CONCLUSIONS

The objective of this dissertation was to determine whether specific mappings of salience techniques would improve accuracy and response time performance to address the potential performance cost of configural displays for hierarchical domains. A handful of the hypotheses were supported or partially supported. Results were generally found for response time performance and not for accuracy performance. Results imply that generally different salience combinations improved performance more than others. Also, multiple salience techniques and combinations were found to significantly improve performance, but the combination of alphanumeric values at low-level data and color at mid-level data appeared to be a significant combination for tasks. This suggest the for the SBT instructor domain, this particular combination is best suited, but additional research is need to confirm this. The lack of significant findings from the accuracy measures and the lack of one superior salience combination indicate that, although these studies provide a good start in addressing the potential performance cost of configural displays for hierarchical domains, it's not perfect. Additional research in other salience combinations and the use of this display in a more complex environment would unveil a greater understanding for display design of hierarchical domains.

Theoretical Implications

The current research has theoretical implications on the application of salience techniques to domains containing multiple levels of data. The findings of this dissertation provide a first step in understanding how hierarchical domain data should be mapped to the display. The results contribute to the literature on the principle of semantic mapping for display design and expand on it. Previous research has applied salience techniques to domains with two levels of data to address the potential performance cost of configural displays. The current studies show that potential performance costs of configural displays exists with hierarchal domains. In general low and mid-level extraction had lower performance compared to high-level extraction. This suggests that the potential cost to performance that is associated with configural displays is not limited to only low-level extraction (or focused attention tasks), and that in hieratical domains that are mapped to configural displays, performance costs exist in data extraction at levels that are lower than the highest level of data.

The existence of potential performance costs in configural displays for hierarchical domains supports the theory of emergent features (Pomerantz et. al., 1977). Salient emergent features are created through the interaction of the individual parts, and in the configural display many emergent features arise. The polar coordinate configural display used in this research demonstrated the nested levels of emergent features and its salience relative to other emergent features in the display. Due to the polar coordinate's inherent nested emergent features, it was thought that this display would map well with hierarchical domains that contained nested levels of data. This research shows the polar coordinate display was suitable for hierarchical domain data (e.g. extraction performances was generally high), but additional research comparing the polar coordinate display with other types of configural displays is needed in order to determine if a particular type of configural display is superior in mapping of hierarchical domain data.

This dissertation further expands on the theory of mapping techniques to display and task constraints to multi-level domains (Bennett & Flach, 2011). It extends upon previous research findings from Bennett and colleagues (2000; 2001) on the impact of salience data extraction via configural displays. The research studies leveraged salience techniques used in previous studies for domains where two data levels were explored and applied those to a hierarchical domain in which three data levels were examined. Findings of this research effort were consistent with the theory of semantic mapping and show that this approach can be used for hierarchical domains. This also research expands previous research on mappings in mapping multiple salience techniques to the display. Findings of the combinations of previously examined salience techniques (see Bennett et al., 2000; Bennett & Walters, 2001) provided a better understanding of how users respond to salience techniques and combinations of salience techniques and which ones are better in supporting multi-level data extraction.

Additionally, dimensions of mapping were manipulated in this research. Particularly, what was mapped and where it was mapped were manipulated. Results of the studies show that the mappings interacted to show specific combinations that were superior in supporting data extraction performance. This indicates that mapping should not be considered as a singular construct, but as a multi-dimensional construct. Mappings should be considered at the sub-dimension level to effectively support extraction performance.

This research adds to the literature of individual differences in display design. Few studies have previously included these measures, and none have examined the impact of individual differences configural displays for hierarchal domains. Although very few significant differences were found for the individual differences measures, this dissertation provides a start for how individual differences may impact data extraction performance for hierarchical domains.

The salience technique combinations used attempted to address the potential cost of configural display for hierarchal domain data. While these combinations did not find improvements in performance for data extraction at all data levels, it was a first step in understanding the mapping of multiple salience techniques to a hierarchical work domain to support multi-level data extraction.

Practical Implications

The findings from this dissertation also have practice implications. Design recommendations can be drawn from these research studies. First, more salience does not necessarily improve performance. The findings of this research are consistent with previously literature in which adding more salience does not improve performance. It is the specific mappings of salience techniques that are important. Second, when using salience combinations, the findings suggest that different salience combinations better support response time performance of multi-level data extraction for hierarchical domains. In general, if more than on salience technique is used, varying the different types of salience would improve performance more than repeatedly using the same salience technique. Third, results show that there are many salience techniques and combinations that improve performance. The specific combinations would depend on the domain constraints, and research would be needed in order to determine which specific combination would be best suited. Specifically, dimensions of mapping should be considered. Fourth, the results have direct implication for designing displays for the SBT instructor domain. The configural display used this research adopted the “at a glance” use case environment where instructors would have to quickly extraction information to make remediations. For this environment, the combination of alphanumeric values mapped to low-level data and color mapped to mid-level data improved performance of multi-level data extraction. Response time performance for most data levels and high-level extraction accuracy performance was improved with this specific salience combination. With additional research, confirmation can be made whether this particular salience combination is superior for improving SBT instructor performance.

Although the SBT instructor domain was used, and study results show which specific salience combination improves performance for this domain, the findings can be applied to other domains. If other domains have similar domain and task constraints, the findings in this research would apply. However, domain analysis would first need to be conducted to determine what the domain constraints are. Then, if they are the same or very close to the SBT instructor domain, the salience techniques found to improve performance in this study could apply. Additionally, the findings of this research also have implication s for other domains. Better displays can support operator tasks, which would in turn increase training effective, performance assessment, and safety. With displays that are deigned better, operator tasks can be completed fore effectively and efficiently, and lead to increase in performance of overall tasks.

This research did find some effects of individual differences on display design. Although a majority of the individual differences measures were not significant, a few were found to have significant impact of data extraction performance. The lack of findings is possibility due to the nature of the tasks not being complex enough, and not because there were no individual differences. However, the significant findings of individual differences measures supports previous literature in that these factors should be considered in display design. In fact, since a few individual differences were found to have significant impact even when the task were challenging, suggests these individual differences would have an even greatly affect in a more complex environment. Therefore, considerations for individual differences should be made when designing displays.

With the understanding of the limitations in the studies, the process of semantic mapping was demonstrated to be effective in supporting some data-level extractions from a hierarchical domain. Additional research is needed to examine the use of semantic mapping for hierarchical domains; however, results generally imply that the process can be used for other hierarchical domains to address the potential performance cost of configural displays.

Study Limitations and Future Research

The studies conducted in this dissertation were done in a controlled, laboratory environment where the only task of the participant was to view the stimulus display and complete the task of extracting information. However, in a more realistic SBT instructor setting, the instructor has many more tasks than just extracting information and deciding when to provide remediation. Additionally, they would have more displays to monitor (e.g. map display and chat), and instructors would actually be providing the remediation either verbally or through a remediation window. In a more complex environment, it is unknown how these salience combinations would impact these instructor tasks. I would hypothesize that since data extraction is used in these instructor tasks, salience technique combinations that result in more accurate and faster data extraction would also lead to more accurate and faster completion of these instructor tasks. However, more research in how the salience technique combination examined in this dissertation would impact instructor tasks in a more complex environment is needed.

Results indicated that although some improvements to accuracy and response time performance was found with some salience technique combinations for some tasks, a lack of significant improvements for other tasks was also found. Additionally, the exploratory results for the response time measures indicate that although salience technique combinations make significant impact for each task, low-level data extraction and mid-level data extraction were generally slower than high-level data extraction and the remediation task. This suggests that although the different alphanumeric salience combination is a good start to addressing the potential cost of configural displays for hierarchical data extraction, other mapped salience combinations may be better. Future research should compare the superior salience combinations found from these studies to other mapped salience techniques to determine if all accuracy and response time performances for all tasks can be improved.

APPENDIX A: UCF IRB APPROVAL LETTER

University of Central Florida

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Audrey W. Fok**

Date: **June 24, 2015**

Dear Researcher:

On 6/24/2015, the IRB approved the following human participant research until 06/23/2016 inclusive:

Type of Review:	UCF Initial Review Submission Form
Project Title:	CONFIGURAL DISPLAYS: THE EFFECTS OF SALIENCE ON MULTI-LEVEL DATA EXTRACTION
Investigator:	Audrey W Fok
IRB Number:	SBE-15-11360
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 06/23/2016, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Joanne Muratori

Signature applied by Joanne Muratori on 06/24/2015 10:07:01 AM EDT

IRB manager

APPENDIX B: DEMOGRAPHICS QUESTIONNAIRE

Demographics

1. Age______________________
2. Biological Sex : _____Male _____ Female
3. Gender you identify with : _____Male _____ Female
4. Handedness: _____LEFT ______RIGHT
5. Year in School:

   ____Freshman ____Sophomore ____Junior ____Senior

   ____Graduate
6. How experienced are you with computers?

   ___ None ___ Novice ___ Average ___ Very Knowledgeable ___ Expert
7. How many hours to you typically use the computer per week?

   ____0-5 hours

   ____6-10 hours

   ____10-15 hours

   _____16-20 hours

   _____20-25 hours

   _____26-30 hours

   _____ Over 30 hours
8. How experienced are you with mobile devices (e.g. smartphones, tablets, iPads)?

   ___ None ___ Novice ___ Average ___ Very Knowledgeable ___ Expert
9. How many hours to you typically use mobile devices per week?

   ____0-5 hours

   ____6-10 hours

   ____10-15 hours

   _____16-20 hours

   _____20-25 hours

   _____26-30 hours

   _____ Over 30 hours

10. Do you currently or have previously served in the military? YES   NO
11. Have you had any military training?   YES   NO
    a. If yes, how many months? _______________
12. Have you had any computer-based military training?   YES   NO
    a. If yes, how experienced are you with computer-based military training?

___ None   ___ Novice   ___ Average   ___ Very Knowledgeable   ___ Expert

13. Do you have any previous experience with Scenario-Based Training (SBT)?   YES   NO
    a. If you answered "YES" on question 13, please rate how experienced are you with Scenario Based Training (SBT)?

___ None   ___ Novice   ___ Average   ___ Very Knowledgeable   ___ Expert

    b. If you answered "YES" on question 13, please describe your SBT experience.

____

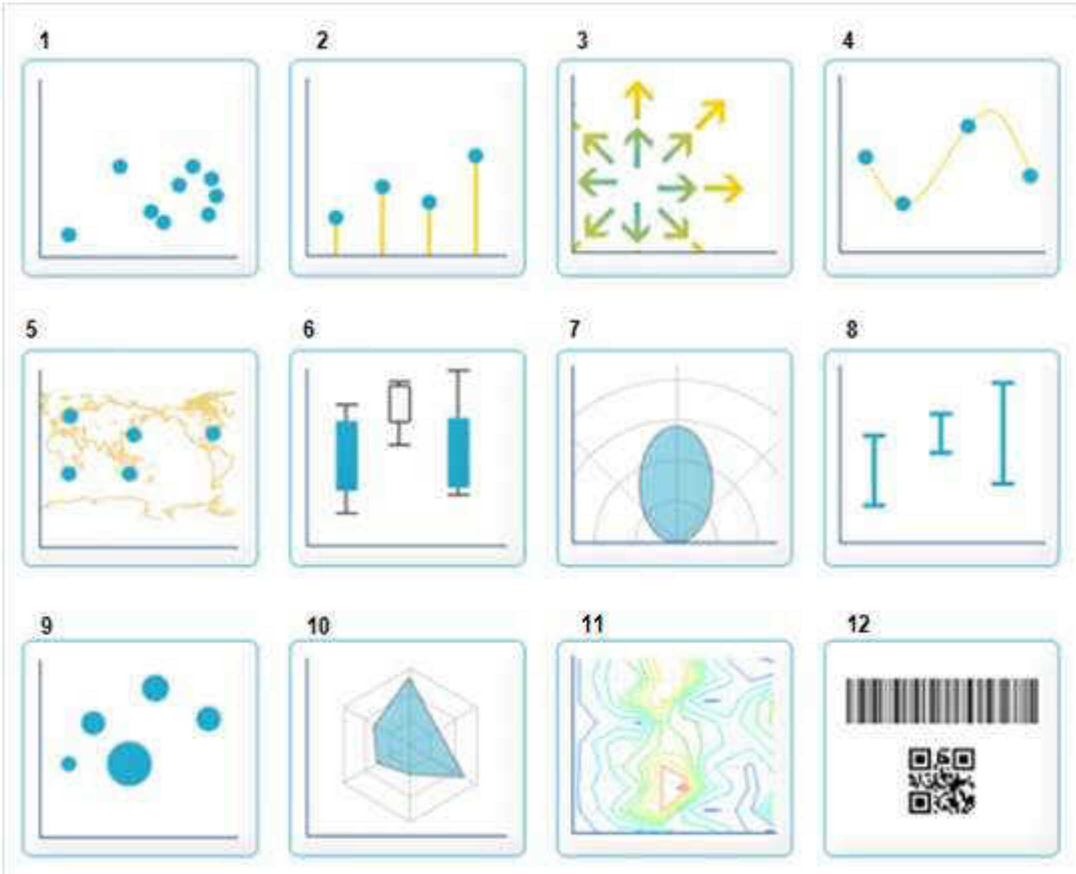
____

____

____

14. Do you have any previous experience using any of the displays shown below? YES   NO

15. Use the picture in question 14 to rate how experienced are you with using each display (#1-12).



#1	___ None	___ Novice	___ Average	___ Very Knowledgeable	___ Expert
#2	___ None	___ Novice	___ Average	___ Very Knowledgeable	___ Expert
#3	___ None	___ Novice	___ Average	___ Very Knowledgeable	___ Expert
#4	___ None	___ Novice	___ Average	___ Very Knowledgeable	___ Expert
#5	___ None	___ Novice	___ Average	___ Very Knowledgeable	___ Expert
#6	___ None	___ Novice	___ Average	___ Very Knowledgeable	___ Expert
#7	___ None	___ Novice	___ Average	___ Very Knowledgeable	___ Expert
#8	___ None	___ Novice	___ Average	___ Very Knowledgeable	___ Expert
#9	___ None	___ Novice	___ Average	___ Very Knowledgeable	___ Expert
#10	___ None	___ Novice	___ Average	___ Very Knowledgeable	___ Expert
#11	___ None	___ Novice	___ Average	___ Very Knowledgeable	___ Expert
#12	___ None	___ Novice	___ Average	___ Very Knowledgeable	___ Expert

16. On a scale from 1 to 5, where 1 is very tired and 5 is fully awake, how would you describe yourself current (circle one)

Very Tired		Somewhat Awake		Fully Awake
1	2	3	4	5

17. On a scale from 1 to 5, where 1 is completely calm and 5 is very stressed, how would you describe yourself current (circle one)

Calm		Somewhat Stressed		Very Stressed
1	2	3	4	5

APPENDIX C: KNOWLEDGE CHECK

Knowledge Check

1. The student communicating orders to navigate own assets (e.g. ship, aircraft) is
    a. Provide Status Updates
    b. Provide Navigational Orders
    c. Communication
    d. Manage Threats

2. The student communicating situational updates to other team members is
    a. Provide Status Updates
    b. Provide Navigational Orders
    c. Resource Usage
    d. Communication

3. How well the student talks to others is
    a. Communication
    b. Provide Status Updates
    c. Team Participation
    d. Manage Threats

4. The student's participation as a team member is
    a. Provide Status Updates
    b. Team Participation
    c. Communication
    d. Resource Usage

5. The student carrying out his role in the mission by following protocol is
    a. Provide Status Updates
    b. Provide Navigational Orders
    c. Follow Mission Plan
    d. Manage Threats

6. The student managing own assets is
    a. Provide Navigational Orders
    b. Manage Threats
    c. Resource Usage

d. Team Participation

7. The student following protocol to deal with threats (i.e. use correct assets and procedures) is
    a. Provide Status Updates
    b. Provide Navigational Orders
    c. Follow Mission Plan
    d. Manage Threats

8. The student's ability to follow tactical protocol is
    a. Team Participation
    b. Follow Mission Plan
    c. Tactical Procedure
    d. Manage Threats

9. The student's overall evaluation is
    a. Manage Threats
    b. Follow Mission Plan
    c. Provide Status Updates
    d. Overall Performance

10. Which of the following are Course of Action Performances? (circle all that apply)

Communication	TAO
Provide Navigational Orders	Team Participation
Provide Status Updates	Follow Mission Plan
Resource Usage	Tactical Procedure
Manage Threats	

11. Which of the following are Learning Objectives Performances? (circle all that apply)

Communication	Provide Status Updates
Provide Navigational Orders	Resource Usage

Manage Threats

TAO

Team Participation

Follow Mission Plan

Tactical Procedure

Use the following list of words to answer questions 12 to 16.

Communication

Provide Navigational Orders

Provide Status Updates

Resource Usage

Manage Threats

TAO

Team Participation

Follow Mission Plan

Tactical Procedure

12. Communication = ______________________________+ ______________________________

13. Resource Usage = _____________________________+________________________________

14. Team Participation =___________________________+______________________________

15. Tactical Procedure = ___________________________+___________________________

16. Overall TAO Performance=______________________+________________________+

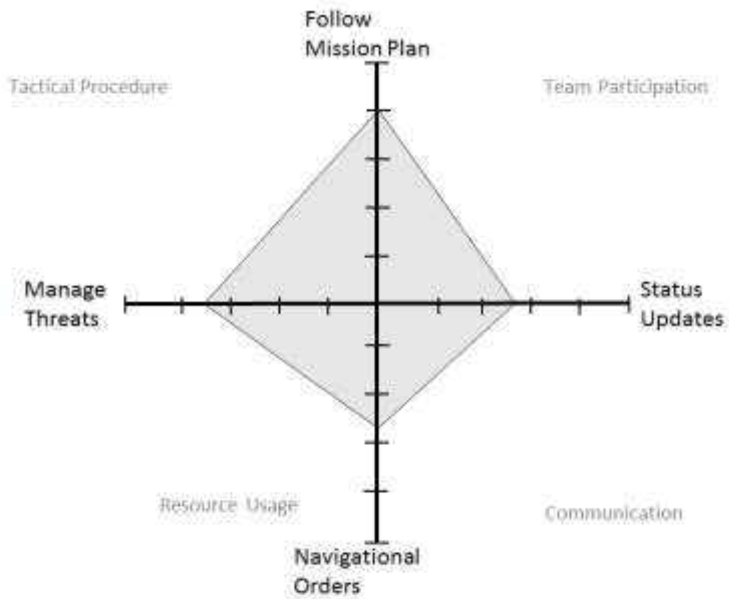
______________________+___________________________

17. Poor learning objectives performance is when
    a. At least one COA Performance is between 0-79%
    b. Two COA Performance is between 0-79%
    c. Three COA Performance is between 0-79%
    d. All COA Performances are equal or greater than 80%

18. The TAO passes when
    a. All Learning Objectives are at good performances
    b. At least one Learning Objective is at poor performance
    c. Two Learning Objective is at poor performance
    d. Three Learning Objective is at poor performance

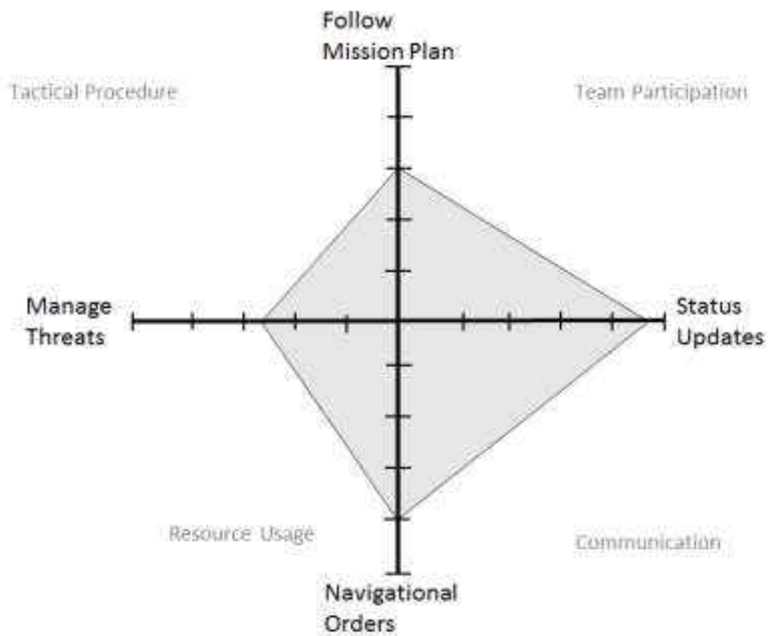
19. You would provide remediation when
    a. At least one COA performance is poor
    b. All COA performances are good

20. Circle the correct answer



What is the performance of Navigational Orders?
1) 80-100%
2) 0-79%

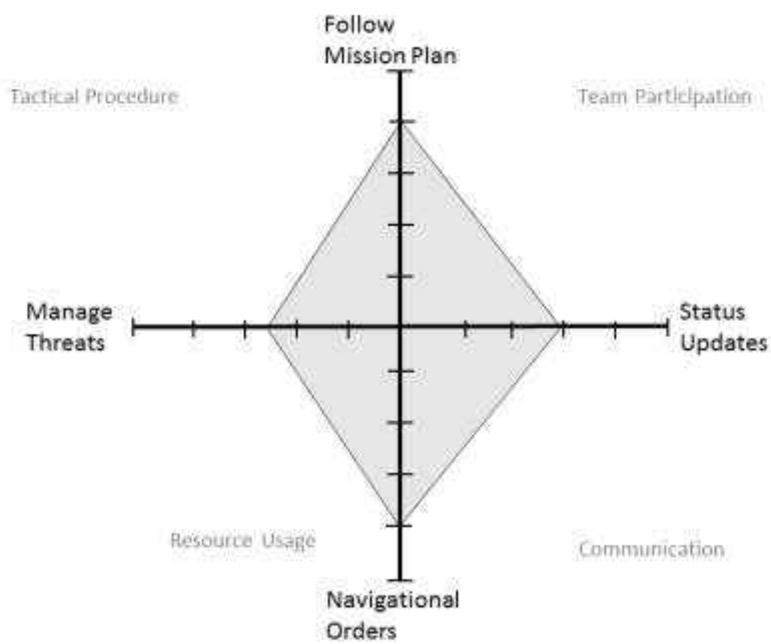
21. Circle the correct answer



What is the performance of Team Participation?

1) Good Performance
2) Poor Performance

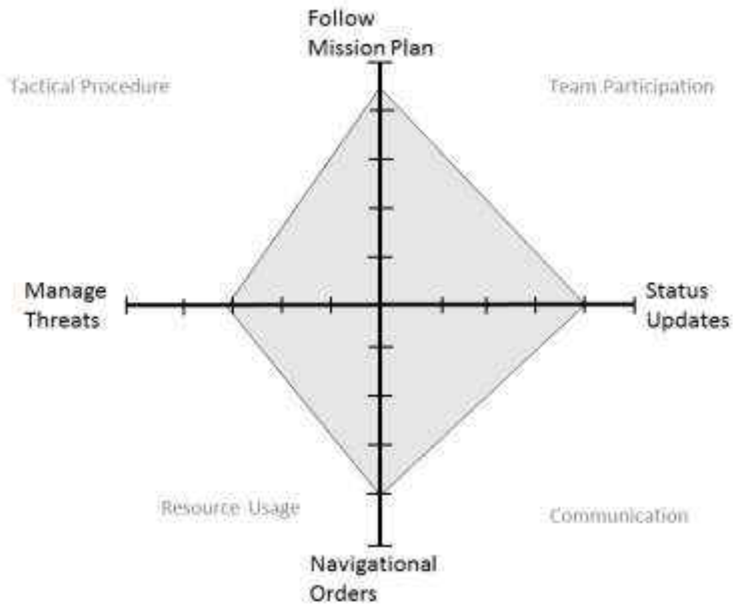
22. Circle the correct answer



What is the performance of the TAO?

1) Pass
2) Fail

23. Circle the correct answer



Would you provide remediation?

1) Yes
2) No

REFERENCES

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer based instruction. *Journal of Educational Computing Research, 13*(2), 11-127.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213-238.

Bennett, K. B., & Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors, 34*, 513–533.

Bennett, K.B. & Flach, J. M. (2011). *Display and interface design: Subtle science, exact art*. Boca Raton, FL: CRC Press.

Bennett, K. B., Nagy, A. L., & Flach, J. M. (2012). Visual displays. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics, 4,* (pp. 1179-1208). New York, NY: John Wiley & Sons.

Bennett, K. B., Payne, M., Calcaterra, J., & Nittoli, B. (2000). An empirical comparison of alternative methodologies for the evaluation of configural displays. *Human Factors, 42*(2), 287–298.

Bennett, K. B., Toms, M. L., & Woods, D. D. (1993). Emergent features and configural elements: Designing more effective configural displays. *Human Factors, 35*, 71–97.

Bennett, K. B., & Walters, B. (2001). Configural display design techniques considered at multiple levels of evaluation. *Human Factors, 43*(3), 415–434.

Burns, C. M., & Hajdukiewicz, J. R. (2004). *Ecological interface design*. Boca Raton, FL: CRC Press.

Buttigieg, M. A., & Sanderson, P. M. (1991). Emergent features in visual display design for two types of failure detection tasks. *Human Factors, 33*(6), 631-651.

Byers, C.J., Bittner, A.C.J. et al. (1988). Workload assessment of a remotely piloted vehicle (RPV) system. *Proceedings of the Human Factors Society's 32nd Annual Meeting*, Santa Monica, California, USA.

Cannon-Bowers, J. A., & Salas, E. (1998). Team performance and training in complex environments: Recent findings from applied research. *Current Directions in Psychological Science, 7*(3), 83−87.

Cannon-Bowers, J. A., Burns, J. J., Salas, E., & Pruitt, J. S. (1998). Advanced technology in scenario-based training. In J. A. Cannon-Bowers & E. Salas (Eds). *Making decisions under stress: Implications for individual and team training* (pp. 365-374). Washington, DC: American Psychological Association.

Carswell, C. M., & Wickens, C. D. (1987). Information integration and the object display. *Ergonomics, 30*, 511–527.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K.R. Boff, L. Kaufman and J.P. Thomas. *Handbook of Perception and Human Performance. Vol. 2*: *Cognitive Processes and Performance* (pp. 41-49). Oxford, England: John Wiley & Sons.

Hansen, J. P. (1995). An experimental investigation of configural, digital, and temporal information on process displays. *Human Factors, 37*, 539–552.

Hart, S.G. and Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload*. P.A.M. Hancock, N. Amsterdam, North-Holland: 139-183

Hill, S.G., Iavecchia, H.P., et al. (1992). Comparison of four subjective workload rating scales. *Human Factors 34*(4): 429-439.

Holt, J., Bennett, K., & Flach, J. (2011). Ambiguity and content mapping among display types. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55*(1), 390-393.

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*(1), 79-97.

MacGregor, D., & Slovic, P. (1986). Graphic representation of judgmental information. *Human–Computer Interaction, 2*, 179–200.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324. doi:10.3758/s13428-011-0168-7.

McCracken, J. H., & Aldrich, T. B. (1984). *Analysis of selected LHX mission functions: Implications for operator workload and system automation goals* (TNA Publication No. AS1479-24-84). Fort Rucker, AL: Anacapa Sciences, Inc.

Milham, L. M., Pharmer, J.A., & Fok, A. W. (2013a). Adaptive instructor operating systems: Design to decrease instructor workload and increase effectiveness. *Proceedings of the 2013 Interservice Industry Training Simulation and Education Conference* (*I/ITSEC).* Orlando, FL.

Milham, L. M., Pharmer, J.A. , & Fok, A. W. (2013b). Training effectiveness evaluation of real-time metacognitive feedback. *Proceedings of the 2013 Interservice Industry Training Simulation and Education Conference* (*I/ITSEC*). Orlando, FL.

Milham, L. M., Pharmer, J.A. , & Fok, A. W. (in press). Adaptive instructor operating systems: Design of support instructor assessment of team performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.

Nygren, T.E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors 33*(1): 17-33.

Nicholson, D., Fiore, S., Vogel-Walcutt, J. J., & Schatz, S. (2009). Advancing the science of training in simulation-based training. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *53,* 1932-1934.

Oser, R. L., Cannon-Bowers, J. A., Salas, E., & Dwyer, D. J. (1999). Enhancing human performance in technology-rich environments: Guidelines for scenario-based training. *Human/technology interaction in complex systems, Vol. 9* (pp. 175-202). Stamford, CT: JAI Press, Inc.

Oser, R. L., Gualtieri, J. W., Cannon-Bowers, J. A., & Salas, E. (1999). Training team problem solving skills: An event-based approach. *Computers in Human Behavior, 15*, 441-462.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition, 28*, 39-58.

Pomerantz, J. R., & Pristach, E. A. (1989). Emergent features, attention, and perceptual glue in visual form perception. *Journal of Experimental Psychology: Human Perception and Performance 15,* 635–649.

Pomerantz, J., Sager, L.C., & Stoever, R.J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance, 3*(3), 422-435.

Rasmussen, J. (1986). *Information processing and human–machine interaction: An approach to cognitive engineering*. New York, NY: Elsevier.

Ross, K. G., Phillips, J. K., Klein, G., & Cohn, J. (2005). *Creating expertise: A framework to guide technology-based training*. (Tech. Rep. for Contract # M67854-04-C-8035 for the Marine

Corps Systems Command, Program Manager for Training Systems). Fairborn, OH: Klein Associates, Inc.

Salas, E., Rhodenizer, L., Bowers, C. A. (2000). The design and delivery of crew resource management training: Exploiting available resources. *Human Factors, 42*(3), 490–511.

Sanderson, P. M., Flach, J. M., Buttigieg, M. A., & Casey, E. J. (1989). Object displays do not always support better integrated task-performance. *Human Factors, 31*(2), 183-198.

Sanderson, P., Pipingas, A., Danieli, F. & Silberstein, R. (2003). Process monitoring using a configural display: Human performance, self-report, and neuroimaging. *Theoretical Issues in Ergonomic Science, 4*, 151–174.

Schatz, S., Oakes, C., Folsom-Kovarik, J. T., & Dolletski-Lazar, R. (2012). ITS + SBT: A review of operational situated tutors. *Military Psychology, 24*(2), 166-193.

Shelton, J. T., Metzger, R. L., & Elliott, E. M. (2007). A group-administered lag task as a measure of working memory. Behavior Research Methods, 39 (3), 482-493.

Steinhauser, N. B., & Dehne, S. (2013). The search for the elusive spidey sense. Proceedings of the 2013 Interservice/Industry Training, Simulation and Education Conference, Orlando, FL.

Steinhauser, N. B., & Dehne, S. (2013). Intuition: Listening to your gut takes skills. Proceedings of the 2014 Interservice/Industry Training, Simulation and Education Conference, Orlando, FL.

Szalma, J.L. (2009). Individual differences in human-technology interaction: Incorporating variation in human characteristics into human factors research and design. Theoretical Issues in Ergonomics Science, 10, 381-397.

Szalma, J. L. (2002). Workload and stress of configural displays in vigilance tasks. *Proceedings of the Human Factors and Ergonomics Society, 46*, 1536-1540.

Szalma, J. L. (2011). Workload and stress in vigilance: The impact of display format and task type. *American Journal of Psychology, 124*, 441-454.

Vandenberg, S. G. & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599-601.

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. Mahwah, NJ: Erlbaum & Associates.

Wickens,C. D., & Andre 1990

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundations and its relevance to display design. *Human Factors, 37*, 473-494.

Woods, D. D., Wise, J. A., & Hanes, L. F. (1981). An evaluation of nuclear power plant safety parameter display systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 110-114.