Use Authorization

In presenting this dissertation in partial fulfillment of the requirements for an advanced degree at Idaho State University, I agree that the Library shall make it freely available for inspection. I further state that the permission to download and/or print my dissertation for scholarly purposes may be granted by the Dean of the Graduate School, Dean of my academic division, or by the University Librarian. It is understood that any copying or publication for financial gain shall not be allowed without my written permission.

Signature ________________________________

Date ______July 19, 2017_______________

COMPARING THE 3D AUGMENTED REALITY SANDBOX AND A 2D PAPER MAP'S EFFECTS ON STUDENT LEARNING AND COGNITIVE LOAD AMONG UNIVERSITY UNDERGRADUATES: APPLYING MULTIMODAL AND EMBODIED INTERACTION THEORIES TO TEACHING TOPOGRAPHIC MAP SKILLS

by

R. Thomas Richardson

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in the

Department of Organizational Leadership and Performance

Idaho State University

Summer 2017

Copyright

Committee Approval

To the Graduate Faculty:

The members of the committee appointed to examine the dissertation of RICHARD THOMAS RICHARDSON find it satisfactory and recommend that it be accepted.

Dr. Dotty Sammons, Major Advisor

Dr. Peter Denner, Committee Member

Dr. Quincy Conley, External Committee Member

Dr. Beverly Ray, Committee Member

Dr. Tina Miyake, Graduate Faculty Representative

Human Subjects Committee Approval

Idaho State UNIVERSITY

Office for Research Integrity
921 South 8th Avenue, Stop 8046 • Pocatello, Idaho 83209-8046

July 27, 2016

Rick Richardson
Organizational Learning - Perf
MS 8059

RE: regarding study number IRB-FY2017-7 : User Performance and Cognitive Load Measures to Affect Spatial Thinking Using Topographic Maps: A Comparative Analysis

Mr. Richardson:

I have reviewed your request for expedited approval of the new study listed above. This is to confirm that I have approved your application.

Notify the HSC of any adverse events. Serious, unexpected adverse events must be reported in writing within 10 business days.

You may conduct your study as described in your application effective immediately. The study is subject to renewal on or before Jul 27, 2017, unless closed before that date.

Please note that any changes to the study as approved must be promptly reported and approved. Some changes may be approved by expedited review; others require full board review. Contact Tom Bailey (208-282-2179; email humsubj@isu.edu) if you have any questions or require further information.

Sincerely,

Ralph Baergen, PhD, MPH, CIP
Human Subjects Chair

Phone: (208) 282-2179 • Fax: (208) 282-4723 • www.isu.edu/research
ISU is an Equal Opportunity Employer

Acknowledgements

I would like to thank Idaho State University's Instructional Design department and my committee members, Dr. Dotty Sammons, Dr. Peter Denner, Dr. Quincy Conley, Dr. Beverly Ray, and Dr. Tina Miyake for their recommendations on and support for this final manuscript. In particular, I would like to acknowledge my committee chair, Dr. Sammons, for her insightful comments and advice on earlier draft versions and Dr. Denner for his recommendations on data analysis and reporting. I am also appreciative of Dr. Conley's willingness to collaborate on the eye-tracking data collected in this study for future publication. Thanks are also extended to Sheldon Harris for providing computer technical support. Dr. Michael Peters was kind enough to provide a copy of his Mental Rotations Test, along with administration and scoring protocols. Dr. Steven Weisberg and his colleagues also deserve thanks for giving permission to modify the Topographic Map Assessment for the purposes of this study.

The scale and scope of this study would not have been possible without my two research assistants, Phillip Nyaku and Ashley Ferguson. Their assistance with conducting experimental sessions, data collation, and data entry was invaluable. I would like to acknowledge that support of The National Science Foundation's Doctoral Dissertation Research Improvement Grant for Spatial Sciences (Award# 1633750) provided funds for research assistants. It also covered hardware, conference attendance fees to disseminate this study's findings, and financial incentives for participants.

Idaho State University's Geoscience department graciously provided access to and technical support of the Augmented Reality Sandbox for the duration of the study.

Thanks go out to Lori Tapanila, Dr. Dave Pearson, Dr. Carrie Bottenburg, Dr. Leif Tapanila, and Dr. Ben Crosby for allowing me to recruit students from their classes. Similar thanks go out to Dr. Yolanda Youngs from the Global Studies department and Dr. Reinalyn Echon and Dr. Tina Miyake from the Psychology department.

Lastly, many thanks to my parents, George and Keitha Richardson, who have always been supportive of my academic aspirations and to my wife, Dr. Donna Delparte, for her sage advice, unwavering encouragement, and technical assistance. Without students in her Geovisualization class building the Augmented Reality Sandbox as a course project, led by Matthew Belt, this study would not have been possible.

Table of Contents

List of Figures

List of Tables

COMPARING THE 3D AUGMENTED REALITY SANDBOX AND A 2D PAPER MAP'S EFFECTS ON STUDENT LEARNING AND COGNITIVE LOAD AMONG UNIVERSITY UNDERGRADUATES: APPLYING MULTIMODAL AND EMBODIED INTERACTION THEORIES TO TEACHING TOPOGRAPHIC MAP SKILLS

Dissertation Abstract—Idaho State University (2017)

This study tested the proposition that the Augmented Reality Sandbox's (ARS) user-interaction from tactile sensory feedback and a realistic 3D perspective improved topographic map comprehension among novice users with reduced cognitive load compared to the same instruction and practice from a 2D topographic map. Undergraduate students were randomly assigned to one of two treatment groups after completing a spatial test that assigned participants to either a low or high spatial ability group. Treatment consisted of multimedia instruction followed by learning practice with either the AR condition or a duplicate 2D paper topographic map (non-AR condition). Learning performance scores and separate cognitive load ratings from the Mental Effort Measurement Scale were collected for each learning practice question and for each question on a subsequent posttest. Treatment group (AR or non-AR), spatial ability, and cognitive load from the posttest did not identify any significant differences between treatments. Applying the same analysis for the learning practice questions that preceded the posttest revealed a significant interaction between instructional condition and performance score and a significant interaction between instructional condition and spatial ability on cognitive load. The AR participants scored significantly higher than non-AR participants during learning practice and did so with less cognitive load. The spatial ability as enhancer hypothesis accounted for the high spatial ability group's higher scores with insignificant differences in cognitive load, regardless of treatment condition. Cognitive load for the low spatial ability group was dependent on treatment condition during learning practice: lower cognitive load and higher scores were associated with the AR condition and higher cognitive load and lower scores with the non-AR condition. This was explained by the spatial ability as enhancer hypothesis. Learning performance with minimal cognitive load was reconciled to embodied interaction and multimodal theories, which suggested that hands-on interaction and tactile feedback from the ARS promoted germane cognitive load, thereby offsetting the extra sensory information (visual plus tactile) that the ARS provided. These findings promote the use of AR instruction for teaching topographic map reading skills to novices. Recommendations for instructional design of embodied interaction and multimodal learning environments and directions for further research are offered.

Keywords: Augmented Reality Sandbox, spatial ability, cognitive load of spatial data, embodied interaction, multimodal learning, and augmented reality.

Chapter I

Introduction

This experimental study compared traditional topographic map instruction techniques to new instructional methods featuring an innovative augmented reality map. The goal was to determine if the latter improved learner performance in novice level topographic map use skills and did so with reduced mental effort. Results were interpreted in terms of multiple theoretical perspectives and informed recommendations on topographic map instructional strategies for digital learning environments.

Theoretical Framework

This study drew upon current research in a range of seemingly disparate fields: cartographic visualization, spatial ability, cognitive load theory, and tangible user interface designs that incorporate tactile feedback as gestures that can support abstract reasoning of multidimensional spaces. Multimodal learning and embodied interaction theories can unify relevant elements within the aforementioned fields as they apply to learner performance and cognitive load measures. Multimodal learning and embodied interaction theories can also interpret this study's research findings by explaining how user interactivity can account for learning gains from topographic map instruction with minimum mental effort. However, multimodal learning theory does not address issues of dimensionality. Cartographic visualization and spatial ability theories offer interpretations of findings related to comparative differences in user performance and mental effort between the three dimensional representation of the Augmented Reality Sandbox (Reed et al. 2014).

Cartographic visualization. A topographic map visually describes the shape and relative elevations of landforms using a series of abstract contour lines. Contour lines represent the shape profile of three dimensional (3D) landscapes on a two dimensional (2D) surface, typically rendered on paper or on a computer screen. Abstract thinking is required to transfer spatial information from one dimension to another; a process that is "neither automatic nor easy" (The National Academies Press, 2006, p. 109). Several studies have concurred on the mental complexity of translating spatial visualizations from 2D to 3D and vice-versa (Ishikawa & Kastens, 2005; Ormand et al., 2014; Titus & Horsman, 2009). Consequently, the literature is rife with conflicting results of learning retention and transfer from 2D and 3D topographic maps (Huk, 2007; Smallman, John, Oonk, & Cowen, 2001; Oulasvirta, Estlander, & Nurminen, 2009; Savage, Wiebe, & Devine, 2004; Schobesberger & Patterson, 2007). This debate over map dimensionality is but one component affecting the usability of topographic maps.

Spatial ability and cognitive load. The consequence of the inherent visual complexity of topographic maps is that a higher degree of spatial ability skill is required, compared to other types of maps that are more widely used by the public (Newcombe, Weisberg, Atit, Jacovina, Ormand, & Shipley, 2015; Piburn, Reynolds, Leedy, & Mcauliffe, 2002). Cognitive load is thus a possible mediating factor on map reading competency as too much spatial information may overwhelm the user's ability to process, store, and retrieve information (Bunch & Lloyd, 2006; Harrower, 2007). This is a situation of cognitive overload that can impede learning (Clark & Mayer, 2011; Sweller, 1988). Consequently, instructional strategies that promote an optimum level of learning will maximize germane cognitive load. High germane cognitive load with minimal extraneous cognitive load (e.g., distractions that can interfere with learning) have the potential to increase topographic map reading ability (Bunch & Lloyd, 2006).

The increasing use of Geographic Information Systems (GIS) as a mapping tool has emphasized the importance of spatial thinking (Goodchild, 2011; Wakabayashi & Ishikawa, 2011) and has provided more sophisticated, detailed, and thus complex, maps. The GIS approach to managing spatial data is to create multiple map layers, each containing spatial attributes in addition to associated metadata that may be queried to identify spatial relationships (The National Academies Press, 2006). A GIS map thus measures, represents, and transforms spatial information. Representing and transforming space are of particular relevance to displaying spatial data as a topographic map. GIS has simplified the process of transferring spatial information from a 2D format to a 3D map that can be viewed from multiple perspectives as geovisualization map (Goodchild, 2011; Nollenburg, 2007; Yun, Yufen, & Yingjie, 2004). Examples include the GeoWall (Johnson, Leigh, Morin, & Van Keken, 2006) and Piburn's Geology instruction modules on spatial visualization and topographic map use (Piburn et al., 2002). Elements of these and other interfaces can facilitate dimensional transfer between 2D and 3D landscape representations.

Tangible user interfaces and tactile feedback. GIS has also inspired other innovations in map visualization and usability. Oliver Kreylos' Augmented Reality Sandbox (ARS) (Reed et al., 2014) is another example of a geovisualization map, which is unique because it is conjoined with a tangible user interface (TUI). The ARS combines 3D landscapes with a tactile interface to allow users to create and alter landforms within a virtual sandbox and to see the changes in real-time. As a TUI, it provides both visual and tactile feedback of a topographically represented surface. In spite of the advances in GIS technology that allow more sophisticated methods to display spatial information with maps, there is a gap in the literature on how tactile feedback may enhance the ARS's 3D perspective in terms of learner performance with optimum cognitive load.

Although dimensionality of maps has been the subject of considerable research, as already noted, there are other larger and unanswered questions about how tactile feedback may be integrated with dimensionality, only some of which were analyzed within the context of this study. For example, does information from the visual and tactile senses from a tangible map, such as the ARS, enhance learning compared to the primarily visual feedback associated with conventional 2D topographic maps? Building on this premise raises another question: are spatial ability skills improved by tactile feedback and what effect might the latter have on cognitive load? Another consideration is whether the ARS' novel multi-sensory interface (visual and tactile feedback) increases user engagement, which may lead to increased performance? On the other hand, is it also plausible that the ARS might provide too much sensory information to the point of cognitive overload? If this were true, would not the (arguably) simpler, more familiar, and commonly used 2D topographic map be a preferable instructional tool, as argued by Savage, Wiebe, & Devine (2004)? Or is the choice of 2D versus 3D dependent on an individual's spatial ability (Clifton et al., 2016; Huk, 2007; Quarles, Lampotang, Fischler, Fishwick, & Lok, 2008b)?

In terms of future applications stemming from this line of inquiry, can recent research on spatial cognition and cognitive load inform new instructional methods that promote hands-on user interactivity with topographic maps? To do so, the map interface and instructional method would need evidence of significant learning gains with, presumably, minimum cognitive load. To date, only one published pilot study (Woods, Reed, Hsi, Woods, & Woods, 2016) has investigated use of the ARS in a university setting, describing how the ARS was integrated into an introductory geology course and how it was perceived by students in terms of learning engagement and preference. Learning gains were not reported (Woods et al., 2016). Additional research studies on learning effectiveness and optimum cognitive load are recommended to ascertain if the ARS is a superior instructional aid for teaching topographic map reading skills and whether spatial ability and/or cognitive load influence the learning process. Multimodal learning theory can address some of these questions in terms of user interactivity effects on learning.

Multimodal learning and embodied interaction. This study was particularly interested in how the addition of tactile feedback in an augmented reality environment could affect learning. User interactivity between visual and tactile feedback offer a possible explanation of differential learning gains and cognitive load measures between 3D augmented reality and conventional 2D paper-based topographic maps. Multimodal learning theory supports multiple forms of instruction (Shams & Seitz, 2008) that can include other senses such as touch. Empirical research on learning in augmented reality (AR) environments that provide multiple sensory cues from the surrounding environment, including touch and sound, have contributed to increased performance

(Skulmowski, Pradel, Kühnert, Brunnett, & Rey, 2016), enhanced performance in less time (Jeong & Gluck, 2003) and with higher learner engagement (Jeong & Gluck, 2003; Woods et al., 2016). As a new paradigm for education, multimodal learning theory is thus amenable to contributions from research on physical and virtual manipulatives that provide both visual and tactile feedback (Wiebe, Minogue, Jones, Cowley & Krebs, 2009) such as the Augmented Reality Sandbox (Reed et al., 2014). Increased engagement associated with working in a multimodal learning environment increases germane load, which resulted in higher learner performance (Van Merriënboer, Schuurman, de Croock, & Paas, 2002). Fostering germane cognitive load also contributed to schema development (Wiebe, Roberts, & Behrend, 2010). Recent research on multimodal learning environments using tangible user interfaces indicated that they can also increase spatial ability skills (Clifton, Chang, Yeboah, Doucette, Chandrasekharan, Nitsche, Welsh, & Mazalek, 2016).

Embodied interaction theory (Streeck, Goodwin & LeBaron, 2011) can address many of the questions already posed. For example, it can offer explanations from evidence that user interactivity can increase performance (Keehner, Montello, Hegarty, & Cohen, 2004) and promote germane cognitive load, which leads to schema development. Robust schemas increase learning retention (Bunch & Lloyd, 2006; Sweller, 1988a; Sweller, Van Merriënboer, & Paas, 1998) that can manifest as enhanced learner engagement, improved learner performance, and reduction in cognitive load due to the availability of multiple sensory cues and opportunities for interactive feedback without violating the dual-channel assumption (Jeong & Gluck, 2003; Skulmowski et al., 2016; Xie, Antle, & Motamedi, 2008).

Statement of the Problem and Study Purpose

From the preceding section, it is evident that theory on multimodal learning and tangible interaction across multiple dimensional representations and the empirical debate over dimensionality of topographic maps are two areas that affect topographic map instruction, but have not been thoroughly examined. There is a shortage of evidence-based research on the learning efficacy of the ARS compared to traditional 2D paper topographic maps in terms of both user performance and cognitive load—hereafter referred to as mental effort. These theoretical and empirical shortcomings provided a rationale for investigating relationships between different types of topographic map interfaces, learning performance, and mental effort.

The purpose of this study was to test the proposition that interactive, tactile sensory feedback and a three dimensional view provided by the ARS improved topographic map comprehension and optimum mental effort relative to the same instruction and practice offered from a conventional 2D paper topographic map. Proof of this claim would suggest that the ARS was a more effective teaching tool for novice topographic map users, particularly those with low spatial ability, when compared to the traditional method of instruction using a 2D paper topographic map. Findings could be interpreted through the lens of embodied interaction and multimodal learning theories.

Research Questions

This study posed six quantitative research questions dealing with the comparisons between the dependent variables of learner performance and mental effort across the independent variables of AR and non-AR treatment conditions and low and high spatial ability groups. Research questions one to three pertained to learner performance while questions four to six applied to cognitive load.

1. Is there a main effect of instructional condition (AR and non-AR treatment) on posttest learner performance, as indicated by a test of topographic map skills?
2. Is there a main effect of spatial ability on posttest learner performance, as indicated by a test of topographic map skills?
3. Is there an interaction effect of spatial ability and instructional condition on posttest learner performance, as indicated by a test of topographic map skills?
4. Is there a main effect of instructional condition (AR treatment and non-AR treatment) on cognitive load, as indicated by the Mental Effort Measurement (Paas & Van Merriënboer, 1993)?
5. Is there a main effect of spatial ability on cognitive load, as indicated by the Mental Effort Measurement Scale (Paas & Van Merriënboer, 1993)?
6. Is there an interaction effect of spatial ability and instructional condition on cognitive load, as indicated by the Mental Effort Scale (Paas & Van Merriënboer, 1993)?

Research Design

This study's primary objective was to compare two indicators of learner performance (user posttest scores and associated mental effort measures) to determine which map type yielded higher posttest scores and lower mean mental effort levels. The ARS (Reed et al., 2014) was defined as the AR treatment condition while the paper topographic was referred to as the non-AR treatment condition. This terminology established the differences between the two types of map interfaces based on dimension

(2D for the non-AR condition and 3D for the AR condition) and the potential for user interactivity with the interface. The AR condition allowed for high user interactivity because of visual and direct tactile sensory inputs that allow for user modification of the map in real-time. In contrast, the non-AR condition offered only low user interactivity—primarily visual plus indirect manipulation of a 2D paper topographic map that did not change the map surface in response to user actions. In the non-AR condition, the participant drew topographic features on a separate sheet of paper and marked locations or traced on the map with a dry erase marker. The learner-reported Mental Effort Measurement Scale (Paas & Van Merriënboer, 1993) was used to quantify mental effort during learning performance testing. A secondary objective was to determine if low or high spatial ability groupings interacted with performance scores and/or mental effort.

This study employed a posttest-only randomized experimental design with matched pairs. The research process is visually illustrated via a flowchart diagram in Figure 1. Two spatial ability tests were administered to participants prior to treatment to determine if spatial ability could affect performance and mental effort. Two conditions (AR treatment and non-AR treatment) were subject to a two-part experimental treatment: (1) common multimedia instruction, followed by (2) a learning practice activity. In the AR condition, the participant was afforded a 3D view of the map with high user interactivity due to tactile feedback. In the non-AR condition, the participant answered the same questions and performed the same tasks as the AR condition using a 2D view with low user interactivity. This interactivity construct was indirectly measured in terms of results on the two dependent variables: learner performance and

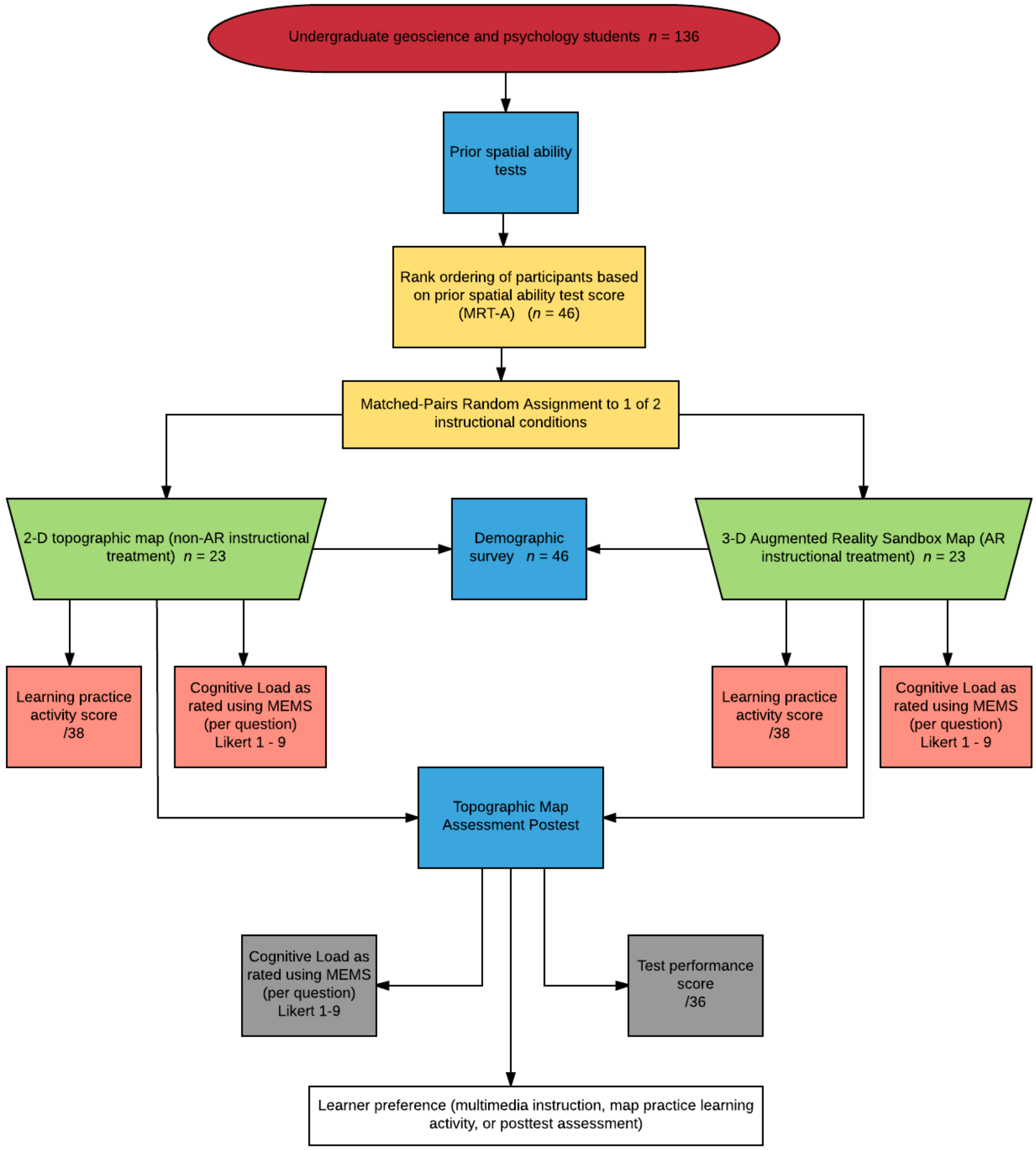


Figure 1. Schematic of the Research Design Process. Flowchart illustrates progression of experiment from participant selection and random assignment to two instructional treatment conditions with data collected during separate learning practice activities, followed by a common posttest assessment.

mental effort. Following instruction, both groups completed a common posttest which measured learning performance scores and mean mental effort.

Undergraduate students from a U.S. Intermountain West university (*n* = 136) were recruited to participate in the study. The sample was tested for spatial ability with a widely-used, standardized spatial ability test: the Mental Rotations test-A (Lovett & Forbus, 2013; Peters et al., 1995). Participants were ranked-ordered based on their MRT score as a matched pair, followed by random assignment to one of the two instructional treatment conditions. Due to a loss of subjects from initial spatial testing to the experimental session, only 83 participants attended and completed the experimental treatment, of which 23 sets of matched pairs (*n* = 46) were available for inferential analysis.

The one-hour experimental treatment session consisted of an online demographic questionnaire. This was followed by 20 minutes of multimedia instruction on how to read topographic maps, a common set of learning practice activities, and a common posttest assessment. Differences in the degree of interactivity that each map offered defined the two constructs of interest: dimensionality (2D for the non-AR condition versus 3D for the AR condition) and interactivity through embodied interaction (high for the AR condition versus low for the non-AR condition). The AR condition would view the topographic map in 3D and could touch and manipulate the sand, which would adjust the topographic map view in real-time. The non-AR condition's topographic map was rendered in 2D and interactivity was limited to line drawings to show topographic features and tracing or marking on the map with a dry erase marker.

The research or a trained assistant scored each learning practice activity and upon completing the task or answering each question, the participant was prompted to give a self-rating of their mental effort expended using the Mental Effort Measurement

Scale (Paas & Van Merriënboer, 1993) The value (rated on a nine point Likert scale—see Table 4 in Chapter 3) was an indication of the mental difficulty and time required to perform the task or answer the question. There were no time limits placed on any of the activities or assessments. The learning practice session consisted of 20 activities averaging 15 minutes to complete.

Participants then completed a standardized test of topographic map reading ability with a modified version of the Topographic Map Assessment test (Newcombe, Weisberg, Atit, Jacovina, Ormand, & Shipley, 2015). This posttest contained 25 questions. After completing each question, there was a space provided to rate the mental effort required to answer the question using the same rating of mental effort as per the learning practice activity. Upon completion of the posttest, participants were then asked which activity: (1) the multimedia instruction, (2) the learning practice using their respective map, or (3) the posttest was the most effective in supporting their learning and increasing their understanding of interpreting topographic maps.

Two 2 x 2 factorial ANOVA’s of between-subjects effects were used to answer research questions one to six. For research questions one to three, a two-way ANOVA on learner performance from a common posttest on topographic map skills included the main effects of treatment condition and spatial ability on posttest learner performance and the interaction effect of instructional condition and spatial ability on posttest learner performance. For research questions four to six, a two-way ANOVA on posttest mental effort values from the posttest on topographic map skills included the main effects of treatment condition and spatial ability on posttest mental effort and the interaction effect of instructional condition and spatial ability on posttest mental effort. A subsequent analysis of learning practice activity scores and their associated mental effort ratings that preceded the posttest assessment revealed significant differences. Two-way ANOVA tests of significant main effects and interactions of instructional condition and spatial ability on learning practice activity scores and on the mean MEMS values for each learning practice activity question yielded findings that could be explained via multimodal learning and embodied interaction theory.

Definition of Key Terms

Augmented reality. Augmented reality (AR) and Virtual reality (VR) are points on a spectrum ranging from reality to something completely imaginary. VR implies a complete immersion in a virtual environment, isolated from outside reality (Azuma, 1997). On the other hand, AR blends aspects of the real environment with virtual objects (Bimber & Raskar, 2005). The ARS's ability to project a moldable sand map and to display changes made to the map surface in real time is a clear example of an AR device.

Augmented reality sandbox (ARS). The Augmented Reality Sandbox is an example of a tangible user interface that allows users to interact with a 3D topographic map through tactile feedback (Reed et al., 2014). A sensor measures the relative height of modelling sand on the map table and a projector displays topographic lines and color to the elevation of the mapped landscape. As users touch and move the sand, the ARS updates the changes to the 3D map surface in real time. The sand is imbued with additional capabilities because of the visual interface that is projected upon it.

Cognitive load versus mental effort. Tory, Kirkpatrick, Atkins, & Moller (2006) provided a succinct definition of cognitive load as "the amount of work needed to acquire and use information" (p. 519). This study measured cognitive load with the Mental Effort Measurement Scale (Paas & Van Merriënboer, 1993), a subjective measure of cognitive load measure (Brünken & Plass, 2003). This study will use the term *mental effort*, when referring to MEMS values. This distinction is appropriate as cognitive load contains both mental load and mental effort constructs (Paas, Merrienboer, & Adam, 1994).

Learning practice activity. After multimedia instruction in basic topographic map reading skills, participants in both treatment groups completed a series of practice activities using their respective map interface. Proficiency scores and mental effort ratings were recorded for each of the 20 questions. The posttest assessment followed learning practice.

Leaner performance. Learning proficiency was based on a participant's individual score totals on the learning practice activity and a separate score on the modified topographic map posttest assessment.

Mental effort measurement scale (MEMS). The Mental Effort Measurement Scale (Paas & Van Merriënboer, 1993) is a widely used metric for measuring mental effort (Gog, Kirschner, Kester, & Paas, 2012). It is qualitative in nature since subjects self-report their relative degree of mental effort expended to complete a cognitive task using a Likert scale of increasing mental effort from one (very, very low mental effort) to nine (extremely high mental effort). Refer to Table 4 in Chapter 3 for the full scale. Two measures of mental effort were collected: (1) mental effort per question during the learning practice activity and (2) mental effort per question during the posttest assessment. For analysis, individual MEMS ratings were aggregated as a mean value per participant for the learning practice activity and for the posttest.

Multimodal learning. According to multimedia learning theory (Clark & Mayer, 2011; Mayer, 2005) verbal and visual channels constitute the two primary sensory input pathways in cognitive load theory. Jeong and Gluck (2003) offer an extension to the dual channel model underpinning multimedia learning theory: "multimodality means more than one communication channel, which is used simultaneously to convey or acquire information" (p. 229). Multimodal learning acts as a bridge between the virtual and physical environments by using other sensory information such as touch. The term *multimodal learning environment* (MLE) (Moreno & Mayer, 2007) will be used to delineate the instructional milieu in which multimodal learning and embodied interaction occurs (Johnson-Glenberg, Birchfield, Tolentino, & Koziupa, 2014).

Spatial ability. Although there is no one single definition for spatial ability, "it is generally accepted to be related to skills involving the retrieval, retention, and transformation of visual information in a spatial context" (Velez, Silver, & Tremaine, 2005, p.12). As a label, spatial ability subsumes multiple attributes such as spatial perception, mental rotation, and spatial visualization (Linn & Petersen, 1985; Voyer & Bryden, 1990), whereas *spatial thinking* is a more specialized term because of its focus on problem-solving using spatial information (The National Academies Press, 2006). Since participants were assessed on their spatial abilities in advance of the treatment session, the measure was based on a snapshot of a participant's spatial ability at the time of testing; it can also be thought of as *prior spatial ability*. Testing spatial ability in advance was necessary to group and then randomly assign participants from the entire sample prior to experimental treatment. Rather than a specific value, participants were coarsely divided to either a low or high spatial ability category based on their spatial test score.

Tangible user interface (TUI). Tangible user interfaces, employ a malleable surface or physical objects that users manipulate through touch (Xie, Antle, & Motamendi, 2008). To this working definition, virtual objects should also be included and a reference to some form of either visual or visual-haptic feedback associated with manipulating the surface or object (Ullmer & Ishii, 2001). A 3D tactile feedback map, such as the ARS (Reed et al., 2014) is an example of an interactive geovisualization display providing visual and tactile feedback.

Topographical map assessment (TMA). The Topographic Map Assessment (Newcombe et al., 2015) was administered as a common posttest to the AR and non-AR treatment conditions following multimedia instruction and learning practice with the map interface assigned to each treatment group. In this study, it measured learner performance and mental effort. The objective was to determine which instruction treatment condition (AR or non-AR) was a more effective learning tool for interpreting topographic maps. In order to cover all of the 17 predetermined learning outcomes for interpreting topographic maps at a novice level and to also incorporate user-reported MEMS values after completing each question (listed in Table 4 of Chapter 3), the researcher contacted the TMA developers (Newcombe et al., 2015) to receive permission to add MEMS rating to the existing 18 questions and to add seven matching questions on identifying topographic features. These two additions distinguished it from the original TMA. When referring to the version of the TMA used in this study, it is identified as the modified TMA (mTMA). Two measures were recorded for the posttest at the end of the experimental session: total score on the 25 mTMA test questions and mental effort for each of these questions.

Treatment/instructional condition. This term refers to the experimental treatment experienced by the two treatment groups: AR (using the ARS) and non-AR (using the 2D paper topographic map). As instruction, learning practice, and posttest assessment are elements shared by both the AR and non-AR groups, instructional condition, and treatment condition are synonymous.

User interactivity and embodied interaction. For the purposes of this study, these two constructs can be used interchangeably as they are defined in terms of the ability of the learner to directly manipulate a topographic map and to see how their actions changed the visual representation of the map surface in real time. Embodied interaction fosters user-interactivity as a two-way communication process between the user and the interface (Moreno & Mayer, 2007). Although both treatment conditions employed user interactivity, the ARS offered higher user interactivity relative to the 2D paper topographic map condition because the ARS dynamically changed in response to user actions, whereas the 2D paper topographic map was static; it did not change in response to user actions. The ARS allowed tactile and visual feedback as the map view altered in response to the user manipulating the sand, whereas the 2D paper topographic map condition offered primarily visual feedback as the user drew or traced topographic features on a separate sheet of paper or on the map. The degree of user interactivity associated with the 2D paper topographic map is the more traditional method associated with topographic map reading instruction (see Christopherson, 2010; Dorling & Fairbairn, 1997; Levin, 1986; Lounsbury & Aldrich, 1986; Petersen, Sack, & Gabler, 2011; Selby, 1985; and Strahler, 1987 for examples of traditional textbook instruction methods).

This dichotomy between high and low interactivity is appropriate as cognitive load theory distinguishes user interaction in terms of " 'active' (requires the use of more elaborate controls) or 'passive' (requires little additional activity from the user) interaction designs" (Skulmowski et al., 2016, p. 65). The effect of the user interactivity-embodied interaction construct could be inferred from higher performance scores of the ARS treatment group with lower mental effort, for example. User-interactivity will be interpreted through the lens of embodied interaction theory, which is defined as learning supplemented with bodily actions such as movement or touch (Price & Jewitt, 2013). These two theories are intertwined in that physical actions, such as tactile feedback, constitutes embodied interaction with a topographic map interface. Interaction promotes germane cognitive load which can lead to increased user-engagement and performance (Bunch & Lloyd, 2006; Harrower, 2007; M. J. Kim & Maher, 2008; Skulmowski et al., 2016; Wakabayashi & Ishikawa, 2011).

Assumptions

This study presumed a link between spatial ability and topographic map reading skills. Three important assumptions arose from this premise. First, it was assumed that spatial ability and topographic map reading skills could be enhanced through deliberate instruction in a particular map dimension (2D or 3D) and that subsequent tests of learner performance would reveal knowledge and skill retention related to the training effect.

Furthermore, mental effort would vary amongst participants: those with high spatial ability scores would likely score higher on the practice and posttest assessments and do so with lower mental effort than those participants who scored lower on the spatial ability tests. The varying degrees of user interactivity through embodied interaction and the dimensional representations of the two map conditions were also expected to influence test scores and mental effort, provided the participants were engaged in the learning activities, and provided MEMS ratings that were in accordance with the difficulty of the question or task. A pilot study found that some participants gave the same MEMS rating for every question. This outcome was not observed in the subsequent implementation of the full study. With respect to multimodal learning and embodied interaction, a combination of visual and tactile feedback using the ARS was presumed to offer a more effective method for learning how to read topographic maps because of the higher degree of tactile feedback and user-interactivity relative to the lower user interactivity effect associated with the non-AR condition.

Limitations

The guidelines for interval validity from Campbell and Stanley (1963) and considerations for external validity by Bracht and Glass (1968) were used to address possible validity threats in this study's implementation, analysis of data, and interpretation of findings. Of the internal validity threats Campbell and Stanley (1963) identified (history, maturation, testing, instrumentation, statistical regression, selection bias, mortality, and selection-maturation interactions), instrumentation and selection bias posed the most probable limitations to this study. As limitations, they could introduce confounding variables that caused the researcher to make incorrect claims about the effects of the independent variable(s) on the dependent variable(s).

Random assignment. As an experimental posttest-only design, random assignment of participants to treatments contributed significantly to overall internal validity: "the most adequate all-purpose assurance of lack of initial biases between groups is randomization. Within the limits of confidence stated by the tests of significance, randomization can suffice without the pretest" (Campbell & Stanley, 1963, p. 25). Since this design did not employ a pretest, the spatial ability tests were not related to the learning outcomes of the experimental treatment. Utilizing matched pairs of participants, based on high or low spatial ability as a blocking factor, mitigated the threat of mortality. Rank ordering and pairing of participants necessitated both participants in a pair to complete the study. If one failed to attend, the other participant's scores were discarded. Matched pairs assignment thus guaranteed an equal distribution of randomly distributed participants to the four cells of a two-way analysis of variance. However, this process also imposed a limitation. Using spatial ability as a blocking factor for matched pairs assignment reduced the usable sample size to 23 sets of matched pairs (n = 46), which impacted the probability of finding statistically significant differences and concurrent effect sizes from a small sample. Consequently, within the parameters of this study random assignment controlled for history, maturation, and selection-maturation interactions, while matched pairs controlled for mortality

Statistical regression. Collecting multiple MEMS scores from both learning practice activities and the posttest increased the likelihood of regression towards the mean. Learning practice following instruction was assumed to positively influence posttest results. Examination of differences between performance scores and mental effort ratings between the learning practice activity and the posttest could identify if a carry-over effect occurred from learning practice to the posttest assessment. One of the advantages of the matched pairs sampling design is that it eliminates carry-over and practice effects (Johnson & Christensen, 2008).

Instrumentation bias. The TMA and MEMS, as validated instruments, controlled for instrumentation effects. The MEMS instrument is regarded as highly reliable ($\alpha > .90$), assuming random assignment (Paas, & Van Merriënboer, 1993). A subsequent study reported similar reliability values (Stark, Mandl, Gruber & Renkl, 2002). The TMA (Newcombe et al., 2015) also exhibited high reliability ($\alpha = .76$). However, multimedia instruction and learning practice activities and modifications to the original TMA were unique to this study and thus received limited independent verification. The mTMA received a face validity review by two geoscience experts prior to implementation. To mitigate this limitation, instruction and assessment for both treatment conditions were aligned to a common set of 17 pre-defined standards (see Table in Chapter 3) that emerged from an instructional design analysis. This ensured that both treatments received the same instruction and were assessed in the same way.

Selection bias. Perhaps this most significant threat to the internal validity of this study, was a probable selection bias that could be only partially remediated by random assignment. A convenience sample of students, mostly at the freshman level and enrolled in geoscience and psychology courses ($N = 477$), created a pool of recruits that were not completely representative of the population of all university students at the study site. A recruitment effect could also bias the selection toward favoring more academically capable and motivated students and/or students who would be less inclined to participate if it were not for the extra credit and financial incentive. Both of these cases could limit generalizability (Taylor & Asmundson, 2007). To control for selection bias, participants were ranked-ordered by spatial test score into matched pairs before random assignment to either the AR or non-AR treatment condition.

Delimitations

Generalizability to other populations (population validity) was limited because a convenience sample of geoscience and psychology students participated in the experiment. Therefore, findings could not be extrapolated to all university students nor to the general population (Bracht & Glass, 1968). Nevertheless, the selection of geoscience and psychology students was a deliberate choice. Training in topographic map reading and related spatial ability is a key skill for geoscientists (Liben & Titus, 2012; Titus & Horsman, 2009),whereas this skill set is less relevant to most psychology students.

Generalizability to other settings or contexts (ecological validity) was also limited due to the specificity of the instructional materials (Bracht & Glass, 1968). For example, the ARS may not be similar enough to other three dimensional TUI's to generalize to other AR devices in terms of interaction through tactile feedback. Moreover, the multimedia instruction and the learning practice activities were specific to the maps used in the experiment. They lacked some of the additional details found on U.S. Geological Survey topographical maps. Consequently, this experiment did not cover all of the skills necessary for full competency in reading topographic maps nor other types of maps. This was a deliberate design choice as excluding latitude, longitude, contour interval numbering, and topographic symbols (e.g., buildings, roads, railway tracks, hachure marks, survey markers, etc.) simplified the map and thus isolated the number of potential intervening variables associated with conflating verbal and spatial ability. Other research has proposed that the decoding of the meaning of symbols is processed through verbal reasoning (Schnotz, 2002). Removing the need to decode topographic symbols permitted analysis of participants' spatial reasoning without the potential confound of simultaneous verbal processing used when dealing with signs and symbols. Furthermore, the presence of multiple map symbols can negatively affect legibility (Phillips & Noyes, 1982).

The use of color was another limitation to implementation of the maps used for instruction and assessment. The ARS base map (and the subsequent 2D paper map derived from it) could only be rendered with color gradation to denote relative elevation, whereas the TMA maps were available in black and white only. Forgoing the use of color versus black and white was a possible mediator when reading topographic maps, other studies found varying effects (Phillips, Lucia, & Skelton, 1975; Phillips & Noyes, 1982a; Potash, Farrell, & Jeffrey, 1978; Shobesberger, 2007).

Several issues associated with the scope of the experiment and other areas of related research were deliberately excluded from this study. As an example, force feedback from tactile interactivity with the ARS was not addressed as a device that could measure the degree of tactile interaction in the ARS does not currently exist. Other measures of cognitive load such as fMRI or EEG data (Antonenko, Paas, Grabner, & van Gog, 2010; Parneet Kaur & Sheveta Vashisht, 2013) were also not within the means of this study and so were excluded as units of analysis. Differentiating between the relative amounts of germane, intrinsic, and extraneous loads could not be directly quantified from examination of a combination of learner performance and subjective ratings of mental effort. Only a total cognitive load, as participant-reported MEMS value could be collected.

Although this study compared the dependent variables of learning performance and mental effort, it could not directly attribute the instructional treatment to any specific, measurable improvement in spatial ability; the duration between instruction and testing was too short to reveal any lasting and significant change. A longitudinal analysis might reveal the degree of learning retention of spatial ability skills over time, which was also not within the scope of this study. Furthermore, any positive effects would likely apply only to topographic map reading, a subset of overall spatial ability.

Significance of the Research

Theoretical significance. There is a shortage of research on recent technological advances in cartographic visualization in multiple dimensions and in use of multi-sensory map interfaces. This study helped to fill this gap in the literature with original research on the ARS as an innovative learning tool and as an example of ARS as more than a novelty. Furthermore, two unique propositions were assessed by this study. First of all, high degrees of embodied interaction would result in increased learner performance due to multi-sensory inputs (vision plus touch). Extending from this first claim, one can conclude that optimum germane cognitive load from multi-sensory inputs would be evident from lower cognitive load measures and high learner performance test scores. This study also contributes to the relative shortage of empirical studies on embodied interactions using tactile feedback; most of the literature in this field focusses on auditory and visual interaction (Minogue & Jones, 2006).

Practical significance. The long-term goal is to use evidence-based research from this study to guide recommendations for effective teaching, learning, and assessment strategies that will enhance topographic map reading skills using AR. Findings will be initially targeted to this institution's geoscience department to support basic topographic map reading and 3D visualization skills instruction for novice geoscientists. Current instruction in this area is lacking according to faculty members at the study site, yet it is regarded as a desirable skillset for aspiring geologists as observed by L. Tapanila, geoscience lab instructor, (personal communication, September 15, 2015) and D. Pearson, structural geologist and introductory geoscience instructor (personal communication, Sept 18, 2015).

In terms of broader impacts, research on multi-sensor interfaces, such as the ARS, can support further inquiry into improving spatial ability skills as well as supporting under-served student populations, such as the visually impaired, who would benefit from tactile feedback. Results from this study can also have direct impacts by informing development of other innovative AR interfaces that are accompanied by teaching and assessment resources informed by instructional design best practices in map instruction for geoscience and K-12 Social Studies curricula. The long term goal is to validate interactive, multimodal map displays using AR and VR as effective learning tools for interpreting complex spatial data that supports geoscience decision making.

Chapter II

Literature Review

The aim of this literature review is to connect theory and empirical evidence on spatial cognition, dimensionality, multimodal learning, and embodied interaction to explain how interactive, 3D visualizations can promote germane cognitive load, such that learners with both high and low spatial ability can optimize learning with a minimum of mental effort. Although this study's research context is narrowly restricted to a comparative analysis of a unique 3D interactive topographic map that uses both visual and tactile sensory cues to teaching basic topographic map reading skills to novices, other studies will be addressed that, although they may use different interfaces and methods, can provide theoretical connections to reconcile multimodal learning, embodied interaction, and spatial cognition perspectives. Due to the multidisciplinary scope of this study, this review of the literature will focus on four key themes that were briefly summarized in Chapter 1: (1) cartographic visualization, (2) spatial ability and cognitive load theory, (3) tangible user interfaces and tactile feedback, and (4) multimodal learning through embodied interaction. This thematic review of recent empirical research and theory will thus provide a foundation for identifying studies that contributed to this study's research design. The advantage of this structure is that it allows for a seamless review of the factors (map type and low or high spatial ability groups) and the dependent measures (mental effort and learner performance) used in this study.

Cartographic Visualization

This section will limit itself to a discussion of the relationships between topographic maps and the spatial skills required to interpret them correctly. A review of cartographic principles that inform design of maps is not relevant to this study. Nor is a discussion of other types of maps pertinent.

Using topographic maps. Maps serve as a permanent record of space and place, acting as “spatial data handling tools” (Dorling & Fairbairn, 1997, p. 1). Maps do this by “reduc(ing) the spatial characteristics of a large area in order to make it observable” (Niedomysl, Elldér, Larsson, Thelin, & Jansund, 2013, p. 88). The ability to read maps is an essential skill for everyone. Yet interpreting a map, like using any tool, presents a variety of cognitive demands on the user. Simple map reading tasks include measuring and comparing distances to identify optimum travel routes or searching for locations with specific features are frequently represented as map symbols. These tasks would be applicable to thematic maps showing roads, cities, political boundaries, etc. More complex tasks may range from identification of similarities and dissimilarities within a landscape as a tool for both determining relative position and direction of travel to measuring elevations and steepness of slopes during wilderness travel. These higher-level spatial skills are typically associated with topographic map use.

Topographic maps represent an abstract representation of the three dimensional (3D) world in a two dimensional (2D) format, whether on paper or a computer screen. Topographic map use contour lines to illustrate shapes and elevation; all points along a contour line are at the same elevation. Hegarty (2013) recommended designing topographic maps with only enough detail to enable the user to access the most relevant information efficiently and with high accuracy. No extraneous details like topographic symbols or labelled index contours should be included as these features add extra, unnecessary detail and thus impose extraneous cognitive load. Colors or shading can be used to simplify a map by showing zones of similar elevation and are recommended (Hegarty, 2013; Muehrcke, 1978).

Translating maps from 2D to 3D representations and vice-versa necessarily introduces some degree of distortion which can affect a user's ability to perform some spatial tasks (Niedomysl, Elldér, Larsson, Thelin & Jasmund, 2013; Smallman, John, Oonk, & Cowen, 2001) such as identifying a precise location, which favors a 2D representation, or visualizing topographic shapes, which favor a 3D representation (Savage, Wiebe, & Devine, 2004; St. John, Cowen, Smallman, & Oonk, 2001). Smallman et al. (2001) posited that users must mentally rotate 2D views to arrive at a 3D perspective; this action was not required for 3D viewing as all three dimensions are integrated, which provides context cues such as perspective, elevation and occlusion (landscape features that block views from one location to another). Judging altitude differences is more complex in 2D maps than 3D maps. Research by Savage et al. (2004) came to the same conclusions.

The dimensionality debate. The debate over map dimensionality and the emergence of tangible user interfaces, such as the Augmented Reality Sandbox (Reed et al., 2014), provided the impetus to compare 2D and 3D maps with varying degrees of user interactivity and feedback. Numerous studies contest learning retention from 2D versus 3D maps (Huk, 2006; Oulasvirta et al., 2009; Smallman et al., 2001; Savage et al., 2004; Schobesberger & Patterson, 2007; Tory et al., 2005). Although some research questions the superiority of 3-D versus 2-D map representations (Cockburn, 2004; Pedersen, Farrell, and McPhee, 2005), other studies (Dalgarno, Hedberg & Harper, 2002; Popelka & Brychtova, 2013; Tavanti & Lind, 2001) disagreed. Some map reading studies examine performance test scores (Collins, 2014; Tavanti & Lind, 2001), while others measure learner preference (Anthamatten & Ziegler, 2006). Based on the diversity of past research on topographic map dimensions, two perspectives are evident: (1) preference for a 2-D map over 3-D (Collins, 2014; Niedomysl et al., 2013), (2) a combination of a 2-D map for measuring distances or place name recall and 3-D for displaying elevation, perspective, or relative location (Koua, MacEachren, & Kraak, 2006; Savage et al., 2004; Smallman et al., 2001). A high level of mental effort imposed by 3-D graphical displays (Amini et al., 2015; Bunch & Lloyd, 2006; Van Der Land, Schouten, Feldberg, Van Den Hooff, & Huysman, 2013) was proposed as one causal factor to account for the unrealized potential of 3-D maps as effective learning tools that also promote spatial thinking skills (Dalgarno et al., 2002; Velez et al., 2005).

Three studies warrant more detailed discussion for their empirical findings and for their theoretical contributions to spatial ability and cognitive load when using 2D or 3D maps. Early research on possible connections between cognitive load and topographic map reading skill, was referred to by Eley (1988) as "cognitive processing" (p. 372). In one of Eley's experiments, expert map users were measured on reaction time and correct interpretation of 3D land surface features drawings extracted from topographic maps. It was proposed that users created a mental model of a sample landscape and compared it to a series of examples to find the correct match and cognitive load was inferred by variability in user response times. Elevation, based on spacing of contour lines, had a significant main effect on correct responses of the 3D landscape depicted at an angle of 30 degrees compared to orientations of zero, sixty, 120, 180, 240, or 300 degrees. This finding suggested that users created a simplified mental model of the mapped landscape with key features (valleys, hills, etc.) identified based on distinctive arrangement of contour lines and the spatial distribution of these distinct features on the map; perspective angle was a significant factor affecting comprehension. Thus, mental rotation of the map examples was an indicator of difficulty, based on response scores. A second experiment (Eley, 1988) proposed that viewing 3D landscapes yielded both faster and more accurate responses than mentally rotating a mental model of a 2D landscape view. Eley concluded that experienced map users created a partial mental representation to identify key features, rather than a holistic representation of the entire mapped landscape from a particular angular perspective. The mental model was thus based on key features based on identifiable landform shapes to aid identification; a process Eley labelled as selective encoding. Pick & Thompson (1991) also reported this effect. In a second study by Eley (1991), expert map users were presented with maps containing more or less detail. Results indicated that there was an upper limit as to the amount of extra detail (either indication of drainage valleys or extra contour lines) required by expert users to correctly match a sample map to the correct target example. Extra information did increase response time, which Eley attributed to the higher cognitive load required to process more visually complex maps. No measure of cognitive load was collected during either of these studies, apart from the inference that extra time was attributed to additional mental processing as working memory limits were reached. The implication is that too much superfluous detail in map displays, beyond identifying key features necessary to complete a defined task, can overload users working memory, thus hindering correct interpretation in a timely manner. Although Eley did not extrapolate this argument to less experienced map users, it is plausible that users with limited topographic map reading experience might find it less helpful to read topographic maps that were rich in detail, as the extra visual information may rapidly overload their more limited working memory capacity compared to experienced users with more developed schemas for interpreting complex topographic maps.

This issue of extra detail in 3D maps identified by Eley (1991) was explored in two separate studies by Niedomysl, Elldér, Larsson, Thelin and Jansund (2013) and (Savage et al., 2004). Niedomsyl et al. proposed that adding a third dimension to show elevation information imposed additional cognitive load when transferring a 2D image to 3D. The maps used in their study were rendered in 2D and 3D formats; subjects in the 2D condition had better recall of information. Savage et al. claimed that topographic maps pose comprehension problems for novices (Gilhooly, Wood, Kinnear, & Green, 1988). Their study found increased accuracy and time for 2D tasks that were not dependent on elevation data. 3D maps were only beneficial when dealing with elevation-type questions yet there was a small enough difference that 3D maps elevation perspective did not justify their use over 2D maps: “(there) is little support in these results using for using this style of 3D topographic maps in problem solving and data extraction tasks.” (Savage et al., 2004, p. 1797). The results of these two studies indicate that cognitive scaffolds are needed to help users increase their comprehension of 3D maps. Multimodal learning and embodied interaction are two strategies that shall be explored to enhance germane cognitive load.

Evaluating topographic map reading skills. Weisberg, Newcombe, & Shipley (2015) tested 261 participants to establish a reliability rating of the Topographic Map Assessment test (TMA). It is important to note that only women were tested, which may have biased the results. The effects of different factors on topographic map skills were examined (hand gestures to represent terrain features in one experiment and a verbal comparison task of 2D and 3D maps in a second experiment). Spatial ability and cognitive load were not measured. A follow-up study (Atit et al., 2016) and also reported in (Newcombe et al., 2015) used a large female-only sample ($n = 272$) and assigned participants to high and low spatial ability groups based on user-reported topographic map experience and results on three spatial ability tests. TMA score was weakly correlated with prior topographic map experience ($r(270) = .16$, $p = .01$) and test results on the Water Level Test of spatial ability ($r(270) = .24$, $p < .001$). High and low spatial ability participants were evenly assigned to four treatments. The effects of pointing and tracing contours, on a 2D topographic map, replicating topographic features with gestures text-only instruction, and no instruction were measured. Participants identified steep and shallow slopes, hills, ridges, and valleys with a set of practice problems followed by administration of the TMA. A mixed-methods ANOVA revealed that the pointing and tracing group, the gestures group, and the text-based group significantly outscored the no instruction control group. The pointing and tracing group had the highest TMA scores; there were no significant differences in scores between the gestures and text-only instruction groups. Further analysis revealed that participants fared better on elevation-type questions compared to questions identifying topographic shapes. The three-dimension step contour models used for each group acted as a tangible user interface for the pointing and tracing group. Therefore, it was plausible that tactile feedback and embodied interaction may have accounted for the highest scores of the point and trace group. This possible explanation was not addressed by the researchers. The degree of embodied interaction was likely high in the point and trace and gestures group, but not as high when compared to the ARS. Since the three-dimensional stepped contour model did not change shape in response to user actions its degree of interactivity would be comparable to this study's 2D topographic map group. A discussion of embodied interaction from tactile feedback as a cognitive scaffold will be addressed in the section on multimodal learning and embodied interaction provided by tangible user interfaces.

Since the TMA covers the topographic map skills found in Table 2 and is a reliable (α =.76) instrument that has been used in two studies (Atit, Weisberg, Newcombe, & Shipley, 2016; Newcombe et al., 2015), if it's questions are aligned with instruction, the TMA, as a posttest assessment, is expected to be an accurate measure of learner performance. Combining the TMA raw score with MEMS ratings for each question would enable overall comparisons of learner performance and mental effort between the AR and non-AR conditions, it could also identify potential correlations between low scoring questions and high mental effort.

Spatial Ability and Cognitive Load Theory

This section will define and identify spatial ability, its characteristics, and how it can be measured. Spatial ability has been used as a predictor of competency when using

2D and 3D representations either as a factor (Hays, 1996; Huk, 2007) or a covariate (Savage et al., 2004)

Spatial ability characteristics. Spatial information is largely dependent on vision and has six unique characteristics that can be experimentally verified (Kimura, 2000). Of these constructs, four are relevant to this study because of their affinity for map reading: (1) spatial orientation (the ability to recognize the orientation of 2-D and 3-D shapes); (2) spatial location memory (recall of a spatially-ordered sequence of objects or features); (3) spatial visualization (a capacity to recognize and quantify orientation changes in a scene); and (4) disembedding (finding a specific object within a more complex field of other subjects). Vision is best for discriminating shape and color, while touch can examine physical properties, such as texture, weight, presence of moisture, elasticity, and viscosity. Touch is rarely considered as a component of spatial ability (Minogue & Jones, 2006). A discussion of touch as a cognitive scaffold for enhancing spatial thinking will be addressed in the section on multimodal learning and embodied interaction.

Researchers have identified and classified a range of spatial ability skills as they apply to spatial cognitive processes as shown in Table 1. A direct, one-to-one correspondence or comparison of spatial ability skills between authors is not intended, Table 1 merely illustrates the range of spatial ability skills identified in multiple studies, and although there is some agreement in terminology across studies (particularly for Condition, Location, and Connection), many spatial abilities are given different descriptors, depending on the researcher's theoretical perspective or based on empirical findings. This diversity of characteristics indicates that a consensus on classification and nomenclature for spatial ability skills has not yet been reached. This review will focus on spatial ability skills associated with reading topographic maps. They are summarized in Table 2. The TMA (Newcombe et al., 2015) measures all of these skills.

Measuring spatial ability. Although some researchers used the term *spatial thinking* or *spatial cognition* rather than spatial ability (Liben & Titus, 2012; Wakabayashi & Ishikawa, 2011). This review will use spatial ability as it encompasses cognitive processing of spatial data. Overall spatial abilities have been measured with a wide variety of tests. Some tests measure a subset of spatial skills, or are more comprehensive in scope. The problem is that the identification of spatial skills to be measured are dependent on how they are identified and categorized. Table 1 highlighted the range of spatial ability characteristics. Only a few that have either been widely used or have been validated as a reliable instrument will be named and/or briefly described in this section.

Spatial visualization/orientation and mental rotation of 3D objects are two broad categories of spatial ability tests. The Paper Folding Test (PFT) (Ekstrom, French, Harman, & Derman, 1976) measured spatial visualization by requiring users to compare a sheet of paper that has been folded and hole-punched multiple time to one of five possible target samples. According to Mohler (2008), it does not appear to favor either gender, unlike many other spatial ability tests. The PFT was used as one measure of spatial ability in a pilot study, but the scores were not normally distributed, based on a Shapiro Wilk test. The distribution was positively skewed and indicate a slight ceiling effect and so the scores were discarded. The Spatial Ability Thinking Test (STAT) (Collins, 2014; Lee & Bednarz, 2012) had the advantage of assessing overall spatial ability and has been used to comparatively measure spatial ability associated with use of a 2D paper map and a 3D virtual globe (Collins, 2014).

Table 1

Comparison of Spatial Skills Characteristics and Classifications

Gersmehl & Gersmehl, 2007	**Golledge et al., 2008**	**Janelle & Goodchild (2009)**	**Kastens & Ishikawa, 2009**	**Bednarz & Lee, 2011**
Condition Location Connection	Identity Location Connectivity Distance Scale	Objects and Fields Location Network Distance Scale	Answering Questions about a Terrain by Referring to a Map Comparing a Map with the Represented Space	Identification and Classification of Map Symbols Map Navigation
Comparison Aura Region	Pattern Matching Buffer Adjacency, Classification	Neighborhood and Region	Comparing a Map with Another Representation Perspective Taking*	Generalized or Abstract Boolean Operations
Hierarchy Transition Analogy	Gradient, Profile Coordinate			
Pattern	Pattern Arrangement, Distribution Order, Sequence		Recognizing Patterns and Shapes*	Recognition of Positive Spatial Correlation
Spatial Association	Spatial Association, Overlay/Dissolve Interpolation Projection Transformation	Spatial Dependence, Spatial Heterogeneity	Mentally Rotating an Object and Envisioning Scenes from Different Viewpoints*	Map Visualization Overlay

Note. This table is based on Table 1 from Bednarz and Lee (2011), p. 17, with recent research from Ishikawa and Kastens (2009) and Bednarz and Lee (2012) added by the author. Ishikawa and Kastens (2009) subdivided spatial ability into categories, each with specific spatial skills (basic spatial abilities for Geoscientists, map use in a real-world setting, understanding where you are relative to a map, and topographic map skills). Items marked with an asterisk (*) are drawn from Ishikawa and Kastens set of basic spatial abilities category as they are relevant to the learning activities and assessments used in this study's design.

Table 2

Spatial Skills Associated with Topographic Map Reading Competency

Estimate the height of a terrain marked on the map
Indicate the height of a terrain marked on the map
Judge which direction a river would flow
Indicate the shortest route between two points on a map without going over a certain hei
Judge whether a person standing at a specific point would be visible at another point on
Mark the highest and lowest points on a map
Compare sketches of different terrain and identify a correct match to a sample map
Identify topographic profiles from line segments on a map

A more recent test assesses spatial orientation instead of spatial visualization. The Spatial Orientation Test (SOT) (Hegarty, Kozhevnikov, & Waller, 2004; Kozhevnikov & Hegarty, 2001) measures allocentric (your spatial location relative to other objects) and egocentric (the location of objects relative to where you are standing) spatial abilities. For example, the subject may be asked to visualize the direction they would be facing and the perspective they would see from that location. As a more recent type of spatial visualization test (the PFT was developed in 1976), it has been used in two other studies that measure spatial ability (Atit. et al., 2016; Weisberg, Nardi, Newcombe, & Shipley, 2014) and which also use the TMA. Less common spatial tests include the Paper Form Board test (Linn & Petersen, 1985), the Water Level test (Hecht & Proffitt, 1995), the Arrow Span, Perspective Taking Ability, and Virtual Navigation tests (Quarles et al., 2008b).

The Mental Rotations test (MRT) has been widely used and validated with large samples (Hegarty & Waller, 2004; Lovett & Forbus, 2013; Peters et al., 1995; Vandenberg & Kuse, 1978). The MRT measures a subject's ability to manipulate and rotate 3D images, which is an important skill for comprehending 3D images and maps (Keehner et al., 2004). The MRT requires the subject to select two of four rotated geometric objects that match the example. The test consists of 24 questions and is timed. Subjects have three minutes to answer questions one to 12 and, following a three minute break, another three minutes to complete questions 13-24. Research has shown that while males tend to perform better on rotational tests, the MRT should differentiate on this ability (Peters et al., 1995).

Once spatial testing has been completed, there are two methods for utilizing the data. Several studies use individual spatial ability test scores as a covariate (Savage et al., 2004). Others categorize subjects into high and low spatial ability categories as factors (Huk, 2007; Kalyuga, Chandler, & Sweller, 2000; Keehner et al., 2004; Mayer & Sims, 1994) although the criteria for designation was not made clear.

Variations in spatial abilities. The ability to comprehend spatial information varies considerably across the general population. Numerous studies have documented differences in spatial thinking skills. Broadly speaking, individuals with low spatial ability tended to score lower on performance assessment, report higher cognitive load, and take more time to complete tasks than individuals with high spatial ability (see Huk, 2007; Jeong & Gluck, 2003; Keehner et al., 2004; Mayer & Sims, 1994; Quarles et al., 2008b). Kimura (2000) summarized studies on spatial ability based on gender: men performed higher at object matching and resizing tasks, with women more competent at

3D alignment of objects. Kimura noted that biological, cultural, and social mediators may have also played a role. Bunch and Lloyd (2006) reported that males generally performed better on spatial ability tests, particularly when transferring images into working memory. Their (assumed) faster processing speed tended to favor the pace of standardized tests of intelligence, which have been sometimes used to assess spatial thinking. Women, on the other hand, were more inclined to score higher at recalling information from long term memory (Bunch & Lloyd, 2006). Other studies compared male and female differences using multiple spatial thinking assessments (Uttal et al., 2013; Velez et al., 2005) Uttal's meta-analysis examined 217 studies of training in spatial skills using sex, age, and different types of training as moderating variables (incidentally, spatial ability improved equally with training for both genders), yet there was no reference to use of gender as a blocking variable. Linn & Petersen (1985) examined effect sizes of differences between males and females on spatial thinking tests, but also did not mention gender as a blocking variable. Research on variations typically treat gender as an independent variable. For example, research on spatial skills measurement in video game play utilized gender as a factor in a multifactorial design (Feng, Spence, & Pratt, 2007).

Spatial ability as enhancer and spatial ability as compensator hypotheses. Numerous experimental studies have used the spatial ability as enhancer and spatial ability as compensator hypotheses (Hays, 1996b; Huk, 2007; Mayer & Sims, 1994; Quarles, Lampotang, Fischler, Fishwick, & Lok, 2008a) to account for differences between high and low spatial ability participants' learning performance results. The spatial ability as enhancer hypothesis states that high spatial ability subjects perform at even higher levels with the support of graphical aids, whereas the spatial ability as compensator hypothesis states the low spatial ability subjects will benefit from using graphical aids. Hence, if some form of learning support is not available, low spatial ability subjects will perform poorly due to cognitive overload, while high spatial ability subjects are not dependent on any learning supports. Their more robust pre-existing schemas enable them to more quickly and easily integrate new information from working memory (WM) to prior knowledge in long term memory (LTM) than low spatial ability subjects who are constructing new schemas while learning, which is less efficient and cognitively more taxing.

Cognitive Load Theory and Mapping

Cognitive load represents the cumulative effect of three types of mental processing associated with cognitive tasks. The three types of cognitive load are intrinsic cognitive load extraneous cognitive load, and germane cognitive load (Clark & Mayer, 2011; Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Sweller, 2010). Intrinsic load is dependent on the mental complexity of the instructional topic and is thus fixed, while extraneous load may vary, depending on environmental distractions during instructions and the quality of the instruction, or lack thereof. Germane cognitive load involves mental processing that contributes to building mental maps or schemas. These schemas contribute to organizing and transferring learning from short-term working memory (WM) into long-term memory (LTM) (Clark & Mayer, 2011; F. Paas, Renkl, & Sweller, 2004; Sweller, Van Merriënboer, & Paas, 1998). These three categories of cognitive load are additive. There is no way to differentiate between extraneous and other types of cognitive load; only total cognitive load can be measured (Bunch &

Lloyd, 2006). It is beyond the scope of this study to differentiate these three types of cognitive load; only total cognitive load will be measured.

Schema formation bridges the gap between WM and LTM. Schema building facilitates learning transfer by structurally organizing large amounts of information into a coherent framework. Schemas can also facilitate map comprehension. Prior knowledge is linked to visual sensory data to produce a cognitive map according to Mayer and Moreno (2003). Therefore, working within memory limits is essential to designing maps that can be easily understood (Bunch & Lloyd, 2006). Since effective instruction and active learner engagement can promote germane cognitive load, designing an easy to read map interface is essential, since only the visual channel is used for maps (Mayer & Moreno, 2003), with exception of the ARS, which also allows for tactile feedback. Oversimplifying a map may help novices, but as learning continues, it may result in an expertise- reversal effect as it does not promote further learning; a very simple map does not generate sufficient germane cognitive load (Mayer & Moreno, 2003). Conversely, an animated map, which bears some resemblance to the ARS map used in this study can add additional complexity, such that cognitive overload is possible (Schnotz & Rasch, 2005). Thus, there is a need to measure and account for the effect of cognitive load, which other studies on mapping have not done.

Cognitive load theory assumes three elements of total cognitive load: mental load, mental effort, and performance (Paas, 1992). Mental load is imposed by the demands of the learning task (akin to intrinsic load), whereas mental effort refers to the total amount of cognitive capacity used for the learning task—the term used in this study. Performance was defined as the end product of the interaction of mental load, mental effort, and any other causal factors. Performance becomes a measure of efficiency and accuracy in performing a task. Thus, measuring student proficiency on a skills or knowledge test is an indirect measure of how much cognitive load was applied to yield a particular score or result. It is not an absolute value, however. The balance between mental effort and performance is referred to as instructional condition efficiency (Paas and Van Merriënboer, 1993) Instructional condition efficiency values are standardized by converting performance and mental effort values to z-scores, with a mean of 0 and a standard deviation of 1. Ratings based on a Likert scale ranging from 1 = very, very low mental effort to 9 = very, very high mental effort (see Table 4 in Chapter 3 for the full MEMS range) were self-reported by test subjects as a gauge of the relative amount of mental effort required to perform a learning or performance task. Students reported little difficulty in self-rating their perception of mental effort during testing. Paas and Van Merriënboer cautioned that this combination scoring of mental effort may not report significant effect sizes. It is possible to have high performance scores with low mental effort expended; this would be the ideal outcome for selecting the most effective map interface. On the other hand, high mental effort scores correlate with lower performance in spite of more mental effort expenditure. This result would be associated in this study with the more cognitively challenging map interface.

MEMS scores were regarded by Paas and Van Merriënboer (1993) to be a “rough estimate” (p. 743), because it assumed a linear relationship between performance and mental effort. This caveat agreed with Brunken, Plass, & Leutner’s (2003) assessment of MEMS as an indirect measure of cognitive load. For this reason, the term mental effort is used only when referring to MEMS scores. The potential limitations of accuracy of MEMS values, based on a self-reported, subjective measure, implies that it should be validated by comparison against a more objective measure of cognitive load.

Measuring Cognitive Load

The goal of effective instruction and learning is minimizing extraneous cognitive load while maximizing germane cognitive load (Antonenko, Paas, Grabner, & van Gog, 2010; Debue & van de Leemput, 2014; Lee & Wong, 2014; Van Merriënboer, Schuuman, de Croock, & Paas, 2002). Although the dual-task method (Brünken & Plass, 2003) is capable of identifying extraneous load by introducing a simple, secondary mental activity that tasks the overall working memory of the subject. The dual-task method has received limited application in multimedia learning studies (Brünken & Plass, 2003) and was not used in this study because of potential limitations on the ecological validity of results (i.e., the intentional distractive effect of extraneous cognitive load on working memory resources would not be normally present in an instructional situation). This was a concern noted by Huk (2007), who eschewed the dual-task method when measuring comprehension of 2D and 3D models in favor of a Likert scale of user-reported cognitive load. The measure of cognitive load chosen for this study, the Mental Effort Measurement Scale (MEMS) (Paas & Van Merriënboer, 1993) assesses total cognitive load (intrinsic, germane, and extraneous) and has high reliability ($\alpha = .90$), although it does not directly specify a state of cognitive overload.

Tactile Feedback and its Unrealized Potential for Multi-Sensory Interfaces

A critical supposition underlying this study is that visual and tactile sensory data are related. As evidence to support this claim, a meta-analysis of 43 studies on multimodal feedback (i.e., visual-tactile sensory data), reported an improvement on reaction time and posttest performance, compared to visual feedback alone (Burke et al., 2006). The difference in sensory information from the ARS—visual and tactile—supports the contention that it will provide additional learning cues from an extra sensory channel compared to the strictly visual cues available to the to the non-AR instructional treatment group.

Despite a sizable body of research on TUI's, few studies have explored whether or not a 3D tactile map interface can reduce cognitive load or enhance learning about topographic map skills and knowledge compared to a strictly graphical 2-D or 3-D map interface. Forlines, Wigdor, Shen, and Balakrishnan (2007), for example, noted that "there is little work that investigates how specific interface properties support users' motor-cognitive processes" (p. 66). The majority of research in this area has been focused on interface development and less on evaluating their effectiveness as a learning tool (Antle & Wang, 2013). Patten and Ishii (2000) ascribed this lack of knowledge as follows: "beyond issues of speed of interaction there is little formal knowledge about the differences between TUIs and GUIs (graphical user interfaces)" (p.41). In spite of a knowledge gap, many different TUIs have been developed and involve some degree of spatial ability testing (Antle & Wang, 2013; Forlines, et al., 2007; Patten & Ishii, 2000; Quarles et al., 2008b; Sharlin et al., 2009). Most TUI studies use physical objects, such as a jigsaw puzzle (Antle & Wang, 2013), a series of wooden blocks (Patten & Ishii, 2000), or a 'stand-in' for another piece of technical equipment (Quarles et al., 2008b). Their effectiveness as a learning support is based on researcher observation and learning performance or retention, but not on cognitive load. These examples are prominent within the literature and bear some theoretical and methodological similarities to this study. Only one study to-date has used tactile feedback to measure learning from an interactive map (Jeong and Gluck, 2003). They added tactile and auditory feedback to a GIS map and discovered that the tactile interface produced the fastest and most accurate responses to the test condition compared to auditory feedback alone.

Tactile systems are multi-sensory (touch plus visual feedback) to perceive and thus help users create a schema or mental map of physical objects and the spatial relationships among them that is built from both tactile and visual data. A meta-analysis of 43 studies on multimodal feedback (i.e., visual-tactile sensory data), reported improved reaction time and posttest performance (g = .77), compared to visual feedback alone (Burke et al., 2006). It is surmised that the difference in sensory information from the ARS—visual and tactile—will provide additional learning cues from an extra sensory channel compared to the strictly visual cue available to a 2D paper topographic map.

There is a lack of research comparing tactile feedback maps against conventional topographic maps and there is currently no reported research on the visual-tactile feedback offered by the ARS (Reed et al., 2014), the TUI chosen for this study. Nevertheless, the research findings on tactile feedback as a learning support for spatial thinking ability have been positive. For example, Wang, Yue, Xiaogang, Yufen, & Meng (2001) proposed that the tangibility of physical objects, through the tactile feedback they provide, are intuitively understandable because of their inherent physical properties; touch provides an extra dimension to facilitate cognitive processing—a process of augmented affordance. Kim & Maher (2005) argue that touch is a spatial modality due to connections between bodily motor functions and our awareness (or lack thereof) of our position in space and in relation to other objects. Tactile feedback enables us to form a schematic map of spatial relationships between physical objects. These positive effect has been most prominent for subjects with low spatial thinking aptitude (Quarles et al., 2008b). This provides a rationale to test for spatial ability as a possible covariate; if pretest and posttest MRT measures were gathered and the rigorous assumption of ANCOVA analysis can be satisfied.

Multimodal Learning

Multimodal learning posits alternative forms of thinking and learning. Although the body of literature is not extensive, it can aid in interpreting this study's results, particularly how the modality of user interaction may aid interpreting experimental data. Other bodily features/functions such as the perceptual and motor systems also enable learning and that mixed-reality environments support learning through multiple senses, including touch and sound.

Multimodal theory has been developed to appeal to diverse learners in elementary school settings. Learning experiences that involve manipulation have shown higher levels of learner engagement among school-aged children (Xie et al., 2008); increased engagement corresponds with higher germane load and thus more effective mental processing through schema building. As previously discussed, these schemas promote transfer of information from WM to LTM. Thus, the visual and tactile sensory inputs of the ARS may promote high cognitive load (which might suppress learning) yet high scores on the TMA posttest. The interactivity of the ARS, in comparison to the low interactivity of the 2-D map, will likely increase germane load in the AR condition and thus form more effective and efficient schemas. The effectiveness of adding a tactile interface to a GIS display, as particularly reported by Jeong and Gluck's (2003) of combining tactile and visual feedback to increase learning performance in less time and with higher engagement, lends support to this study's experimental rationale. A meta-analysis of 43 multimodal feedback studies that supplementing visual feedback with visual-audio and visual-tactile feedback reduced reaction times, led to higher testing scores, but did not reduce error rates. Visual-tactile feedback allowed for multiple, concurrent tasks (Burke et al., 2006)

Embodied Interaction

There is limited evidence-based research on how interfaces facilitate motor-cognitive skill development when problem solving (Antle & Wang, 2013; Price & Jewitt, 2013). Embodied interaction and user-interactivity are synonymous constructs within this study and leverage multimodal feedback through use of TUI's. Linking these constructs forms a bridge between cognitive processing, action, response, and learning.

Embodied interaction implies two-way feedback between the learner and the educational tool. The learner changes the tool and the tool changes the learner, otherwise learning becomes a passive process, which does not promote engagement and germane cognitive load. Bodily movement such as touch and gestures are central to embodied interaction; "by incorporating physical movement and tangible feedback in digital systems, TUIs can leverage the relationship between the body and spatial cognition to engage, support, or improve spatial skills"(Clifton et al., 2016, p. 1). The ARS has tremendous educational potential because the map surface changes in response to user actions in real time and provides tactile feedback to the user.

The embodied interaction inherent to TUI's allow movement and placement of objects through manual or remote control into spatial arrangements as noted in Patten and Ishii's (2000) exploration of a simple TUI that compared how participants spatially arranged information to users performing the same task with a graphical user interface. Touch, moving, and arranging information blocks improved spatial recall compared to the other group. Gestures, as a form of tactile interaction, are typically associated with speech, yet the act of gesturing may be indicative of moving cognitive load processing from verbal to visual-spatial storage (Goldin-Meadow & Wagner, 2005). Thus, embodied interaction from tactile feedback could promote evidence of understanding through explaining concepts or giving oral responses to questions. As already discussed, Atit et al's (2016) TMA scores supported the positive effect of gestures and pointing and tracing on topographic map comprehension, which further support the learning potential of embodied interaction as an aid to cognition.

Contributions of Prior Research to the Design of this Study

The implementation of this study is built upon the research designs of several studies. In terms of general implementation, Moreno & Mayer (2007) recommended a comparative study of interactive and non-interactive learning environments. Pick & Thompson (1991) informed the design of multimedia instruction by using a side-by-side comparison of photos or drawing of a landscape with its match shown as a topographic map. For assessing mental effort, Huk (2007) had participant self-report their cognitive load, using a five point Likert scale of +2 to -2. It was not clear if a verbal descriptor was associated with each value, as per the MEMS rating scale, nor was the frequency of cognitive load measures report. This study will also measure cognitive load, but will use

Paas and Van Merriënboer's (1993) MEMS rating scale instead. Categorizing participants into low and high spatial ability groups for factorial analysis is common practice (Atit, et al., 2016; Huk, 2007; Kalyuga et al., 2000; Keehner et al., 2004; Mayer & Sims, 1994) and will be used in this study. Although it was tempting to use spatial ability as a covariate as per Savage et al. (2004), the more rigorous assumptions associated with ANCOVA, particularly with the assumption of independence of the covariate and the dependent variables (mental effort and learning performance) would be more difficult to realize. Using a block design would likely be more powerful (Maxwell & Delaney, 2004; Myers, Well, & Lorch Jr., 2010) and much simpler to implement. The TMA (Newcombe et al., 2015) is a suitable posttest assessment as its questions cover all of the topographic map skills listed in Table 2. With regard to procedures, Savage et al. (2004) and Niedomysl et al.(2013b), compared 2-D and 3-D maps. In both studies, the conclusion reached was a negligible difference between 2-D and 3-D maps. In contrast, this study's goal was to determine if these there was a significant difference between the non-AR 2D topographic map instruction condition and the AR instruction condition. Since the ARS offers user-interactivity through embodied interaction, the comparison of 2D and 3D maps may yield different results. Atit et al's (2016) three-dimension step contour model is a pseudo-tangible interface that is somewhat similar to the ARS, except their model was static, offering no embodied interaction. Nevertheless, it does use TMA as an assessment tool. For data analysis and interpretation, Huk (2007) and Hays (1996) used a two-way ANOVA with high and low spatial ability groups and a 2D and 3D model as factors. Kalyuga et al. (2000) also employed a factorial design with mental effort ratings and learner performance scores as dependent measures. The spatial ability as enhancer and spatial ability as compensator hypotheses, which have been used to account for performance and cognitive load between low and high spatial ability subjects is a useful explanatory tool due to its use in several similar studies (Hays, 1996; Huk, 2007; Mayer & Sims, 1994; Quarles, Lampotang, Fischler, Fishwick, & Lok, 2008a).

Chapter III

Methods

This study's purpose was to investigate the learning effectiveness of the Augmented Reality Sandbox (ARS)(Reed et al., 2014). A shortage of empirical research on learner performance and cognitive load of the augmented reality (AR) maps, compared to traditional, non –AR paper topographic maps, provided the impetus for this study. Consequently, the study's primary goal was to ascertain if the ARS's higher user interactivity increased learner performance with less mental effort compared to participants using a 2D topographic paper map. In this chapter, the research design is reviewed and descriptions of the sampling, instruction, experimental procedures, data collection, and data analysis are provided. A two-way ANOVA tests potential interactions of instructional condition, spatial ability, learner performance, and mental effort, using the MEMS rating system (Paas & Van Merriënboer, 1993).

Research Design

The research design selected was a posttest-only randomized experiment using matched pairs to answer the six research questions. This study utilized two independent variables: spatial ability (high and low) and instructional condition (non-AR or AR treatments). The two dependent variables were learner performance, based on a participant's total score on the modified Topographic Map Assessment posttest (mTMA) and mental effort, as determined by participant self-reporting their perceived level of mental effort using the Mental Effort Measurement Scale (MEMS) (Paas & Van Merriënboer, 1993).

Population and Sampling

Undergraduate students enrolled at an Intermountain West U.S. university formed a convenience sample (*N* = 477) from six geoscience and two psychology classes. These classes were chosen based on their large enrollment as well as minimal familiarity of topographic maps. This sample is not representative of the university's undergraduate population as it is limited to only these two disciplines. The researcher and two assistants visited each class to explain the purpose of the study, to provide an overview of the process, and to solicit volunteers. A total of 136 students (67 geoscience students and 69 psychology students) volunteered for the study and were randomly assigned to either an AR or non-AR treatment session.

Materials

The following five materials were used in the experimental treatments: multimedia instruction, learning practice activities, the 2D paper topographic map (for the non-AR condition), the Augmented Reality Sandbox map (for the AR treatment condition), and computer and software to support the multimedia instruction and learning practice activities. Multimedia instruction, and the computer and software to support it, were identical for both treatments. The learning practice activities were identical in terms of prompts and outcomes, but were accomplished with either a 2D paper maps or with the ARS.

The process that informed the selection and deign of instructional content, sequencing of instruction, and assessment strategies was the Kemp Model of instructional design (Morrison, Ross, Kemp, & Kalman, 2011). The instructional design rationale and planning steps are can be found in Appendix G. Following a review of contemporary instructional methods for teaching novices to read topographic maps, instruction and assessment tools were selected based on alignment with 17 learning outcomes identified during the instructional design process (Table 3).

Multimedia instruction. Both the AR and non-AR condition received 20 minutes of multimedia instruction on basic skills for reading topographic maps. The entire slideshow plus a script of relevant narration for each slide may be found in Appendix H. Instruction was oriented to novices; it did not assume any prior knowledge. Design of the slideshow adhered to the contiguity, modality, and redundancy principles of multimedia learning theory (Clark & Mayer, 2011; Mayer, 2005). Additional discussion of each principle and how it was fulfilled in the multimedia instruction can be found in Appendix G.

Thirty-two slides were designed using Microsoft Office Mix to cover the 17 instructional learning outcomes described listed in Table 3. Office Mix is an add-on feature to Microsoft PowerPoint; it enabled a seamless and user-friendly method to add narration, a pedagogical agent, screen drawing, and on–screen questions to standard PowerPoint slides. Data analytics could also be collected and grouped by user and by question response

Movement through the slide show was linear—after the narration was completed, the presentation automatically advanced to the next slide. A sample slide is shown in Figure 2. A scroll bar at the bottom of the viewing window and a timer indicated progress. Each slide featured user-controls. For example, learners could adjust narration volume, pause playback, advance to the next slide, revert to the previous slide, or use the table of contents tab to move to any other slide by clicking on the slide sorter

Table 3

Learning Outcomes for Multimedia Instruction, Learning Practice Activities, and the Modified Topographic Map Assessment Posttest

Learners will correctly identify topographic contour lines on a map (Core competency)
Learners will correctly identify topographic contours as a series of concentric circles, with the same elevation at each point on the contour line (Core competency)
Learners will calculate elevations using contour line data
Learners will indicate lowest and highest points on a topographic map
Learners will determine elevation differences between one or more features
Learners will differentiate between low angle and steep slopes, and cliffs
Learners will correctly identify geographical features based on an understanding of shapes and spacing of contour lines: (a) low angle vs. steep slopes; (b) hill vs. mountain; (c) ridge and ridgeline; (d) saddle; (e) mesa vs. butte; (f) valley; (g) spur; (h) depression/basin
Learners will draw (non-AR condition) or sculpt (AR condition) each of these features during learning practice only
Determine if one location can be seen from another based on either (a) obstructing, higher elevation feature, or (b) clear view (no obstruction)
Learners will draw or trace direction water will flow in valleys and drainages and where it will settle to form a basin
Learners will draw or trace a path from one point to another on a topographic map to indicate the least steep (easiest) line of travel
Learners will correctly match a cross-section on a 2D map to the correct elevation profile
Learners will correctly identify 3D views of a given landscape when given a 2D plan view from a sample topographic map
Learners will correctly match a sample 2D plan view to the corresponding 3D landscape view
Learners will correctly match a sample 3D landscape view plan view to the corresponding 2D plan view
Learners will correctly identify the direction they are facing on a 2D map when given a 3D scene
Learners will correctly identify the direction they are facing on a 3D scene when given a 2D map

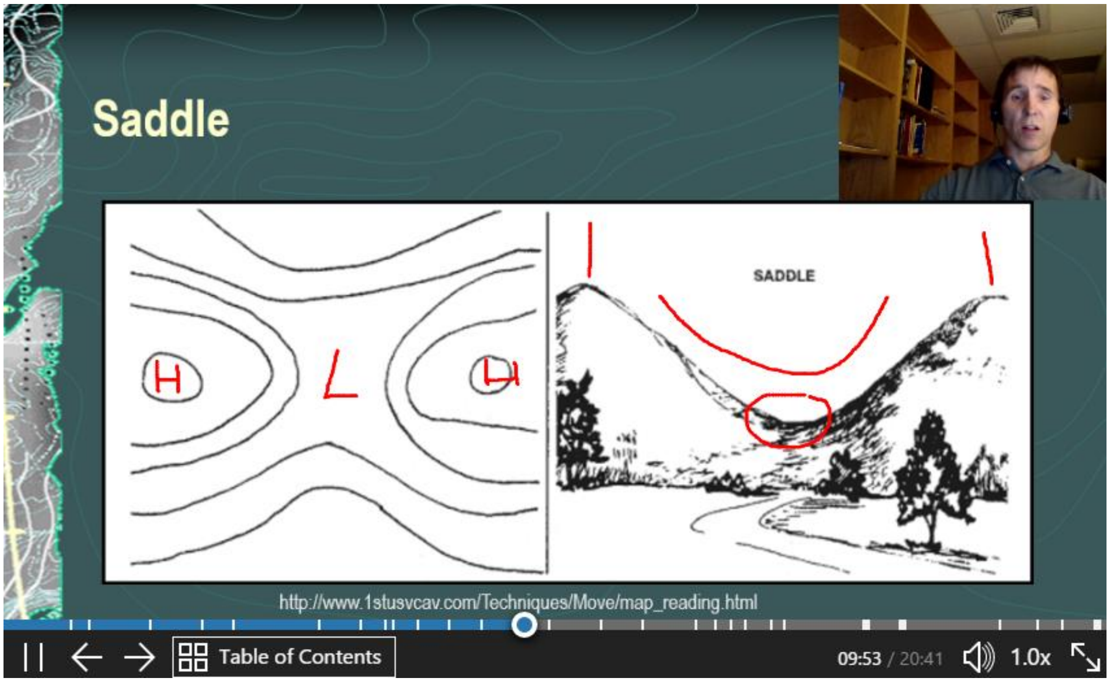


Figure 2. Sample Instructional Slide (slide 13) from Multimedia Slideshow. Each slide featured a pedagogical agent in the top right of the slide. Instruction was provided via narration, supported by accompanying drawing (in red) on the images to identify key visual features in each image. A common set of navigation tools are located at the bottom of each slide within the black tool bar.

The remaining slides, if they displayed any text, were limited to labels, thereby minimizing the amount of reading required and not violating the redundancy principle of multimedia learning (Clark & Mayer, 2011; Mayer, 2005).

The learning progression moved from concrete to abstract topics by starting with simple features, such as gentle and steep slopes, before introducing more complex features such as spurs, which combine ridges and valleys. The example provided in Figure 2 presents a slide with a narrated description of the term ‘saddle.’ This slide featured two side-by-side views of the geographic term: the left image was a 2D view, as would be found on a conventional paper topographic map, whereas the right-hand image illustrated how the topographic feature would appear on the landscape. A simple line drawing was selected for the landscape view to minimize distractive details that might impose extraneous cognitive load.

During pre-scripted narration created and read by the researcher, Office Mix's drawing tool facilitated screen markup using a red drawing pen to direct the learner's attention to key features of the two images. For example, in the 2D topographic contour line example on the left hand side of the slide, the bold letter 'L' indicated the lowest point on the saddle. The researcher timed the explanation to correspond with drawing on the slide. This was followed by narration while drawing a u-shaped line on the landscape view to correspond with the low point. Next, the letter 'H' was applied to label two locations on the 2D topographic contour view as the high points of the saddle and drawing two vertical lines at the high points in the landscape view.

Learner engagement during the instruction was facilitated via two interactive features: questions during instruction and MEMS ratings. The narrator would frequently pose questions when instructing. For example, the narration paused after the narrator asked the learner to identify the high and low points on the saddle. This short delay allowed the learner to mentally locate these points before they were marked on the screen, rather than receiving the correct answer immediately. Six slides at the end of multimedia instruction asked the learner to identify specific features on either a 2D or 3D map view. The correct answer was then marked on the map and accompanied by appropriate narration. Since these questions were rated by the researcher as cognitively challenging, after answering the question and receiving feedback, the learner rated their mental effort from one to nine using the MEMS (Paas & Van Merriënboer, 1993). The purpose of using the MEMS was to ascertain learner adeptness at mentally translating 2D paper topographic map views to 3D views and vice versa.

Learning practice activities. After viewing the instructional slideshow, all learners interacted with their map (either the paper map or the ARS) through a series of common practice learning activities designed to reinforce the multimedia instruction. As the primary interactive task, each participant re-created a representation of specific topographical features that had been covered in the multimedia instruction (e.g., gentle slope, steep slope, hill, mountain, valley, ridge, mesa, basin, and spur). In the non-AR treatment condition, participants sketched the requested topographic feature on a separate piece of paper. For the AR treatment condition, the participant recreated the feature by sculpting the sand surface in the ARS interaction space (see Figure 7). Learning practice also included tasks such as asking the participant to identify the highest and lowest points on the map, locating basins, calculating elevations, tracing travel routes, and indicating water flow following rain. The complete list of 20 learning practice activities are listed in Appendix I with the answer key in Appendix J. After each question, space was provided on the protocol form for scoring and for the researcher or assistant to record the participants self-rated mental effort using the MEMS rating scale of one to nine (Paas & Van Merriënboer, 1993).

2D topographic paper map. To ensure fidelity between the maps used in each instructional treatment condition, the 2D topographic paper map used in the non-AR instructional condition were identical. A screen shot of the ARS map was printed onto poster paper that was the same dimensions as the ARS. The map was laminated to allow participants to draw upon it with a dry erase marker. During the learning practices activities portion of the treatment, the non-AR participants were provided with separate sheets of paper for drawing topographic features. After drawing a feature, they would then circle or draw an X to identify similar examples on the map. Other learning practice questions required tracing lines on the map. Markings on the map were erased after completing each question. The sample 2D topographic paper map is show in Figure 3.

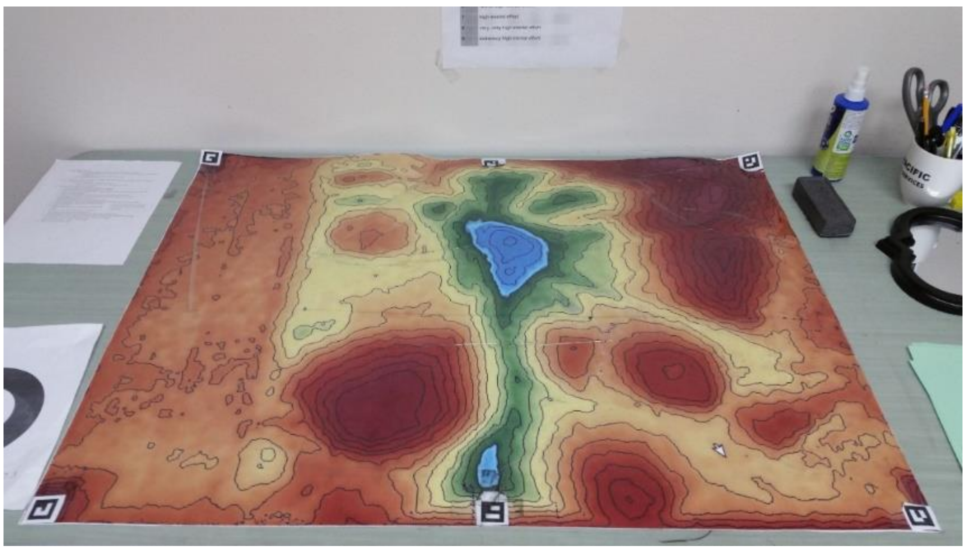


Figure 3. Laminated Paper Map used for the non-AR Treatment Condition. Note that size, relative scale, and colors are identical to the ARS map.

3D Augmented Reality Sandbox (ARS). The Augmented Reality Sandbox (Reed et al., 2014) used in this study was originally constructed as a student project for a geovisualization course. Appendix F provides a detailed setup, specifications, and instructions file that were followed for constructing the ARS. Figure 4 shows the five parts of the ARS: (1) a Microsoft Kinect sensor, (2) a BenQ MX631ST digital projector, (3) an Intel Core i5 PC running a Linux operating system with the Mate GUI,

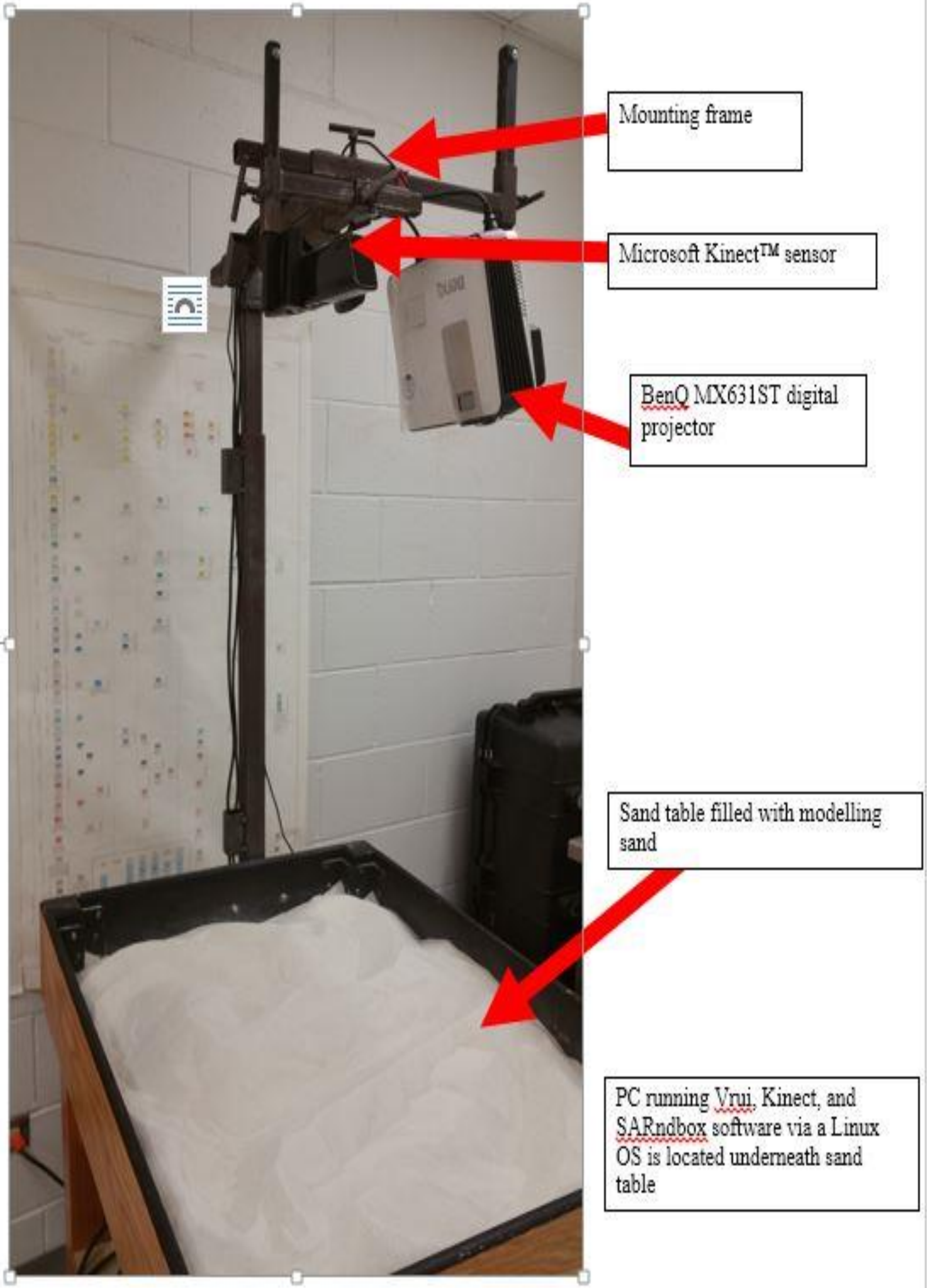


Figure 4. Components of the Augmented Reality Sandbox. In this photo, the system is not activated so that the modelling sand can be shown.

(4) frame for mounting the Kinect and projector at a height of 1.016 m above the mean sand surface, and (5) a 1 m x 0.75 m wooden sandbox filled with 100 dm^3 of modelling sand. This quantity of sand was higher than specified in the ARS design instructions (Appendix F), to permit adding extra vertical relief when modelling hills and additional negative relief for shaping basins. The Kinect's sensor used laser-ranging to measure the height of the sand level and relayed this information through the Vrui, Kinect, and SARndbox open-source Linux software for processing. The resulting image, which corresponded to the relative heights of the sand, was then displayed on the sand surface via the BenQ MX631ST digital projector. Locations below the mean sand 'height' were displayed as low elevation areas (valleys, plains, basins) while locations above the mean height were displayed as high elevation features (hills, ridges, and mountains).

As shown in Figure 5, the map used color (water in blue, green for low elevations, progressing through shades of beige, brown, and red to show increasing elevation, with white as the highest elevation). Gradations of elevation were represented by superimposed black contour lines. Index contours, which indicated relative elevation value with a numerical label, were not included, as is the norm for most U. S. Geological Survey and Canadian National Topographic System maps.

A 20 cm x 70 cm strip of the ARS surface was left bare and defined as the *interaction space* where participants could use the sand within this area to build topographic features without altering the base map (see Figure 5). This feature was not included with the 2D topographic paper map. Instead, participants in the non-AR condition were provided with separate sheets of paper for drawing as the equivalent to the ARS' interaction space.

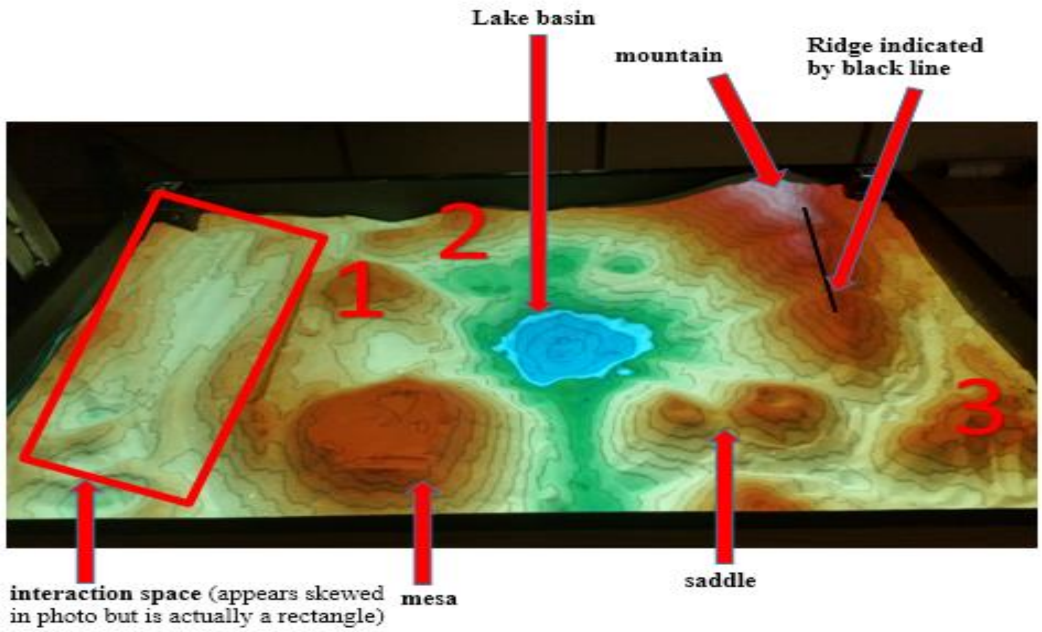


Figure 5. ARS Base Map The interaction space is located on the left side of the sandbox; a ruler was placed in the sand to delineate the boundary of the interaction space and the base map. The base map configuration remained throughout the treatment session, thereby allowing participants to sculpt topographic map features in the interaction space without changing the base map. Labels indicate sample topographic features. The numbers refer to three examples of hills. The black lines represent topographic contour lines. Although not labelled with elevations, during learning practice activities the contour interval (distance between successive contour lines) was 50 feet.

Rainfall was simulated by holding a hand above any spot, thus forming a virtual cloud. Simulated rain fell at the point below the virtual cloud and then flowed downslope, along valleys to settle in basins (Figure 6*)*. This effect demonstrated the path of least resistance water would follow.

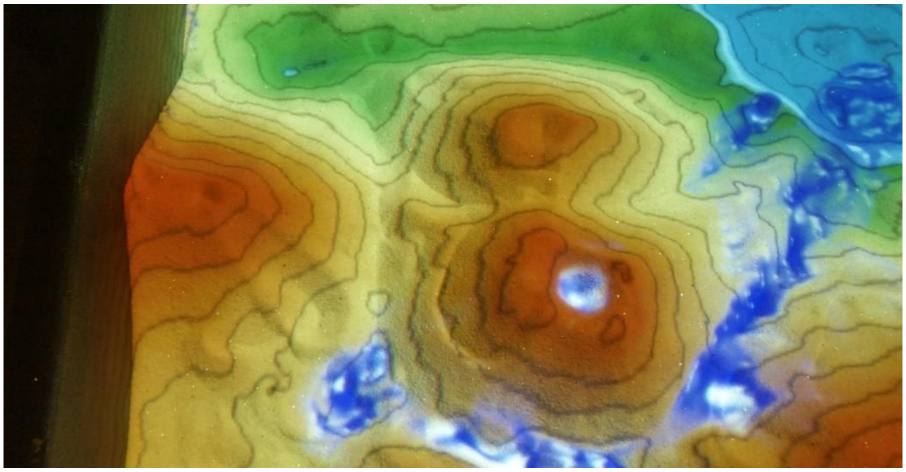


Figure 6. Simulating Virtual Rainfall the ARS. Positioning a hand over the surface creates a virtual raincloud and the resulting surface water flow.

Figure 7 and Figure 8 visually emphasize that the ARS map is a three-dimensional and malleable representation of the treatment map, features that are not available with the non-AR map. Note that Figure 7 shows the opposite view from Figure 8. These views provide perspective in terms of heights and shapes of topographic features.

Computers and software. The instructional slideshow for both treatment conditions was displayed on identical 11” x 17” VGA PC monitors with a screen-mounted speaker. Each monitor was connected two identical desktop PC systems, each with an i7 processor, 32 GB RAM, and a Linux Ubuntu operating system, mouse, and keyboard. An Ethernet connection provided internet connectivity. Loaded applications included Mozilla Firefox to host the Google Docs online demographic survey and a cloud link to Office Mix for the multimedia instruction slideshow.

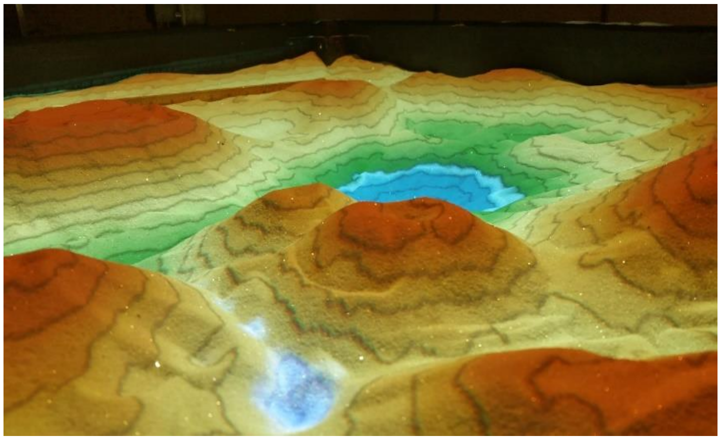


Figure 7. 3D Perspectives from the ARS Map 1. (Looking from bottom right to top left)

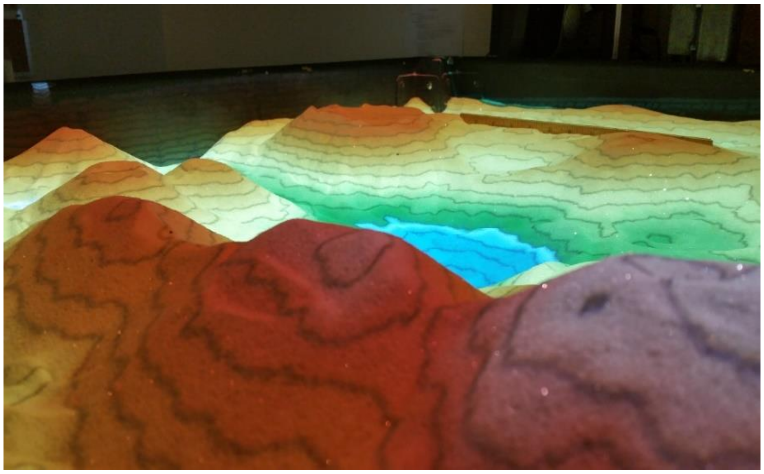


Figure 8. 3D perspectives from the ARS Map 2. (Looking from top left to bottom right)

Instruments

Four instruments were used in the study to provide data on participant demographics (descriptive), two measures of spatial ability (a contributing variable), performance on the mTMA post-test (dependent variable), and mental effort during the posttest (dependent variable).

Demographic survey. The survey was developed using the Google Docs online questionnaire tool. The survey (Appendix C). collected the following demographic information: academic major, academic standing (freshman, sophomore, junior, senior, graduate, or mature/returning student), gender, enrollment in courses that may include use of advanced spatial thinking skills (such as math or engineering), experience with activities that typically use topographic maps (such as hiking, orienteering, hunting, military, etc.), prior use/knowledge of topographic maps (rated on a five point Likert scale), completion of a geography or geography-related course (including the number of years since taking such a course, if applicable), and comfort level or experience with using technology (rated on a five point Likert scale). The intent of including the preceding questions on prior experience with related spatial skills was to identify possible outliers and to aid in interpreting potential relationships between prior experience on performance scores and mental effort ratings.

Spatial thinking ability test (MRT-A). Two widely-used tests associated with measuring spatial ability were used in this study. The Mental Rotations Test (MRT-A) (Hegarty & Waller, 2004; Lovett & Forbus, 2013; Peters, Laeng, Latham, Jackson, Zaiyouna, & Richardson, 1995; Vandenberg & Kuse, 1978) assessed spatial orientation while. This was not a pre-test. Spatial orientation has been identified as important pre-requisites for reading topographic maps (Bednarz & Lee, 2011; Gersmehl & Gersmehl, 2006). This study examined whether participants' spatial ability might have an interaction effect with treatment condition on participants' mental effort or posttest learning performance.

The participants completed the 24 question MRT-A as a measure of ability to perform mental translations associated with varying spatial orientations of 3D objects. The Peters, Laeng, Latham, Jackson, Zaiyouna, and Richardson (1995) variation of the MRT was used in this study; it is an updated version of Vandenburg and Kuse's (1978) mental rotation test. Of the four versions of the MRT test, version A was recommended by Peters, et al., as it yielded the most consistent distribution of scores compared to version B and was less difficult than versions C and D. For each question on the MRT-A, a sample 3D object was depicted on the left side of the page; participants compared this sample with four rotated examples of the target. Two of the examples were correct. One point was given only if participants marked both correct figures; questions that were not completed received no points. Three minutes were allotted for participants to attempt questions one to 12, followed by a three minute break, and then three minutes to answer questions 13 to 24. Specifications, tests procedures, and scoring of the MRT-A are provided in Appendix E.

Mental effort measurement scale (MEMS). The MEMS scale is a subjective rating of mental effort when engaging in cognitive tasks (Paas & Van Merriënboer, 1993). Mental effort is ranked on a scale of one (very, very low mental effort) to nine (extremely high mental effort). Table 4 provides the entire scale and the descriptors used for each number value. The MEMS scale has proven to be a reliable ($\alpha = .90$), albeit somewhat subjective, assessment of mental effort (Paas & Van Merriënboer, 1993; Stark, Mandl, Gruber, & Renkl, 2002). In the current study, participants rated their mental effort using ratings MEMS after completing each task during the learning practice activity and after answering each question on the mTMA posttest.

Table 4

Mental Effort Measurement Scale (MEMS) values (Paas & Van Merriënboer, 1993)

1	very, very low mental effort
2	very low mental effort
3	low mental effort
4	rather low mental effort
5	neither low nor high mental effort
6	rather high mental effort
7	high mental effort
8	very, very high mental effort
9	extremely high mental effort

Modified topographic map assessment (mTMA). The mTMA was used as a posttest assessment of learner performance. The test was not time-limited. The paper topographic maps used in the mTMA focused user attention on the arrangement and spacing of contour lines to assess understanding of slope, gradient, perspective, travel planning, and how typical landforms were depicted. The mTMA is provided in Appendix K with commentary on its structure.

The instructional design process defined 17 learning outcomes aligned with the multimedia instruction, learning practice activities, and the modified TMA (mTMA) posttest assessment (Table 3). The mTMA contained seven general question categories: relative elevation, gradient/slope, perspective, relative direction, stream flow, line of easiest travel, and topographic profiles. The mTMA used a series of 18 unique 2D topographic maps that, unlike the treatment maps used in this study, did not use color to delineate relative elevation. Fourteen questions used reference contours with elevation marks. Four questions provided 3D perspectives along with a 2D representation from which a particular location or perspective was to be identified. Participants were to label 2D topographic contour line views from a word bank of topographic features. These first seven questions were created by the researcher as the mTMA did not require participants to identify topographic features, which was an important skill that was identified during the instructional design process (objective seven of Table 3). Of the total of 25 questions in the mTMA used for this study, 14 questions could be answered with a 2D perspective and eight with a 3D perspective. Within this distribution of questions, there were seven 2D topo map questions to assess distances, elevations, travel routes, or locations; seven questions required identification of topographic features from contour line patterns; and four questions involved labelling stream flow locations and direction of water flow. The researcher added a box after each question where the participant was instructed to rate the mental effort required to answer the respective question.

Procedures

Experimental session times were arranged at one-hour intervals from 8 AM to 6 PM, November 1st 2016 through December 6th, 2016 (additional times were added from Jan 26th through February 1st, 2017 to accommodate another class that elected to participate). A wide range of dates were selected to encompass the entirety of the sample and to provide enough flexibility to accommodate students' schedules. Participants were given a choice of four dates and times that they could commit to attending. An individual email was sent out in advance to each participant indicating their scheduled time along with directions to the Spatial Cognition and Visualization lab. A reminder email was also sent one day prior. Participants who did not arrive for their scheduled time were contacted again by email the same day to arrange an alternate session.

The researcher trained two assistants to arrange the schedule of treatment sessions. They were also trained in the protocols to score and record the spatial ability tests and to implement the protocols for the two experimental treatment sessions under the researcher's supervision. The researcher conducted over 50% of the experimental treatments. While conducting the learning practice activities, the researcher or the assistants followed an answer key for consistent scoring of the learning practice activity. The answer key for the learning practice activity is in Appendix J.

Structure of the experiment. Table 5 provides an overview of the seven phases of the experiment. An experimental protocol form guided the researcher or assistant as they administered the treatment. One form was completed for each participant. A random number was used to identify participants for data collation and scoring.

Recruitment. The researcher and the two assistants visited six geoscience and two psychology classes to recruit volunteers. At each visit, a cover letter (Appendix A) describing the study was distributed to students while the researcher narrated a PowerPoint slideshow outlining the purpose of the experiment, what was expected of participants, how they could benefit by volunteering, and how they would be selected. Those who elected to participate (n = 136) completed a consent waiver (Appendix B). The cover letter, presentation, waiver, and testing was approved by the study site's Institutional Review Board prior to the study's implementation.

Table 5

Procedural Flowchart of the Experiment

	Experimental Phase	Procedure	Data Collected
Prior to Treatment Session (Phases 1-3)	**1.Recruitment**		Consent Form
	2. Spatial skills assessment	Complete spatial skills tests	MRT-A
	3.Random assignment	Participants ranked by spatial skills combined score; pairs randomly distributed to one of two treatment conditions (AR or non-AR) in high or low spatial ability categories	N/A
During Treatment Session (Phases 4- 7)	**4. Demographic data**	At start of experimental session, participants provide demographic data	Demographic Survey
	5. Multimedia instruction	Participants watch instructional slideshow (same slideshow for both conditions)	MEMS collected for select slides
	6. Learning practice activity	Participants complete 20 practice activities with either the 2D paper topographic map (non-AR condition) or the 3D sandbox (AR condition)	Interim performance score on each practice activity MEMs reported after each activity
	7. Performance assessment	Participants complete 2D, paper-based performance assessment (same test for both conditions)	mTMA MEMs reported after each item on mTMA

Spatial ability assessment. The researcher then administered the MRT-A spatial ability test (24 questions) to participants at the end of the class visit. Each test was broken in two sections of three minutes in duration with a three-minute break after each section. Testing of spatial ability allowed participants to be randomly assigned based on test scores and on a rating of high or low spatial ability. Appendix D contains a sample of the MRT-A sample. Appendix E details protocols for administering the test and scoring procedures.

Random assignment. Assignment to treatment condition involved three steps. First, participants were rank ordered based on their score total from highest score to lowest score. Second, participants were sorted into high or low spatial ability categories. A median split of the spatial tests score was used to differentiate those with high spatial ability as above the median score and those who were below the median score were categorized as low spatial ability. Third, the highest ranked participant was randomly assigned to either the AR or non-AR treatment condition using a Samsung Galaxy Note 4 random number app. The next highest score was then assigned to the other condition, forming a matched pair with the same or a very similar MRT-A test score. This process continued until all participants were assigned to a treatment condition. Since each pair had similar spatial ability test scores, they also shared either a low or high spatial ability group. The objective of the matched pairs process was to assign an equal number of participants across the experiment's four conditions: (1) high spatial ability in the ARS condition, (2) low spatial ability in the ARS condition, (3) high spatial ability in the non-AR condition, and (4) low spatial ability in the non-AR condition. The net result of random assignment was 69 participants for the AR treatment condition and 67 for the non-AR treatment condition.

Following assignment, participants were contacted by email to schedule an experimental session. These sessions occurred between November 1st 2016 through December 6th, 2016 and January 20th to January25th, 2017 and were conducted by the researcher or the assistants. The duration of each session was approximately 60 minutes.

Demographic data collection. The experiment began with the researcher or assistant explaining the purpose of the study. Participants then completed the online Google demographic survey. After completing the survey, participants were asked if they were color blind (which would exclude them from the study), and if they normally wore glasses or contacts for reading and/or distance vision. This information was recorded on their experimental protocol form.

Multimedia instruction. After completing the demographic survey, the participant logged into Office Mix via their institution's Google email account to access the online instruction module. The 20 minutes of multimedia instruction was covered by 32 slides. Users could pause, advance, or reverse the playback, if necessary. Slides were narrated by the researcher, keeping the amount of text to an absolute minimum, as per the recommendations for multimedia learning theory (Clark & Mayer, 2011; Mayer, 2005). The duration of each slide was preset; nine of the slides posed a question and after a delay, the response was given. Slides 23, 25, 28, 29, and 30 contained questions that required mentally translating 2D and 3D map perspectives. After answering these questions, participants reported the amount of mental effort required to answer each question with the MEMS rating that was displayed on the next screen. Upon conclusion of the slideshow, the researcher or assistant asked the participants if they had any questions. Learning practice activities followed the multimedia instruction.

Learning practice activities. This phase of instructional practice was designed to reinforce the concepts that were introduced during multimedia instruction through active engagement with either the AR or non-AR map. The participant's attention was drawn to a copy of the MEMS ratings to refer to during practice. Although the learning practice activities were identical, the degree of user interactivity varied for the two treatment conditions. The learning practice activities consisted of 20 question tasks with a maximum score of 28 points, as shown in Appendix I with the answer key in Appendix J. The researcher or assistant read through the questions in order, asking the participant to complete a task or answer a question and then immediately prompted for a MEMS rating (Paas & Van Merriënboer, 1993). Participants were reminded that mental effort was related to the difficulty of the task or question and the amount of time required to respond.

Apart from clarifying the meaning of the question and what was required during learning activity practice, the researcher or assistant only provided corrective feedback if the response was grossly incorrect (i.e., a score of zero for the task or question response). The participants were not informed that they were being scored on the questions or tasks nor did the researcher or assistant reveal their score. The participants were only aware that their mental effort rating was being recorded.

Of the 20 learning practice items, five were responses to a question. Questions followed a specific routine. The participant were asked a yes or no response question (e.g., question 13: "Can you see Point B from Point A?"). The researcher or assistant entered either correct with a check mark or incorrect with an X on the experimental protocol recording form based on the participant's response. The researcher or assistant then asked the participant to rate their mental effort and recorded the MEMS value. The other 15 questions involved an interactive task. For the AR condition, participants were asked to 'build' a specific topographic feature, such as a gentle slope, mountain, valley, etc. in the interaction space of the sandbox. The resulting feature was marked as either correct or incorrect. They were then asked to point to examples of the feature they just created in the map area by touching the sand surface; they were reminded to physically insert their finger(s) into the sand, rather than point with a finger. The score for correct identification was based on the complexity of the topographic feature and how readily identifiable it was on the map. For example, gentle slopes, steep slopes, valleys, saddles, and hills were plentiful and so one point was given for tracing a finger along at least three examples of each. A more complex feature such as a spur, which was more difficult to identify, required only tracing two examples and was awarded two points. Participants were asked to provide a MEMS value after shaping the topographic feature and touching examples for each task question.

The non-AR group sketched topographic lines on a separate sheet of paper to represent the requested topographic feature and then marked examples with a dry erase marker on the laminated paper map using a circle or X. All marks were erased by the researcher or assistant before moving on to the next question. Scoring and MEMS rating procedures were identical to those for the AR condition.

Two questions required tracing a low angle route from one point to another. In the AR condition, participants traced a finger in the sand to show a route, whereas in the non-AR condition, a dry erase marker was used to draw a route on the map. Question 20 involved tracing the flow of water from rainfall at five standardized locations (marked with small flags placed in the sandbox or by letters on the paper map). Participants in the AR condition were given the option to use the virtual rainfall feature, as previously shown in Figure 6, to visualize and then trace water flow from source to outlet. For the non-AR condition, participants were asked to visualize and then draw arrows with a dry erase marker to indicate water flow. Again, participants were asked to subjectively rate their mental effort after tracing or marking all five locations.

Posttest mTMA assessment. Upon completion of the learning practice phase, participants were given the option of a five minute break in a separate location. This was followed by administration of the paper-based posttest. The posttest measured learning performance (as determined by score on the mTMA) and mental effort required to answer each question. Participants were provided with a separate copy of the MEMS scale and were asked to fill-in their rating in a box at the bottom right corner of each question and to do so before advancing to the next question. The mTMA was not a timed-test; participants received as much time as necessary to attempt all 25 questions. See Appendix K for a sample of the mTMA posttest.

After completing the posttest, the researcher or assistant answered any participant questions, thanked them for volunteering, and reminded them not to discuss the experiment with other potential recruits. They were then asked which element of the experiment most increased their understanding of how to read topographic maps: (1) the multimedia instruction, (2) the learning practice activities associated with their respective map interface (2D paper topographic map or the ARS), or (3) the mTMA posttest. The researcher or research assistant recorded the response and provided a $10 honorarium. Since two of the classes did not award extra credit, those students received a $15 reimbursement. Participants entered and their name, student number, and class on a tally sheet to ensure financial accountability and as a record for extra credit.

Data Collection and Analysis

Testing data was collated for each participant in an Excel 2013 template created by the researcher. One assistant collated and entered the practice activity scores and their MEMS values into Excel (the researcher or research assistant recorded totals after each treatment session), while the other used an answer key to score and enter the mTMA scores and the associated MEMS values into a separate Excel template. The intent of assigning research assistants to different scoring and data entry responsibilities was to increase inter-rater reliability.

Four sessions (two for each treatment condition) were randomly selected for inter-reliability comparisons of consistency of researcher and research assistants scoring of the learning practice activities. The researcher conducted the treatment while being observed by both research assistants. They simultaneously and independently scored the participant. After the session, the researcher and the two assistants compared scores. Following completion of all experimental sessions, the researcher randomly selected 15 sets of participant data sets from each treatment collection for data quality in terms of missing scores and accuracy of data entry and tabulation.

Means and standard deviations for the dependent variables posttest mTMA score and posttest mean mental effort were calculated using the Explore analysis functions in SPSS v. 24.0. To satisfy assumptions of a univariate ANOVA analysis, Shapiro-Wilk normality test values were generated for MRT-A test score, posttest mTMA, and posttest mean values, supplemented with visual inspections of Normal Q-Q plots. Boxplots were created to check for outliers and Levene's test was used for assessing homogeneity of variances as part of a two-way ANOVA.

Separate two-way ANOVA tests for between-subjects effects were conducted on learner performance and on mental effort from the mTMA posttest. The factors for the both ANOVA's were treatment condition (AR or non-AR) and spatial ability (high or low). For the first ANOVA on posttest learner performance, the main effects of treatment condition (AR or non-AR) and spatial ability (high or low) and the interaction effect of instructional condition and spatial ability on learner performance were calculated to address research questions one to three. The second ANOVA on posttest mTMA mental effort (mean MEMS values) determined the main effects of treatment condition (AR or non-AR) and spatial ability (high or low) and the interaction effect of instructional condition and spatial ability on mental effort to address research questions four to six. Any statistically significant interaction effects were subjected to simple main effects tests for each factor on the dependent variable. This would be followed by Bonferroni-adjusted pairwise comparisons to determine significance between mean differences and a graph of the interaction to aid in interpreting the effects of each factor. All tests of significance were set at an alpha level of $\alpha = .05$.

Comparison of mean differences between the posttest mTMA scores and mean MEMS values and the corresponding values from the learning practice activity prompted further analysis. The same two-way ANOVA analysis structure used for research questions one to six was replicated using learning practice scores and learning practice MEMS as the dependent measures.

Chapter IV

Results

This experimental study measured learner performance and mental effort associated with two types of map interfaces: the 3D Augmented Reality Sandbox (AR treatment condition) and a 2D paper topographic map (non-AR treatment condition). The study's purpose was to determine whether a high interactivity 3D augmented reality topographic map would be a superior learning tool when compared to the traditional method of using a low interactivity 2D map for teaching novice-level topographic map reading skills. High learner performance on assessments with a minimum of mental effort was the criterion for determining which type of map was more effective as an instructional learning tool.

During the treatment sessions, participants received 20 minutes of multimedia instruction on topographic map skills followed by a learning practice activity using the map interface associated with their assigned treatment. The learning practice activity was guided by 20 question prompts. Both treatment conditions then completed a 25 question topographic map skills posttest (mTMA), which was based on the TMA (Newcombe et al., 2015). Although the questions were different for the learning practice activity and the mTMA, they were based on a common set of 17 learning outcomes. Participants used the nine-point scale of the Mental Effort Measurement Scale (MEMS) to subjectively rate the amount of mental effort, in terms of difficulty or complexity of the question or task, as well as the time required to respond (Paas & Van Merriënboer, 1993) following answering each learning practice activity question and each mTMA question. Forty-six participants, comprising 23 matched pairs, completed all elements of the study for descriptive and inferential analysis. A two-way ANOVA was used to assess if there was a significant main effect of spatial ability, a significant main effect for instructional treatment condition and/or an interaction effect on posttest mTMA scores, with a second two-way ANOVA of the same factors on posttest mean MEMS values. Subsequent examination of the differences in scores and mental effort between treatment conditions during the learning practice activity prompted further analysis. Using the same structure as the six primary research questions, learning practice activity score and learning practice activity mean MEMS value were substituted as dependent measures that would assess performance and mental effort during the learning practice activity session.

Participant Characteristics

Selecting from undergraduate students enrolled in geoscience and psychology classes at an Intermountain West U.S. university (n = 477), a convenience sample (n = 136) volunteered to participate in this study, which was conducted in the fall and winter of 2016 - 2017. Students completed a test of spatial ability and were scheduled for a subsequent treatment session. Of the original sample of 136 enrolled students, 61% (n = 83) attended their scheduled experimental treatment session; those who did not attend were excluded from the study. Due to the matched pairs design, if one of the pair did not report for their treatment session, the other member was excluded from the sample. Therefore, the results from 23 pairs (n = 46) who completed all parts of the study were usable. A median cut score of 10 out of 24 on the Mental Rotations spatial ability test was used to create the high spatial ability and low spatial ability groups. The results of the MRT-A spatial by treatment condition were virtual identical: 53% for the AR condition (*M* = 12.70, *SD* = 4.11) and 51% for the non-AR condition (*M* = 12.22, *SD* = 3.87). There were 12 high spatial ability pairs for the AR and non-AR conditions and 11 low spatial ability pairs for the AR and non-AR conditions. The matched pairs sample (*n* = 46) consisted of 27 females (61%) and 19 males (39 %). Thirty-seven percent of participants were enrolled in an entry level geoscience class (*n* = 17) and 63% were enrolled in an entry level psychology class (*n* = 29). Table 6 summarizes the distribution of participants by course and gender in the two treatment conditions. There was an even distribution of males and females for the two treatment conditions, with a larger proportion of students from psychology courses in the AR condition (63.0% of the sample).

Table 6

Distribution of Participants by Gender and Class from Random Assignment to Treatment Condition

Treatment	**Gender**			**Class**		
	Male	Female	Total	Geoscience	Psychology	Total
AR	9	14	23	6	17	23
Non-AR	10	13	23	11	12	23
TOTAL *n*			46			46

Table 7 data collected included academic major. There was no pattern to the distribution of academic major, apart from health science majors as the largest category

(32.6% of the sample). Academic standing was recorded in Table 8, indicating that freshmen were the largest group in terms of undergraduate academic standing (17), followed by sophomores (11), juniors (6), and seniors (5,) and returning or unclassified students (3). Freshmen and sophomores were near equally represented in the two treatment; there was a small imbalance for juniors and a large imbalance of all five seniors in the non-AR treatment condition.

Table 7

Academic Major by Treatment Condition

Treatment	**Academic Major**										
	HS	A&L	ED	SS	PS	E	CS.	FA	M	Undecided/no response	Total
AR	9	1	3	1	2	4	1	0	0	2	23
Non-AR	6	1	3	0	3	5	1	1	1	2	23

Note. HS = Health Science; A&L = Arts and Letters; ED = Education; SS = Sports Science; PS = Physical Science (Chemistry, Biology, Physics): E = Engineering; CS= Computer Science; FA = Fine Arts; M = Mathematics.

Table 8

Academic Standing by Treatment Condition

Treatment	**Academic Standing**						
	Freshman	Sophomore	Junior	Senior	Mature	Undecided or no response	Total
AR	9	5	4	0	3	2	23
Non-AR	8	6	2	5	0	2	23

As indicated in Table 9, participants rated their experience with using topographic maps on a five point Likert scale: 1 = *I have never used a topographic map* to 5 = *I use topographic maps daily*.

Table 9

Frequency, Percentage, and Median Responses on Topographic Map and Technology Experience by Treatment Condition

	Topographic map experience			**Technology experience**		
	Frequency	**%**	**Median**	**Frequency**	**%**	**Median**
AR			2.00			3.00
1	8	35.0		0	N/A	
2	8	35.0		3	13.0	
3	1	4.3		8	35.0	
4	4	17.4		9	38	
5	0	0.0		1	4.3	
No value reported	2	9.0		2	9.0	

	Frequency	**%**	**Median**	**Frequency**	**%**	**Median**
Non-AR			2.00			3.50
1	7	30.4		0	N/A	
2	7	30.4		4	17.4	
3	5	22.0		7	30.4	
4	1	4.3		5	22.0	
5	2	9.0		6	26.1	
No value reported	1	4.3		1	4.3	

Note. Frequency refers to the number of responses for each value on the five point Likert scale. Percentages were rounded to one decimal place.

The median value and frequencies for each response from one to five are summarized by treatment condition. Ratings were very similar and indicated that most participants had limited familiarity with using topographic maps (Median = 2.00 for AR and non-AR conditions). Likewise, experience using technology was rated on a five-point Likert scale (1 = *not comfortable at all using technology* to 5 = *very comfortable using* technology).

Experience with technology was also very similar between the two groups (Median = 3.00 for AR, and 3.50 non-AR conditions). Both groups reported values that indicated participants had moderate to high comfort levels in using technology for learning. There were no participants who indicated they were not comfortable with using computers. Also, only four participants required glasses for reading and there were no color blind individuals.

Effects of Treatment Condition and Spatial Ability on Posttest Scores

Research questions one to three used a two-way ANOVA with treatment condition and spatial ability as factors with learner performance (posttest mTMA score) as the dependent measure. A copy of the 25 question mTMA posttest can be found in Appendix K. Table 10 summarizes the means and standard deviations of posttest learner performance in terms of posttest mTMA score by condition and spatial ability group. All two-way ANOVA assumptions were met for research questions one to three. There was randomized assignment of ranked pairs of participants to the two treatment conditions (AR and non-AR treatment conditions) with continuous dependent variables (mTMA score for research questions one to three and mean MEMS value for research questions four to six). There was independence of observations between the two treatment groups as participants were exposed to only one treatment and did not interact with participants in the other group. Shapiro-Wilk tests reported posttest mTMA scores as violating the assumption of normality ($p < .001$), which was further confirmed by visual inspection of Normal Q-Q plots, as shown in Appendix L.

Table 10

Mean Posttest Performance Scores by Treatment Condition and Spatial Ability

Treatment Condition and Spatial Group	**Posttest mTMA Score**		
	n	*M*	*SD*
AR treatment	23	26.04	7.02
Non-AR treatment	23	26.48	5.49
High spatial ability group	24	26.88	5.93
Low spatial ability group	22	26.00	6.43
High spatial ability group, AR treatment	12	26.75	9.48
High spatial ability group, non-AR treatment	12	22.58	13.11
Low spatial ability group, AR treatment	11	26.27	8.06
Low spatial ability group, non-AR treatment	11	25.91	5.04

Note. MTMA maximum score was 36; higher scores indicated higher learner performance.

Although outliers were present (all outliers were the lowest posttest scores), ANOVA is robust to violations of normality, so no values were discarded. There was homogeneity of variances, as reported by a Levene's tests for posttest mTMA score (p = .57). Statistical significance for all tests was set at $\alpha = .05$. A copy of the two-way ANOVA summary table for mTMA posttest score is presented in Appendix M.

Research question 1. *Is there a main effect of instructional condition (AR and non-AR treatment) on posttest learner performance, as indicated by a test of topographic map skills?* As indicated in Table 10, the posttest mTMA means between the two treatment conditions were very similar with the non-AR group scoring marginally higher on the posttest. Two-way analysis of variance indicated no statistically significant main effect of instructional condition on posttest learning performance scores, $F(1, 42) = 0.05$, $p = .83$. The size of this non-significant relationship (partial $\eta^2 = .001$) was below the threshold of a small effect size of partial $\eta^2 = .01$ (Cohen, 1988; Meyers, Gamst, & Guarino, 2013).This finding indicated that neither treatment condition made a significant difference on posttest learning performance as treatment condition accounted for only .1 percent of the variance on mTMA scores. It was anticipated that the AR condition would have scored higher on the posttest.

Research question 2. *Is there a main effect of spatial ability on posttest learner performance, as indicated by a test of topographic map skills?* As shown in Table 10, the mean posttest mTMA scores by spatial ability grouping were very similar between those labeled as "high spatial ability" and those with "low spatial ability," with the high spatial ability group scoring marginally higher on the posttest. Two-way analysis of variance indicated no statistically significant main effect of low spatial ability on posttest learning performance scores, $F(1, 42) = 0.03$, $p = .87$. The size of this non-significant relationship (partial $\eta^2 = .001$) was below the threshold of a small effect size of partial $\eta^2 = .01$ (Cohen, 1988; Meyers et al., 2013). This finding disconfirmed the expectation that high spatial group participants would score significantly higher on the mTMA than the low spatial group participants; spatial ability measure accounted for only .1 percent of the variance on mTMA scores.

Research question 3. *Is there an interaction effect of spatial ability and instructional condition on posttest learner performance, as indicated by a test of topographic map skills*? Means and standards deviations for mTMA by instructional condition and spatial ability, as shown in Table 10 varied depending on treatment condition. The high spatial ability AR group scored substantially higher on the posttest than the high spatial ability non-AR group, whereas the low spatial ability AR group scored slightly higher than the low spatial ability non-AR group. Two-way analysis of variance indicated no statistically significant interaction between spatial ability (low or high) and treatment condition (AR or non-AR) on posttest mTMA scores, $F(1, 42) = 0.16$, $p = .69$. The size of this non-significant relationship (partial $\eta^2 = .001$) was below the threshold of a small effect size of partial $\eta^2 = .01$ (Cohen, 1988; Meyers et al., 2013). The interaction effect accounted for only .1 percent of the variance of spatial ability and instructional treatment on posttest mTMA scores. It was anticipated that the two factors of spatial ability group and treatment condition might have a combined effect on learner performance, as suggested by a mean 4.17 point mTMA score difference between the two high spatial ability groups on posttest score, yet this difference was not statistically significant. The findings from research questions one to three suggested that posttest mTMA score, as a measure of learner performance, did not vary significantly based on treatment condition, spatial ability, or a combination of the two factors.

Effects of Treatment Condition and Spatial Ability on Mental Effort

Research questions four to six utilized the same factors in a two-way ANOVA, but with mean posttest MEMS value as the dependent measure. A summary table of means and standard deviations is provided in Table 11. As per research questions one to three, research questions four to six also met the same assumptions of a two-way ANOVA: MEMS values were normally distributed (p = .92) (see Appendix L for Normal Q-Q plots), outliers consisted of the lowest mean MEMS values) so no values were discarded, and there was homogeneity of variances on the posttest mean MEMS values (p = .16). A copy of the two-way ANOVA summary table for mean posttest mental effort (MEMS) value is presented in Appendix M.

Research question 4. Is there a main effect of instructional condition (AR treatment and non-AR treatment) on cognitive load, as indicated by the Mental Effort Measurement Scale? Participants reported their mental effort, using the nine-point Mental Effort Measurement Scale (Paas & Van Merriënboer, 1993), after completing each item on the mTMA. The means and standard deviations for the MEMS by treatment condition are displayed in Table 11 with mental effort slightly lower for the AR treatment group during the posttest. Two-way analysis of variance indicated no statistically significant main effect of treatment condition on mean posttest cognitive load (MEMS) values, $F(1, 42) = 1.63$, p = .21. The size of this non-significant relationship (partial $\eta^2 = .04$) exceeded the threshold of a small effect size of partial η^2 =

.01 (Cohen, 1988; Meyers et al., 2013). This finding was consistent with the expectation that the AR group would use less mental effort during the posttest yet treatment condition accounted for less than four percent of the variance on cognitive load score.

Table 11

Mental Effort Ratings by Treatment Condition and Spatial Ability

Treatment Condition and Spatial Group	**Posttest MEMS Value**		
	n	*M*	*SD*
AR treatment	23	3.88	1.42
Non-AR treatment	23	4.31	1.09
High spatial ability group	24	3.88	1.12
Low spatial ability group	22	4.29	1.41
High spatial ability group, AR treatment	12	4.05	1.28
High spatial ability group, non-AR treatment	12	3.78	.99
Low spatial ability group, AR treatment	11	3.76	1.54
Low spatial ability group, non-AR treatment	11	4.84	.88

Note. Mental Effort Measurement Scale (Paas & Van Merriënboer, 1993) quasi-interval values rated on a Likert scale of one to nine. Higher MEMS values indicate higher mental effort.

Research question 5. *Is there a main effect of spatial ability on cognitive load, as indicated by the Mental Effort Measurement Scale?* The means and standard deviations for MEMS values on the mTMA posttest, as displayed, in Table 11indicated a very small difference between low and high spatial ability groups with the high spatial ability group expending slightly less mental effort during the posttest. Two-way analysis of variance indicated no statistically significant main effect of spatial ability on posttest cognitive load (MEMS) values, $F(1, 42) = 1.07$, $p = .31$. The size of this non-significant relationship (partial $\eta^2 = .03$) exceeded the threshold of a small effect size of partial $\eta^2 = .01$ (Cohen, 1988; Meyers et al., 2013). Spatial ability did not affect mental effort values when completing the posttest questions as spatial ability explained only three percent of the variance on cognitive load score. It was anticipated that the high spatial ability group would use less mental effort relative to the low spatial ability group on the posttest.

Research question 6. Is there an interaction effect of spatial ability and instructional condition on cognitive load, as indicated by the Mental Effort Scale (Paas & Van Merriënboer, 1993)? Means and standards deviations for mental effort by instructional condition and spatial ability, as shown in Table 11, were quite similar, with the high spatial ability participants in the non-AR treatment and the low spatial ability participants in the AR treatment reporting lower on mental effort during the posttest. Two-way analysis of variance indicated no statistically significant interaction of spatial ability (low or high) and instructional condition (AR or non-AR) on cognitive load (MEMS), $F(1, 42) = 4.02$, $p = .051$. The size of this non-significant relationship (partial $\eta^2 = .09$) exceeded the threshold of a medium effect size of partial $\eta^2 = .06$ (Cohen,

1988; Meyers et al., 2013). The interaction effect accounted for only nine percent of the variance of spatial ability and instructional treatment on posttest MEMS values.
The interaction was marginally outside the .05 alpha criterion for statistical significance. The small sample size (n = 46) may have reduced statistical power to detect this interaction. The findings from research questions four to six suggested that mean MEMS values, as a measure of cognitive load, did not appear to vary significantly based on treatment condition, spatial ability, or a combination of the two factors.

Ancillary Analyses: Learning Practice Activity Performance and Mental Effort

After multimedia instruction on basic topographic map reading skills and before taking the mTMA, participants completed a 20 question learning practice activity. (The list of questions and scores assigned to each are shown in Appendix I and the assessment key in Appendix J). Participants were scored on their performance and were prompted by the research team to provide a MEMS rating for each question. Although the learning practice activity and mTMA posttest questions differed, each assessment was aligned to the same set of 17 learning outcomes for novice-level topographic map reading skills (Table 3). The primary difference between the learning practice activity and the mTMA posttest was that participants used their respective treatment maps (either the Augmented Reality Sandbox or the non-AR 2D paper map) during the learning practice activity, while the mTMA posttest was a traditional paper-and-pencil test (see Appendix K). Examination of data collected during the learning practice activity suggested that a substantial difference might exist in learner performance scores and mental effort between the two treatment conditions and between the mTMA results. Therefore, it was determined that ancillary analysis should be conducted on the learning practice activity measures, following the structure and procedures of the six original research questions.

Effects of instructional condition and spatial ability on performance during the learning practice activity. Three research questions guided this ancillary analysis, asking whether there were main and interaction effects of treatment condition and spatial ability on performance, as measured by the 20 question learning practice activity formative assessment. Table 12 summarizes the means and standard deviations of learning practice activity scores by treatment condition and spatial ability. All two-way ANOVA assumptions were satisfied, as per research questions one to six. A Shapiro-Wilk test reported learning practice activity scores as normally distributed (p =.10). These findings were further confirmed by a visual inspection of Normal Q-Q plots (see Appendix L). There were no outliers for learning practice activity scores. There was homogeneity of variances for the learning practice activity, as assessed by Levene's test for equality of variances for practice score ($p = .10$). Statistical significance for all tests were set at $\alpha = .05$. Copies of the two-way ANOVA summary tables for learning practice activity score are presented in Appendix M.

Is there a main effect of instructional condition (AR or non-AR treatment) on learner performance, as indicated by a learning practice assessment? As indicated in Table 12, performance score means during the learning practice activity were substantially higher for the AR condition. Two-way analysis of variance indicated a statistically significant main effect of instructional condition on learning practice activity scores, $F(1, 42) = 19.25$, $p < .001$. The size of this significant relationship (partial η^2 = .31) exceeded the threshold of a large effect size of partial η^2 = .14) (Cohen, 1988;

Meyers et al., 2013). Instructional condition accounted for 31% of the variance in the learning practice activity performance scores. Participants in the AR treatment condition (M = 31.35, SD = 3.90) scored a mean of 5.28 points higher than the non-AR treatment group (M = 26.07, SD = 4.60) on the learning practice activity assessment; this was an anticipated outcome. This result contrasted with the findings of research question one which did not find a significant difference on posttest mTMA scores by treatment condition.

Table 12

Learning Practice Activity Performance Scores by Treatment Condition and Spatial Ability

Treatment Condition and Spatial Group	**Learning Practice Activity Assessment Score**		
	n	*M*	*SD*
AR treatment	23	31.35	3.90
Non-AR treatment	23	26.07	4.60
High spatial ability group	24	28.88	5.05
Low spatial ability group	22	28.02	5.37
High spatial ability group, AR treatment	12	30.00	5.13
High spatial ability group, non-AR treatment	12	27.75	4.92
Low spatial ability group, AR treatment	11	31.82	4.00
Low spatial ability group, non-AR treatment	11	23.54	3.68

Note. Learning practice activity assessment maximum possible score was 38.

Is there a main effect of spatial ability on learner performance, as indicated by a learning practice assessment? As shown in Table 12, learning practice activity score means were very similar between those labeled as “high spatial ability” and those with “low spatial ability,” with the high spatial ability group scoring marginally higher on the learning practice activity. Two-way analysis of variance indicated no statistically significant main effect of spatial ability on learning practice activity scores, $F(1, 42) = 1.14$, $p = .29$. The size of this non-significant relationship (partial $\eta^2 = .03$) exceeded the threshold of a small effect size of partial $\eta^2 = .01$ (Cohen, 1988; Meyers et al., 2013). Spatial ability accounted for three percent of the variance in learning practice activity scores. This finding of no effect of spatial ability on learning practice activity scores paralleled the results of research question two, which found no effect of spatial ability on posttest mTMA scores.

Is there an interaction effect of spatial ability and instructional condition on learner performance, as indicated by a learning practice assessment? Means and standards deviations for learning practice activity performance scores by instructional condition and spatial ability, as shown in Table 12 varied depending on treatment condition. The high spatial ability AR group scored slightly higher on the posttest than the high spatial ability non-AR group, whereas the low spatial ability AR group scored substantially higher than the low spatial ability non-AR group on the learning practice activity assessment. Two-way analysis of variance indicated that there was no statistically significant interaction between spatial ability (low or high) and instructional condition (AR or non-AR) on learning practice activity scores, $F(1, 42) = 3.26$, $p = .08$. The size of this non-significant relationship (partial $\eta^2 = .07$) exceeded the threshold of a medium effect size of partial $\eta^2 = .06$ (Cohen, 1988; Meyers et al., 2013). Considering the large effect associated with instructional condition ($\eta^2 = .31$), the partial η^2 value associated with this interaction suggested that the spatial ability factor may have confounded the main effect of instructional condition, since the combination of spatial ability and instructional condition accounted for only seven percent of the variance on the learning assessment scores. Similar to research question three, it was conjectured that spatial ability group and treatment condition might have a combined effect on learner performance, but this was not evident in spite of the AR group scoring a mean of 8.28 points higher than the non-AR group on the learning practice activity assessment which was statistically significant. The findings on learning practice activity performance appeared to vary significantly based on treatment condition, but not on spatial ability, nor a combination of the two factors.

Effects of instructional condition and spatial ability on cognitive load during the learning practice activity. Three research questions guided this ancillary analysis, asking whether there were main and interaction effects of instructional condition and spatial ability on cognitive load, as measured by the Mental Effort Measurement Scale (MEMS) (Paas & Van Merriënboer, 1993) during the learning practice activity. Means and standard deviations for learning practice mental effort values are summarized in Table 13.

The mean mental effort measurement scale ratings were normally distributed ($p = .40$) (see Appendix L for Normal Q-Q plots). There was one outlier, which was the lowest practice mean MEMS value within the sample so it was not discarded. Levene's test indicated homogeneity of variances on mean learning practice activity MEMS values (p = .823). Copies of the two-way ANOVA summary tables for learning practice activity mean mental effort (MEMS) value are presented in Appendix M.

Table 13

Learning Practice Activity Mental Effort by Treatment Condition and Spatial Ability

Treatment Condition and Spatial Group	**Learning Practice Activity MEMS Value**		
	n	*M*	*SD*
AR treatment	23	3.36	1.03
Non-AR treatment	23	4.23	1.16
High spatial ability group	24	3.46	.87
Low spatial ability group	22	4.16	1.35
High spatial ability group, AR treatment	12	3.40	.99
High spatial ability group, non-AR treatment	12	3.52	.77
Low spatial ability group, AR treatment	11	3.31	1.12
Low spatial ability group, non-AR treatment	11	5.01	1.01

Note. Mental Effort Measurement Scale (Paas & Van Merriënboer, 1993) quasi-interval values rated on a Likert scale of one to nine. Higher MEMS values indicate higher mental effort. The scores of the low and high spatial ability groups differed between learning practice and posttest, hence different values are recorded here.

Is there a main effect of instructional condition on cognitive load during the learning practice activity, as indicated by the Mental Effort Measurement Scale?

Participants reported their mental effort, using the nine-point Mental Effort

Measurement Scale, after completing each question of the learning practice activity assessment. The means and standard deviations for the MEMS by treatment condition are displayed in Table 13 with mental effort lower for the AR treatment condition. Two-way analysis of variance indicated a statistically significant main effect of treatment condition on learning practice activity cognitive load (MEMS) values, $F(1, 42) = 9.93$, $p = .003$. The size of this significant relationship (partial $\eta^2 = .19$) exceeded the threshold of a small effect size of partial $\eta^2 = .14$ (Cohen, 1988; Meyers et al., 2013). Instructional condition accounted for 19% of the variance on mean MEMS values. The AR treatment condition ($M = 3.36$, $SD = 1.03$) used significantly less mental effort (averaging .87 points lower on the MEMS nine point Likert scale) than the non-AR treatment condition ($M = 4.23$, $SD = 1.16$) during the learning practice activity. This effect was not evident for research question four on mTMA mean MEMS values.

Is there a main effect of spatial ability on cognitive load during the learning practice activity, as indicated by the Mental Effort Measurement Scale? The means and standard deviations for MEMS values on the learning practice activity assessment, as displayed Table 13, indicated a difference between low and high spatial ability groups with the high spatial ability group expending less mental effort during the learning practice activity. Two-way analysis of variance indicated a statistically significant main effect of spatial ability on learning practice activity cognitive load (MEMS) values $F(1, 42) = 5.90$, $p = .02$. The size of this significant relationship (partial $\eta^2 = .12$) exceeded the threshold of a medium effect size of partial $\eta^2 = .06$ (Cohen, 1988; Meyers et al., 2013). Spatial ability accounted for 12% of the variance on mean MEMS values. The high spatial ability group ($M = 3.46$, $SD = .87$) used significantly less mental effort (mean -.70 points lower on the MEMS nine point Likert scale) than the low spatial ability group (M = 4.16, SD = 1.35) during the learning practice activity. Research question five did not identify a similar statistically significant difference between spatial ability groups on posttest MEMS values.

Is there an interaction effect of spatial ability and instructional condition on cognitive load, as indicated by a practice assessment? Means and standards deviations for mental effort by instructional condition and spatial ability, as shown in Table 13, varied depending on treatment condition. The high spatial ability participants in the AR treatment condition reported slightly lower mean mental effort than the high spatial ability non-AR group during the learning practice activity. The low spatial ability participants in the AR treatment reported substantially lower mental effort than the low spatial ability non-AR group. Two-way analysis of variance indicated a statistically significant interaction of spatial ability (low and high) and treatment condition (AR or non-AR) on learning practice activity mental effort values, $F(1, 42) = 7.46$, $p = .009$. The size of this significant relationship (partial $\eta^2 = .15$) exceeded the threshold of a large effect size of partial $\eta^2 = .14$ (Cohen, 1988; Meyers et al., 2013).

A simple main effect test for spatial ability indicated a statistically significant difference between spatial ability groups on mean MEMS values, $F(1, 42) = 13.31$, $p = .001$. The size of this significant relationship (partial $\eta^2 = .24$) exceeded the threshold of a large effect size of partial $\eta^2 = .14$ (Cohen, 1988; Meyers et al., 2013). A second simple main effect test for instructional condition indicated a statistically significant difference between AR and non-AR instruction, $F(1, 42) = 16.58$, $p > .001$. The size of this significant relationship (partial $\eta^2 = .29$) exceeded the threshold of a large effect size of partial $\eta^2 = .14$ (Cohen, 1988; Meyers et al., 2013).

Bonferroni-adjusted pairwise comparisons indicated that, among high spatial ability participants in the AR condition (*M* = 3.40, *SD* = .99), the difference in mean MEMS value was .9 Likert scale points higher (95% CI [-.74, .91]) than the low spatial ability participants in the AR condition (*M* = 3.31, *SD* = 1.12), which was not a statistically significant difference, *p* = .831. For high spatial ability participants in the non-AR condition (*M* = 3.52, *SD* = .77) the difference in mean MEMS values was 1.49 Likert scale points lower (95% CI [-.66, 2.31]) than the low spatial ability participants in the non-AR condition (*M* = 5.01, *SD* = 1.01), which was a significantly significant difference, *p* = .001. Recall that lower mean MEMS values equated with less mental effort expended.

As observed in the AR condition, the low spatial ability group's MEMS values were slightly lower than the high spatial ability group, which was a non-significant difference. In the non-AR condition, MEMS values of the low spatial group were significantly higher than the high spatial ability group. This finding indicated that spatial ability mediated mean MEMS values during the learning practice activity depending on which treatment condition participants experienced. High spatial ability participants' mental effort varied less between the two treatment conditions; this was not the case for the low spatial ability group, which reported only slightly less mental effort than the high spatial ability group in the AR condition yet significantly higher mental effort in the non-AR condition. These results suggest that the non-AR condition required significantly more mental effort among low spatial ability participants when compared to high spatial ability participants during the learning practice activity. Yet the AR condition required approximately the same amount of mental effort regardless of spatial ability group.

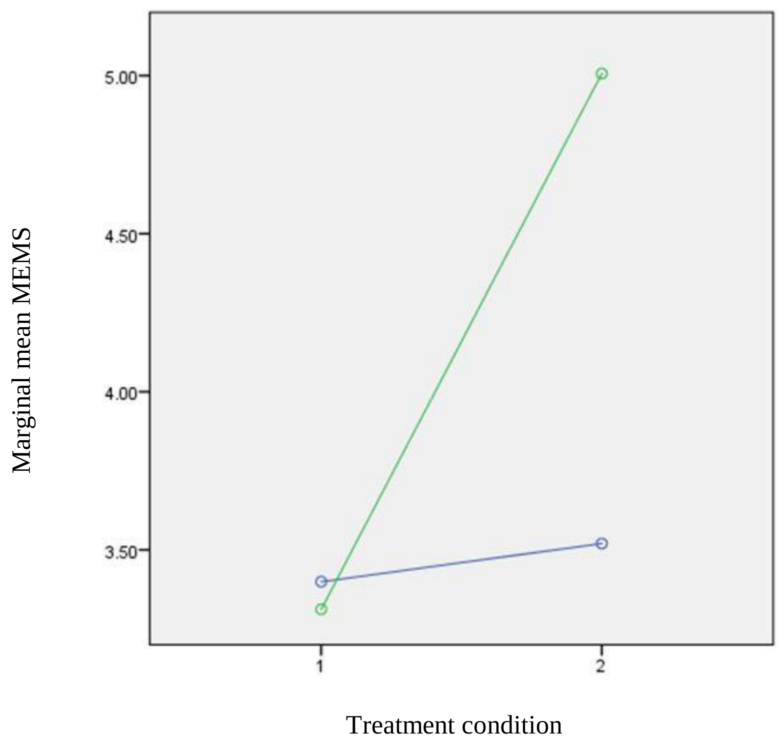


Figure 9. Graph of Disordinal Interaction of Treatment Condition and Learning Practice Activity MEMS (Paas & Van Merriënboer, 1993). 1= AR treatment condition; 2= non-AR treatment condition. The mean MEMS values of the high spatial ability group is shown with a blue line; the low spatial ability line in green. Values in the y-axis are MEMS marginal means from the learning practice activity.

Summary of Results

This study's six primary research questions failed to find statistically significant main effects or interactions of spatial ability and treatment condition on either posttest learner performance or posttest mean mental effort; there were no significant differences on performance and mental effort between the two treatment conditions. The same was not true for the learning practice activity that preceded the posttest assessment. The AR participants scored significantly higher than the non-AR participants on learning practice activity performance scores and did so with less mental effort. Spatial ability and treatment condition interacted to affect mean MEMS values on the learning practice activity. High spatial ability participants' mental effort varied only slightly regardless of treatment condition. This was not the case for the low spatial ability group. Mental effort for the low spatial ability group was dependent on treatment condition during the learning practice activity; lower MEMS values were associated with the AR condition and higher MEMS with the non-AR condition. Consequently, spatial ability and treatment condition experienced during the learning practice activity affected learning performance and cognitive load results.

Chapter V

Discussion

The purpose of this study was to compare two topographic map interfaces to determine if the Augmented Reality Sandbox's (ARS) multimodal feedback and embodied interaction coupled with a 3D map perspective improved topographic map learning performance, with less mental effort compared to the same instruction and practice offered using a 2D paper topographic map. This study provided empirical evidence that the ARS was a more effective teaching tool for novice topographic map users, particularly those with low spatial ability, when compared to the traditional method of instruction using a 2D paper topographic map. This also demonstrates the value of using AR to teach topographic map skills. Results for this study's six primary and six ancillary research questions will be interpreted in terms of spatial ability, cognitive load, multimodal learning, and embodied interaction theories. This will be followed by a discussion of internal validity issues associated with this study, recommendations for using tangible user interfaces for instruction in topographic map reading to support STEM engagement and learning, and areas for future research.

Summary of Findings from Research Questions 1 - 6

Research questions one to six examined performance scores and mental effort ratings associated with mTMA posttest results. There was no significant differences between mental effort and performance for each treatment condition. No main effects or interactions were found among the factors of instructional condition (AR or non-AR), performance score, and mean mental effort. Consequently, neither instructional condition nor spatial ability had a differential effect on performance or mental effort. The AR group's mean score on the mTMA was 72.3% and 73.6% for the non-AR group. Mean mental effort was neutral on the MEMS scale, between *rather low mental effort to neither low and high mental effort.* Instructional condition and spatial ability had a near significant (p = .051) interaction effect on mental effort scores. The next section will describe possible reasons the non-significant findings.

Interpretation of Findings from Research Questions 1 - 6

There are several potential explanations for the results of research questions one to six. First of all, with regard to test validity, instruction and assessment were aligned to a common set of 17 learning outcomes (see Table 3), so it was reasonable to assume that the mTMA measured what was taught. Although one of two previous studies using the TMA examined correlations between spatial ability and prior topographic map experience (Atit, et al., 2016), it did not study investigate the effects of cognitive load on either spatial ability or performance. A second issue concerned the structure of several TMA questions. Five questions required translating 3D landscape views to locations on a 2D topographic map and vice versa. Participant scores were lower on these questions, which could have skewed the results. Future analysis of participant scores on a per question basis can determine if these questions had a significant effect on mTMA scores. A third issue concerned the context of this experiment with Atit et al (2016) and Newcombe et al. (2015) using different instructional conditions and materials than were applied in this study and the potential effect it might have on the sensitivity of the TMA. A fourth consideration was sample size and representativeness of the target population. A sample size of 58 participants in Atit et al. (2016) was similar to the sample size of this study (n = 46), but was much smaller than Newcombe et al.’s (2015) study (n = 272). Since Atit et al’s study was more comparable in terms of examining high and low spatial ability groupings and TMA performance, it is plausible to assume that this study’s small sample size was somewhat representative of the larger population of undergraduate students with novice level topographic map skills. However, caution is urged in this assumption as effect sizes were small for all but one of the six research questions.

It is plausible that performance was lower and mental effort was higher on the mTMA than on the learning practice activity due to limited opportunities for embodied interaction and feedback when completing the posttest. Learning practice provided varying degrees of user-interactivity and feedback, depending on instructional condition. The TMA was akin to the non-AR condition since embodied interaction was limited to drawing. Yet non-significant differences in performance scores for the AR and non-AR conditions on the mTMA suggested that this might not have been a contributing factor. As shown in Chapter 2, embodied interaction and multisensory inputs can enhance performance by reducing overall cognitive load. When considering TMA test validity, use of different materials, measuring only performance and not cognitive load, a small sample size, and limited embodied interaction with a test instrument that was more akin to the non-AR condition than the AR condition, it is not clear whether or not the TMA structure or its questions may have affected either performance or mental effort.

The marginally-significant interaction of spatial ability and instructional condition, on cognitive load (p = .051) further highlighted that the mTMA was potentially sensitive to these dependent measures. Since the interaction of these three factors were statistically significant during the learning practice activity suggests that instructional condition can affect mental effort values from learning practice and mTMA posttest assessment since these values may be moderated by spatial ability. As instructional condition was a significant main effect the learning practice activity performance, it is also possible that it was the main contributor to this near-interaction of the TMA.

Summary of Findings from Ancillary Research Questions 1 - 6

Whereas the previous six research questions did not identify any significant main effects or interactions of instructional condition or spatial ability on either mTMA posttest performance or mTMA posttest mean mental effort, the same was not true when examining performance and mental effort during the learning practice activity. Instructional condition was a significant predictor of learning practice scores, with the AR condition scoring significantly higher (87.1%) than the non-AR condition (72.4%). The size of this effect was very large. There was no main effect of spatial ability on performance nor was there an interaction of instructional condition and spatial ability. This was unexpected given the large effect size of the instructional condition main effect.

Mean mental effort values were significantly lower for the AR condition (*low mental effort to rather low mental effort*) than for the non-AR condition (rather low to neither low nor high mental effort) during learner practice and there was a significant difference between high and low spatial ability participants mean mental effort ratings. There was a statistically significant interaction of spatial ability (low and high) and treatment condition (AR or non-AR) on learning practice activity mental effort values.

Multiple comparison testing revealed that the AR participants used less mental effort than the non-AR participants and scored higher, but this effect was strongly mediated by a participant’s spatial ability.

Interpretation of Findings from Ancillary Research Questions 1 - 6

Differences between low and high spatial ability group performance and mental effort can be accounted for in terms of the spatial ability as enhancer and spatial ability as compensator hypotheses. Performance and mental effort differences favoring the AR condition can be interpreted by multimodal and embodied cognition theories.

Spatial ability and cognitive load interpretations of results. Experimental results suggested that spatial ability was an indicator of relative performance and mental effort during the learning activity in which the ARS was more effective (lower mental effort and higher scores) than the traditional 2D paper map. There were medium to large effect sizes in spite of a small sample. However, performance on learning practice activity was not associated with combined effects of spatial ability and instructional condition. Spatial ability alone did not explain learning performance.

Spatial ability and instructional condition did interact to explain differences in mental effort during the learning practice activity. The spatial ability as enhancer hypothesis (Hays, 1996; Huk, 2007; Mayer & Sims, 1994) can account for the high spatial ability group’s ability to use less mental effort than the low spatial ability group regardless of instructional condition. High spatial ability participants leverage their prior experience and/or innate spatial cognitive ability which provides them with more developed schemas for incorporating new spatial information with their prior knowledge. With additional space in working memory, they are able to incorporate new information with less mental effort. Figure 9 illustrated that there is little difference in mental effort for the high spatial ability group in the AR and non-AR conditions so the effect of the spatial ability as enhancer hypothesis is not fully realized; one would expect high spatial ability participants' mental effort would be much lower than the mental effort of low spatial ability participants in the non-AR condition. The conclusion is that high spatial ability participants could use either a 2D topographic map or the ARS to learn topographic skills, as they had the highest scores in both treatment conditions and do so with the least mental effort.

The spatial ability as compensator hypothesis (Hays, 1996; Huk, 2007; Mayer & Sims, 1994; Quarles, Lampotang, Fischler, Fishwick, & Lok, 2008) can be used to interpret the mental effort of the low spatial ability group. Participants with low spatial ability benefitted from the tactile feedback from the ARS and its ability to promote embodied interaction. User-interactivity between the participant and the ARS increases germane cognitive load, but not at the expense of total mental effort. As Figure 7 shows, low spatial ability participants' mean mental effort was substantially higher than the high spatial ability participant group when using the 2D topographic map, but much lower and nearly equal to the mental effort of the high spatial ability group when using the ARS. This significant difference suggests that the ARS' 3D graphical and tangible features provide cognitive scaffolds that enable them to learn topographic map skills with less mental effort than if they were to use a 2D topographic map. Huk (2007) reported a significant interaction between spatial ability and 2D or 3D model (instructional condition) with different effects for high and low spatial ability. Only high spatial ability students benefitted, unlike in this study, where low spatial ability participants benefitted more than high spatial ability participants. The primary differences between Huk's findings and the anticipated results of this study is that there were no direct physical (tactile) interactions in his study and he speculated that the 3D interface was too complicated for low spatial ability students, such that they experienced cognitive overload.

Embodied interaction and multimodal interpretations of results. The higher degree of user interactivity the ARS afforded also provided additional sensory cues from tactile interaction. The ARS' real-time updates to the map surface in response to user interaction likely increased germane cognitive load, which is central to embodied interaction theory. The 2D topographic map was static; it did not change in response to user interactions. Drawing likely provided some degree of embodied cognition, yet it lacked the interactive response that was a key feature of the ARS. Since this study did not directly measure embodied interaction (it merely inferred it based on performance and mental effort) it is plausible that drawing topographic features on paper and marking them on a 2D paper map might result in cognitive overload among the low spatial ability group. Using the MEMS ratings, cognitive overload would likely equate to a mean value of seven (*high mental effort*) to nine (*extremely high mental effort*) across the assessment. To mitigate this potential limitation, the number of tasks requiring drawing of topographic features was limited to 10 of the 20 learning practice activity questions.

Tactile sensory cues provided an extra sensory input, although how much this extra sensory information contributed to learning could not be measured. Some combination of dimensionality and tactile interaction clearly had an effect on performance scores and mental effort during the learning practice activity and varied significantly depending on whether both constructs were present (AR condition) or absent (non-AR condition). It is proposed that these constructs increased germane cognitive load through user interaction with the map. The novelty of the ARS also promoted more user engagement and thus may have positively affected results. For example, 70% of AR condition participants reported that the ARS was more effective than either the multimedia instruction (20%) or posttest (10%) as a learning tool, whereas among non-AR condition participants, 47% preferred the multimedia instruction for learning topographic map reading skills over use of the 2D paper map (44%) but less than the posttest (9%).

Dimensionality interpretations of results. Ascertaining the learning effect of 2D versus 3D is more problematic because learner performance and mental effort during both the learning practice and the posttest is conflated with the simultaneous effects of spatial ability as enhancer and spatial ability as compensator effects as well as the effects of multimodal feedback and embodied interaction. The study lacked a means to determine the effects of dimensionality on performance and mental effort. This is further complicated by conflicting reports in the literature on the learning benefits of 2D versus 3D maps. As an example, the literature indicates that the degree of mental abstraction required to visualize a 2D map in 3D would be less for high spatial ability participants (Mayer & Moreno, 1998), but to what degree does each construct (multimodal feedback or embodied interaction) contribute to performance? Or is there an interactive effect of these two constructs? The significantly higher performance of low spatial ability participants in the AR condition with less mental effort suggested that some combination of factors could be attributed to the results. Therefore, one can only claim that the ARS afforded a combination of multimodal sensory feedback and embodied interaction, plus a more realistic 3D landscape where differences in elevation are readily evident, and do not have to be inferred when using a 2D topographic map. Lower performance scores and higher mental effort values in the non-AR condition is indicative that an absence of these AR affordances (embodied interaction, multimodal sensory inputs, and dimensionality differences) is detrimental to low spatial ability participants. These affordances are not required for high spatial ability participants.

Summary of Results: Comparing Posttest and Learning Practice Values

This study demonstrated the importance of hands-on use when novices learn how to use a topographic map, in spite of a small sample size. Results indicated that the mTMA posttest scores and mental effort ratings did not significantly differ. There were, however, significant differences in scores and mental effort during learning practice. Participants scored higher and with less mental effort during the learning practice activity than during the posttest assessment. The difference between learning practice scores (M = 5.30, SD = 1.26) and posttest scores (M = -.43, SD = 1.90) was statistically significant with a large Cohen's d effect size, $t(44) = 4.20$, $p < .001$, $d = 1.04$, 95% CI [.90, 4.0] The difference between mean mental effort values during learning practice (M = -.87, SD = .32) and mean mental effort posttest scores (M = -.43, SD = .37) was statistically significant with a small to medium Cohen's d effect size, $t(44) = -2.71$, $p = .010$, $d = .40$, 95% CI [-1.52, -.22]. Interactive practice was anticipated to improve performance with reduced mental effort, yet a lack of evidence of retention and transfer from multimedia instruction and learning practice to a posttest assessment was an unexpected result. Nevertheless, the overall findings indicate the learning benefits of interacting directly with an AR map; learning practice in this experimental context yielded significantly higher learner performance with significantly less mental effort. It is plausible that a combination of AR affordances (embodied interaction, multimodal sensory inputs, and dimensionality differences) promoted germane cognitive load, as evidenced by higher performance scores from interactive practice during the learning activity such that cognitive overload did not occur.

Addressing Threats to Validity

Clearly the small sample size tempers the findings and applicability of this study's findings. An *a post hoc* statistical power calculation using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) predicted that a .10 Cohen's *f* effect size with an alpha level of .05 and a sample size of 46 participants would yield a power value of .07, which was well below the recommended threshold of 80% statistical power to detect a significant different between experimental treatments. With a larger sample, research question 6 on the interaction of spatial ability and cognitive load on posttest performance (p = .051) may have been significant.

Creating a third experimental control group exposed to the instructional slideshow and then given the posttest only could establish if instruction without practice was more or less effective than instruction, practice, and then posttest assessment. This addition of a true control group to this study's research design was initially considered. This approach would offer an independent measure of the validity and reliability of the TMA as a performance assessment. Such a design would necessitate an unrealistically large sample size, however, and so was rejected.

This study cannot definitively claim that drawing was rated as ‘low’ interactivity compared to sculpting sand with the ARS. There is no current research that compares the relative degrees of germane load from these two types of actions. Lower results of the 2D topographic map group could, at least in part, be explained by extraneous cognitive load associated with drawing of topographic map features, particularly if the participant drew a feature incorrectly and then used it as a clue to identify similar features that would also be, by analogy, incorrect. Skulmowski et al., 2016 inferred that any physical interaction must be appropriate to the learning task; it was possible that drawing topographic features did not scaffold learning to a degree necessary to increase topographic map reading skill when compared to the ARS group.

Since this study did not directly measure embodied interaction (it merely inferred it based on performance and mental effort) it is plausible that drawing topographic features on paper and marking them on a 2D paper map might result in cognitive overload among the low spatial ability group. To mitigate this potential limitation, the number of tasks requiring drawing of topographic features was limited to 10 of the 20 learning practice activity questions.

Any individual conclusions about the relative effects of multimodal feedback, or embodied interaction, or dimensionality on the AR or non-AR conditions are not warranted. All one can surmise from this study’s results was that their cumulative effect was maximized in the AR condition and minimized in the non-AR condition; spatial ability was an important determining factor as to how much interaction and feedback was necessary to support learning.

Recommendations for Instructional Design of Multimodal Learning Environments

This experimental study has demonstrated the learning benefits of multimodal learning environments (MLE's) using AR. Highly interactive learning environments are characterized by a large degree of two-way communication, implying the learner receives information in response to an action (learner-instructional interface), which allows for knowledge construction as learner creates a schema to integrate the new information. User-interactivity promotes embodied interaction, which can increase germane cognitive load through active engagement. Instructional design of MLE's using spatial data are based on five key principles. (1) To maximize embodied interaction, two-way communication between the interface and the user is essential to foster learner engagement. The one-way communication of many digital learning environments does not leverage the benefits of embodied interaction. (2) Embodied interaction implies 'hands-on' activity to build robust schemas, so combine content delivery with practice opportunities. (3) Incorporate other sensory modalities whenever possible to provide extra non-verbal information. Auditory-tactile feedback interfaces show promise (Jeong & Gluck, 2003). (4) Identify students' spatial ability as part of the learner characteristics phase of the Kemp instructional design process (Morrison, Ross, Kalman, & Kemp, 2011) if students will be working with spatial information. Spatial ability testing can inform how much AR affordance (user-interaction, interface feedback, multidimensional representations, etc.) are required to support students with low spatial ability. Although students with high spatial ability can excel with AR affordances, they are not dependent on them. (5) Training and practice to improve spatial ability is beneficial to all students, particularly for STEM students and when dealing with spatial information. It is essential to identify as part of the instructional problem if a MLE's purpose is to provide spatial skills training as a stand-alone product or if a certain base-level of spatial ability is required to use the interface effectively. If the latter is the case, then pre-training of spatial skills required for the task at hand is recommended. Simply practicing a variety of spatial ability tests has been shown to have a positive learning effect (Peters et al., 1995).

To measure learning, eye-tracking using mobile headsets that are equipped with electroencephalography (EEG) sensors can inform instructional design of multimodal learning environments. They can measure leaner engagement through gaze tracking, cognitive load based on pupil size variations (Klingner, Kumar & Hanrahan, 2088: Klingner, 2010) and cognitive load from EEG patterns (Antonenko, Paas, Grabner, & van Gog, 2010a; Whelan, 2007). These systems are recommended for advanced analysis of instructional design products and for assessing interface usability, engagement, and learning benefits with optimal cognitive load.

Future Research

There are many possible avenues of research that emerged from this study. They are summarized below into five themes. The heading 'unanswered questions' addresses issues uncovered in the design and execution of this study and offers recommendations for additional research.

Eye-tracking. Mobile headsets were worn by participants during the multimedia instruction and learning practice phases of the experiment. Although not included in this study, eye fixation, saccades, and pupil diameter changes are direct measures of cognitive load. A possible correlation between pupil size changes and

MEMS ratings would further validate the MEMS scale as a simple, yet accurate measure of cognitive load. Eye fixation and saccades can be mapped onto the computer display during multimedia instruction and on the ARS surface to ascertain where participants are looking, how long they fix their gaze on certain areas of the map or computer display, and if their gaze tracks or follows their hand movements. Eye tracking can also be used to identify areas of attention or non-attention in interface design.

Measurement. Future efforts at quantifying and thus validating measures of learning performance, mental effort during instruction, and mental effort during testing are needed to inform instructional design practices. The instructional condition efficiency (Paas & Van Merriënboer, 1993) and multidimensional condition efficiency approaches (Tuovinen & Paas, 2004) can quantify these values and so is worthy of additional research. This study did not examine individual scores and associated mental effort values on a per-question basis to identify questions that require more abstract and complex mental translation from 2D to 3D perspectives and vice-versa—a possible learning dimensionality effect? Designing questions and tasks based on Gersmehl & Gersmehl's (2006) spatial ability hierarchy could determine if a learning dimensionality construct is valid. Future research is also needed to separate learning effects based on dimension, multimodality, and embodied interaction and how these constructs could be operationalized as measurable independent variables.

Spatial ability. Future analysis should differentiate results based on gender; and use a longitudinal study to measure changes in spatial ability over time associated with using the AR vs. non-AR devices. It would also be worthwhile to determine if these spatial skill transfer to different situations or when using different interfaces.

Instructional design. Findings on positive learning effects with reduced mental effort can inform design of spatial-based augmented reality TUI's and specific teaching and assessment strategies to improve spatial thinking The latter will have broader impacts on enhancing spatial skills necessary for success in many STEM fields such as, but not limited to, geoscience. The intuitive nature of the ARS and its appeal to children (Xie, Antle & Motamedi, 2008) suggests that it can and should be used to supplement map reading instruction, in K-12 Social Studies curricula, with the added potential to develop students' spatial abilities.

Unanswered questions. Additional research is needed to differentiate between intrinsic, extraneous, and germane load and how to measure each. For example, item response theory which differentiates between mental effort mental load shows promise (Krella, 2015), as do recent advances in EEG analysis of cognitive data (Antonenko, et al., 2010a; Whelan, 2007). The debate over whether 2D or 3D spatial maps are more effective is still an unanswered question. It would appear that a user's spatial ability is a mediating factor. Further inquiry is needed to measure how multimodal feedback and embodied interaction can support the mental abstractions of thinking in both 2D and 3D as well as finding a method to independently measure the learning differences between 2D and 3D topographic maps. Otherwise, the effects of multimodal feedback and embodied interaction can confound the question of dimensionality effects.

Another question is whether dual coding theory (Clark & Paivio, 1991; Mayer & Sims, 1994) can accommodate more than visual and verbal channels. Is other sensory data, such as touch, encoded on one of these channels or are there other, yet unknown, cognitive pathways? Moreno & Mayer's (2007) work on MLE's suggest a modification to dual coding theory as it has been adapted to multimedia learning environments. They list five sensory memory inputs: auditory, visual, tactile, olfactory, and gustatory (Figure 1, p. 314) yet discuss only the traditional verbal and visual channels. How the remaining three senses could be incorporated into the multimedia learning/dual coding theory models has not yet been addressed.

Conclusion

This study has provided empirical evidence on the learning effectiveness of the ARS compared to traditional topographic map instruction. Augmented reality affordances (embodied interaction, multimodal sensory inputs, and dimensionality differences) accounted for performance and mental effort differences during learning practice when such affordances were either available (ARS instructional condition) or absent (2D topographic map instructional condition). Lower total cognitive load in the ARS condition allowed for more information processing with spatial ability as a significant mediator. Participants with high spatial ability did not require AR affordances as their performance and mental effort did not vary significantly between the two instructional conditions. This group scored higher and with less mental effort in both conditions than did the low spatial ability participants. This group benefitted from the AR affordances as their performance was higher and mental effort lower only when they were available in the ARS treatment condition.

A lack of a significant learning difference between the two instructional condition groups could be also explained by a lack of AR affordances. This contention was further supported by significant differences between higher learning practice and lower mental effort values compared to posttest results. The overall findings of this study suggest that using AR learning to teach topographic map skills to novices in this experimental context is more effective than more traditional non-AR instruction and assessment, as AR learning can promote germane cognitive load through active engagement in hands-on learning activities.

These findings will contribute to multimodal learning, embodied interaction, and dimensionality theories to improve spatial ability. The ultimate goal of this line of research is to emphasize the importance of spatial thinking skills and to promote their development in all learners through design and testing of multimodal learning environments that include AR affordances.

References

Amini, F., Rufiange, S., Hossain, Z., Ventura, Q., Irani, P. & McGuffin, M. J. (2015). The impact of interactivity on comprehending 2D and 3D visualizations of movement data. *Visualization and Computer Graphics, IEEE Transactions on, 21*(1), 122–135. IEEE.

Anthamatten, P., & Ziegler, S. S. (2006). Teaching geography with 3-D visualization technology. *Journal of Geography*, *105*(6), 231–237. http://doi.org/10.1080/00221340608978692

Antle, A. N., & Wang, S. (2013). Comparing motor-cognitive strategies for spatial problem solving with tangible and multi-touch interfaces. *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction - TEI '13*, 65. http://doi.org/10.1145/2460625.2460635

Antle, A. N., & Wang, S. (2013). Comparing motor-cognitive strategies for spatial problem solving with tangible and multi-touch interfaces. *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction - TEI '13*, 65. http://doi.org/10.1145/2460625.2460635

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010a). Using Electroencephalography to Measure Cognitive Load. *Educational Psychology Review*, *22*(4), 425–438. http://doi.org/10.1007/s10648-010-9130-y

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010b). Using Electroencephalography to Measure Cognitive Load. *Educational Psychology Review*. http://doi.org/10.1007/s10648-010-9130-y

Atit, K., Weisberg, S. M., Newcombe, N. S., & Shipley, T. F. (2016). Learning to interpret topographic maps: Understanding layered spatial information. *Cogntive Research: Principles and Implications*, *1*(1), 50. http://doi.org/10.1186/s41235-016-0002-y

Atit, K., Weisberg, S. M., Newcombe, N. S., & Shipley, T. F. (2016). Learning to interpret topographic maps: Understanding layered spatial information. *Cognitive Research: Principles and Implications*, *1*(1), 2. http://doi.org/10.1186/s41235-016-0002-y

Ayres, P., & Paas, F. (2012). Cognitive load theory: New directions and challenges. *Applied Cognitive Psychology*, *26*(6), 827–832. http://doi.org/10.1002/acp.2882

Bednarz, R. S., & Lee, J. (2011). The components of spatial thinking: Empirical evidence. *Procedia - Social and Behavioral Sciences*, *21*, 103–107. http://doi.org/10.1016/j.sbspro.2011.07.048

Bimber, O., & Raskar, R. (2005). *Spatial Augmented Reality Merging Real and Virtual Worlds*. *SpatialAR.com* (Vol. 6). http://doi.org/10.1260/147807708784640126

Bracht, G. H., & Glass, G. V. (1968). The External Validity of Experiments. *American Educational Research Journal*, *5*(4), 437–474. http://doi.org/10.3102/00028312005004437

Brünken, R., & Plass, J. L. (2003). Direct Measurement of Cognitive Load in Multimedia Learning, *38*(1), 53–61.

Bunch, R. L., & Lloyd, R. E. (2006). The cognitive load of geographic information. *Professional Geographer*, *58*(2), 209–220. http://doi.org/10.1111/j.1467-9272.2006.00527.x

Burke, J. L., Prewett, M. S., Gray, A. a., Yang, L., Stilson, F. R. B., Coovert, M. D., … Redden, E. (2006). Comparing the effects of visual-auditory and visual-tactile feedback on user performance: A Meta-analysis. *Proceedings of the 8th International Conference on Multimodal Interfaces - ICMI '06*, 108. http://doi.org/10.1145/1180995.1181017

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experiment al designs for research*.

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, *8*(4), 293–332. http://doi.org/10.1207/s1532690xci0804_2

Chang, J. S., Yeboah, G., Doucette, A., Clifton, P., Nitsche, M., Welsh, T., & Mazalek, A. (2017). TASC : Combining Virtual Reality with Tangible and Embodied Interactions to Support Spatial Cognition. http://doi.org/10.1145/3064663.3064675

Christopherson, R. W. (2010). *Elemental Geosystems* (6th ed.). Upper Saddle River, NJ: Pearson.

Clark, D., Reynolds, S., Lemanowski, V., Stiles, T., Yasar, S., Proctor, S., … Corkins, J. (2008). University Students' Conceptualization and Interpretation of Topographic Maps. *International Journal of Science Education*, *30*(January 2015), 377–408. http://doi.org/10.1080/09500690701191433

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

Clark, R. C., & Mayer, R. E. (2011). *E-Learning and the Science of Instruction* (3rd ed.). San Franscisco, CA: Pfeiffer.

Clifton, P. G., Chang, J. S.-K., Yeboah, G., Doucette, A., Chandrasekharan, S., Nitsche, M., … Mazalek, A. (2016). Design of embodied interfaces for engaging spatial cognition. *Cognitive Research: Principles and Implications*, *1*(1), 24. http://doi.org/10.1186/s41235-016-0032-5

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

Collins, L. (2014). Tradition or Technology?: the Impact of Paper Versus Digital Map Technology on Students' Spatial Thinking Skill Acquisition. *Theses and Dissertations*. Retrieved from http://scholarcommons.sc.edu/etd/2800

Debue, N., & van de Leemput, C. (2014). What does germane load mean? An empirical contribution to the cognitive load theory. *Frontiers in Psychology*, *5*(SEP), 1–12. http://doi.org/10.3389/fpsyg.2014.01099

Dorling, D., & Fairbairn, D. (1997). *Mapping: Ways of Representing the World*. Dorchester: Prentice Hall.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Manual for kit of factor-related cognitive tests*. Princetpn, NJ: Educational Testing Service.

Eley, M. G. (1991). Selective encoding in the interpretation of topographic maps. *Applied Cognitive Psychology*, *5*, 403–422.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–91. http://doi.org/10.3758/BF03193146

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychol Sci*, *18*(10), 850–855. http://doi.org/10.1111/j.1467-9280.2007.01990.x

Gerjets, P., & Scheiter, K. (2010). Goal Configurations and Processing Strategies as Moderators between Instructional Design and Cognitive Load: Evidence from Hypertext-Based Instruction. *Educational Psychologist*, (38), 43–52. http://doi.org/10.1207/S15326985EP3801

Gersmehl, P. J., & Gersmehl, C. A. (2006). Wanted: A concise list of neurologically defensible and assessable spatial thinking skills. *Research in Geographic Education*, *8*(1), 5–38.

Gilhooly, K. J., Wood, M., Kinnear, P. R., & Green, C. (1988). Skill in Map Reading and Memory for Maps. *The Quarterly Journal of Experimental Psychology Section A*, *40*(1), 87–107. Retrieved from http://dx.doi.org/10.1080/14640748808402284

Gog, T. van, Kirschner, F., Kester, L., & Paas, F. (2012). Timing and Frequency of mental effort measurement: Evidence in favor of repeated measures. *Applied Cognitive Psychology*, *26*(6), 833–839.

Goodchild, M. F. (2011). Spatial thinking and the GIS user interface. *Procedia - Social and Behavioral Sciences*, *21*, 3–9. http://doi.org/10.1016/j.sbspro.2011.07.002

Harrower, M. (2007). The Cognitive Limits of Animated Maps. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *42*(4), 349–357. http://doi.org/10.3138/carto.42.4.349

Hays, T. A. (1996a). Spatial abilities and the effect of computer animation on short-term and long-term comprehension. *Journal of Educational Computing Research*, *14*(2), 139–155. http://doi.org/10.1146/annurev.eco

Hays, T. A. (1996b). Spatial Abilities and the Effects of Computer Animation on Short-Term and Long-Term Comprehension. *Journal of Educational Computing Research*, *14*(2), 139–155. Retrieved from https://doi.org/10.2190/60Y9-BQG9-80HX-UEML

Hecht, H., & Proffitt, D. R. (1995). THE PRICE OF EXPERTISE: Effects of Experience on the Water???Level Task. *Psychological Science*, *6*(2), 90–95. http://doi.org/10.1111/j.1467-9280.1995.tb00312.x

Hegarty, M. (2013). Cognition, Metacognition, and the Design of Maps. *Current Directions in Psychological Science*, *22*(1), 3–9. http://doi.org/10.1177/0963721412469395

Hegarty, M., Kozhevnikov, M., & Waller, D. (2004). Perspective taking / spatial orientation test. *Intelligence*, *32*(January), 175–191.

Huk, T. (2007). “Who benefits from learning with 3D models? The case of spatial ability”: Erratum. *Journal of Computer Assisted Learning*, *23*(2), 169. http://doi.org/10.1111/j.1365-2729.2007.00231.x

Ishikawa, T., & Kastens, K. A. (2005). Why some students have trouble with maps and other spatial representations. *Journal of Geoscience Education*, *53*(2), 184–197. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-17744370094&partnerID=40&md5=fd3c7c97779ca443818f634295e8311e

Jeong, W., & Gluck, M. (2003). Multimodal geographic information systems: Adding haptic and auditory display. *Journal of the American Society for Information Science and Technology*, *54*(3), 229–242. http://doi.org/10.1002/asi.10202

Johnson, A., Leigh, J., Morin, P., & Van Keken, P. (2006). GeoWall: Stereoscopic visualization for geoscience research and education. *IEEE Computer Graphics and Applications*, *26*(6), 10–14. http://doi.org/10.1109/MCG.2006.127

Johnson, B., & Christensen, L. (2008). *Educational Research* (Vol. 3rd). Thousand Oaks, CA: Sage Publications.

Kalyuga, S. (2007). Enhancing instructional efficiency of interactive e-learning environments: A cognitive load perspective. *Educational Psychology Review*, *19*(3), 387–399. http://doi.org/10.1007/s10648-007-9051-6

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, *92*(1), 126–136. http://doi.org/10.1037/0022-0663.92.1.126

Keehner, M., Montello, D. R., Hegarty, M., & Cohen, C. A. (2004). Effects of interactivity and spatial ability on the comprehension of spatial relations in a 3D computer visualization. *26Th Annual Conference of Th Cognitive Science Society*, *1*(1978), 1576.

Kim, M. J., & Maher, M. Lou. (2008). The impact of tangible user interfaces on spatial cognition during collaborative design. *Design Studies*, *29*(3), 222–253. http://doi.org/10.1016/j.destud.2007.12.006

Kim, M., & Maher, M. L. O. (2005). Comparison of designers using a tangible user interface & graphical user interface and impact on spatial cognition. *Proc Human Behaviour in Design 5*. Retrieved from https://books.google.com/books?hl=en&lr=&id=ZArouQGtlPcC&oi=fnd&pg=PA81&dq=tangible+user+interface+vs+graphical&ots=hPh3oprs_4&sig=U9u5AjFs3hDkYjMp4bn0QRDtvLk

Kimura, D. (2000). *Sex and Cognition*. Boston, MA: MIT Press.

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition*, *29*(5), 745–756. http://doi.org/10.3758/BF03200477

Krella, M. (2015). Evaluating an instrument to measure mental load and mental effort using Item Response Theory, *2015*(3), 1–6.

Lee, E. A.-L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers & Education*, *79*, 49–58.

Lee, J., & Bednarz, R. (2012). Components of Spatial Thinking: Evidence from a Spatial Thinking Ability Test. *Journal of Geography*, *111*(1), 15–26. http://doi.org/10.1080/00221341.2011.583262

Levin, H. L. (1986). *Contemporary Physical Geology* (2nd ed.). New York, NY: Saunders College Publishing.

Liben, L. S., & Titus, S. J. (2012). The importance of spatial thinking for geoscience education: Insights from the crossroads of geoscience and cognitive science. *Geological Society of America Special Papers*, *486*(10), 51–70. http://doi.org/10.1130/2012.2486(10).

Linn, M. C., & Petersen, A. C. (1985). Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis. *Child Development*, *56*(6), 1479–1498. http://doi.org/10.2307/1130467

Linn, M., & Petersen, A. (1985). Emergence and Characterization of Sex Differences in Spatial Ability: a Meta-Analysis. *Child Development*, *56*(6), 1479–1498.

Lounsbury, J. F., & Aldrich, F. T. (1986). *Introduction to Geographic Field Methods and Techniques* (2nd ed.). Columbus, OH: Charles E. Merrill Publishing.

Maxwell, S. E., & Delaney, H. D. (2004). Research methods and experimantal design: Model Comparison Pespective. *Designing and Analyzing Data*.

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. *The Cambridge Handbook of Multimedia Learning*, 31–48. http://doi.org/10.1207/s15326985ep4102_2

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, *20*(2), 167–171. http://doi.org/10.1016/j.learninstruc.2009.02.012

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*(2), 312–320. http://doi.org/10.1037/0022-0663.90.2.312

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, *86*(3), 389–401. http://doi.org/10.1037/0022-0663.86.3.389

Meyers, L. S., Gamst, G., & Guarino, A. J. (2013). *Applied Multivariate Research: Design and Implementation*. (Sage Publications, Ed.) (2nd ed.). Thousand Oaks, CA.

Minogue, J., & Jones, M. G. (2006). Haptics in Education: Exploring an Untapped Sensory Modality. *Review of Educational Research*, *76*(3), 317–348. http://doi.org/10.3102/00346543076003317

Modjeska, D., & Chignell, M. (2003). Individual differences in exploration using desktop VR. *Journal of the American Society for Information Science and Technology*, *54*(3), 216–228. http://doi.org/10.1002/asi.10197

Mohler, J. L. (2008). A review of spatial ability research. *Engineering Design Graphics Journal*, *72*(2), 19–30.

Moreno, R., & Mayer, R. (2007). Interactive Multimodal Learning Environments. *Educational Psychology Review*, *19*(3), 309–326. http://doi.org/10.1007/s10648-007-9047-2

Morrison, G. R., Ross, S. M., Kalman, H. K., & Kemp, J. E. (2011). *Designing Effective Instruction* (6th ed.). Hoboken, NJ: John Wiley & Sons.

Muehrcke, P. C. (1978). *Map Use: Reading, Analysis, and Interpretation*. Madison, WI: JP Publications.

Myers, J. L., Well, A. D., & Lorch Jr., R. F. (2010). *Research Design and Statistical Analysis* (3rd ed.). New York, NY: Routledge.

Nam, C. S., Li, Y., Yamaguchi, T., & Smith-Jackson, T. L. (2012). Haptic User Interfaces for the Visually Impaired: Implications for Haptically Enhanced Science Learning Systems. *International Journal of Human-Computer Interaction*, *28*(12), 784–798. http://doi.org/10.1080/10447318.2012.661357

Newcombe, N. S., Weisberg, S. M., Atit, K., Jacovina, M., Ormond, C. J., & Shipley, T. F. (2015). The lay of the land: Sensing and representing topography. In *The Baltic Yearbook of Cognition, Logic and Communication* (Vol. 10, pp. 1–57). http://doi.org/http://dx.doi.org/10.4148/1994-3676.1099

Niedomysl, T., Elldér, E., Larsson, A., Thelin, M., & Jansund, B. (2013). Learning Benefits of Using 2D Versus 3D Maps: Evidence from a Randomized Controlled Experiment. *Journal of Geography, 112*(3), 87–96. http://doi.org/10.1080/00221341.2012.709876

Nollenburg, M. (2007). Geographic Visualization. *Human-Centered Visualization Environments*, 257–294. http://doi.org/10.1007/978-3-540-71949-6_6

Ormand, C. J., Shipley, T. F., Tikoff, J., Harwood, C. L., Atit, K., & Boone, A. P. (2014). Evaluating Geoscience Students' Spatial Thinking Skills in a Multi-Institutional Classroom Study. *Journal of Geoscience Education*, *62*(1), 146–154. http://doi.org//10.5408/13-027.1

Paas, F. G. W. C., Merriënboer, J. J. G. Van, & Adam, J. J. (1994). Measurement of Cognitive Load in Instructional Research. *Perceptual and Motor Skills*, *79*(1), 419–430. http://doi.org/10.2466/pms.1994.79.1.419

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1993). The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *35*(4), 737–743. http://doi.org/10.1177/001872089303500412

Paas, F. G. W. C, Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, *38*(1), 1–4. http://doi.org/10.1207/S15326985EP3801_1

Parneet Kaur & Sheveta Vashisht. (2013). Analyzing EEG data for cognitive behaviour using EEGLAB through data mining. *International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR)*, *3*(2), 19–26. Retrieved from http://www.tjprc.org/view_archives.php?year=2013&jtype=2&id=14&details=archives

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). Peters (1995) a Redrawn Vandenberg and Kuse Mental Rotations Test. Different Versions and Factors That Affect Performance.Pdf. *Brain and Cognition*, *28*, 39–58.

Petersen, J. F., Sack, D., & Gabler, R. E. (2011). *Fundamentals of Physical Geography* (Instructor). Belmont, CA: Brooks/Cole.

Phillips, R. J., Lucia, A., & Skelton, N. (1975). Objective tests of the legibility of relief maps. *The Cartographic Journal*, *12*(1), 39–46.

Phillips, R. J., & Noyes, L. (1982a). An Investigation of Visual Clutter in the Topographic Base of a Geological Map. *The Cartographic Journal*. http://doi.org/10.1179/000870482787073225

Phillips, R. J., & Noyes, L. (1982b). An investigation of visual clutter in the topographic bases of a geological map. *The Cartographic Journal*, *19*(2), 122–132.

Piburn, M. D., Reynolds, S. J., Leedy, D. E., & Mcauliffe, C. M. (2002). The Hidden Earth : Visualization of Geologic Features and their Subsurface Geometry.

Piburn, M. D., Reynolds, S. J., Leedy, D. E., McAuliffe, C. M., Birk, J. P., & Johnson, J. K. (2002). The hidden earth: Visualization of geologic features and their subsurface geometry. *In Proceedings of the Annual Meeting of the National Association for Research in Science Teaching*, (JANUARY). Retrieved from citeulike-article-id:11979729

Pick, H. L., & Thompson, W. B. (1991). *Topographic map reading*. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA238026

Potash, L. M., Farrell, J. P., & Jeffrey, T. S. (1978). A Technique for Assessing Map Relief Legibility. *Cartographic Journal, The*, *15*(1), 28–35. http://doi.org/10.1179/000870478787338866

Price, S., & Jewitt, C. (2013). A multimodal approach to examining “embodiment” in tangible learning environments. *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction - TEI ’13*, 43–50. http://doi.org/10.1145/2460625.2460632

Quarles, J., Lampotang, S., Fischler, I., Fishwick, P., & Lok, B. (2008a). A mixed reality approach for merging abstract and concrete knowledge. *Proceedings - IEEE Virtual Reality*, (March), 27–34. http://doi.org/10.1109/VR.2008.4480746

Quarles, J., Lampotang, S., Fischler, I., Fishwick, P., & Lok, B. (2008b). Tangible User Interfaces compensate for low spatial cognition. *3DUI - IEEE Symposium on 3D User Interfaces 2008*, (June 2017), 11–18. http://doi.org/10.1109/3DUI.2008.4476585

Reed, S., Kreylos, O., Hsi, S., Kellogg, L., Schladow, G., Yikilmaz, M. B., … Sato, E. (2014). Shaping Watersheds Exhibit: An Interactive, Augmented Reality Sandbox for Advancing Earth Science Education. In *American Geophysical Union (AGU) Fall Meeting* (Vol. Abstract n). Retrieved from http://adsabs.harvard.edu/abs/2014AGUFMED34A..01R

Savage, D. M., Wiebe, E. N., & Devine, H. a. (2004). Performance of 2D versus 3D Topographic Representations for Different Task Types. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(16), 1793–1797. http://doi.org/10.1177/154193120404801601

Schnotz, W. (2002). Commentary: Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, *14*(1), 101–120. http://doi.org/10.1023/A:1013136727916

Selby, M. J. (1985). *Earth's Changing Surface: An Introduction to Geomorphology*. Oxford, UK: Clarendon Press.

Shobesberger, D. (2007). *Evaluating the effectiveness of 2D vs. 3D trailhead maps: A study conducted at Zion National Park, Utah*. *Report for the National Park Service*.

Skulmowski, A., Pradel, S., Kühnert, T., Brunnett, G., & Rey, D. (2016). Embodied learning using a tangible user interface: The effects of haptic perception and selective pointing on a spatial learning task. *Computers & Education*, *92–93*, 64–75. http://doi.org/10.1016/j.compedu.2015.10.011

Smallman, H. S., John, M. S., Oonk, H. M., & Cowen, M. B. (2001). Information availability in 2D and 3D displays. *IEEE Computer Graphics and Applications*, *21*(5), 51–57. http://doi.org/10.1109/38.946631

Song, H. S., Pusic, M., Nick, M. W., Sarpel, U., Plass, J. L., & Kalet, A. L. (2014). The cognitive impact of interactive design features for learning complex materials in medical education. *Computers & Education*, *71*, 198–205. http://doi.org/10.1016/j.compedu.2013.09.017

St John, M., Cowen, M. B., Smallman, H. S., & Oonk, H. M. (2001). The use of 2D and 3D displays for shape-understanding versus relative-position tasks. *Human Factors*, *43*(1), 79–98. http://doi.org/10.1518/001872001775992534

Stieff, M., & Uttal, D. (2015). How Much Can Spatial Training Improve STEM Achievement? *Educational Psychology Review*. http://doi.org/10.1007/s10648-015-9304-8

Strahler, A. N. (1987). *Introduction to Physical Geography*. Hoboken, NJ: Wiley.

Sweller, J. (1988a). Cognitive Load During Problem Solving : Effects on Learning, *285*, 257–285.

Sweller, J. (1988b). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. http://doi.org/10.1207/s15516709cog1202_4

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296. http://doi.org/10.1023/A:1022193728205

Taylor, H. A., & Hutton, A. (2013). Think3d!: Training Spatial Thinking Fundamental to STEM Education. *Cognition and Instruction*, *31*(4). http://doi.org/10.1080/07370008.2013.828727

Taylor, S., & Asmundson, G. J. G. (2007). Internal and external validity in clinical research. In D. McKay (Ed.), *Handbook of Research Methods in Abnormal and Clinical Psychology* (p. 586). Los Angeles, CA: Sage Publications.

The National Academies Press. (2006). *Learning to Think Spatially*. The National Academies Press. Retrieved from http://books.google.com/books?id=TJsmcnvrWVMC&pgis=1

Titus, S., & Horsman, E. (2009). Characterizing and Improving Spatial Visualization Skills. *Journal of Geoscience Education*, *57*(4), 242–254. http://doi.org/10.5408/1.3559671

Tuovinen, J. E., & Paas, F. (2004). Exploring Multidimensional Approaches to the Efficiency of Instructional Conditions. *Instructional Science*, *32*(1), 133–152. http://doi.org/10.1023/b:truc.0000021813.24669.62

Ullmer, B., & Ishii, H. (2001). Emerging frameworks for tangible user interfaces. *IBM Systems Journal*, *39*(3.4), 915–931. http://doi.org/10.1147/sj.393.0915

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2). http://doi.org/10.1037/a0028446

Van Merriënboer, J. ., Schuurman, J. ., de Croock, M. B. ., & Paas, F. G. W. . (2002). Redirecting learners' attention during training: effects on cognitive load, transfer test performance and training efficiency. *Learning and Instruction*, *12*(1), 11–37. http://doi.org/10.1016/S0959-4752(01)00020-2

Velez, M. C., Silver, D., & Tremaine, M. (2005). Understanding visualization through spatial ability differences. *Proceedings of the IEEE Visualization Conference*, 65. http://doi.org/10.1109/VIS.2005.108

Voyer, D., & Bryden, M. P. (1990). Gender, Level of Spatial Ability, and Lateralization of Mental Rotation. *Brain and Cognition*, *13*(1), 18–29. http://doi.org/10.1016/0278-2626(90)90037-O

Wakabayashi, Y., & Ishikawa, T. (2011). Spatial thinking in geographic information science: A review of past studies and prospects for the future. *Procedia - Social and Behavioral Sciences*, *21*, 304–313. http://doi.org/10.1016/j.sbspro.2011.07.031

Weisberg, S. M., Nardi, D., Newcombe, N. S., & Shipley, T. F. (2014). Up by upwest: Is slope like north? *The Quarterly Journal of Experimental Psychology*, *67*(10), 1959–1976. http://doi.org/10.1080/17470218.2014.880122

Weisberg, S. M., Newcombe, N. S., & Shipley, T. F. (n.d.). *What Predicts Understanding of Topographic Maps ? What predicts topo map comprehension ?*

Whelan, R. (2007). Neuroimaging of Cognitive Load in Instructional Multimedia. *Educational Research Review*, *2*(1), 1–12.

Wiebe, E. N., Roberts, E., & Behrend, T. S. (2010). An examination of two mental workload measurement approaches to understanding multimedia learning. *Computers in Human Behavior*, *26*(3), 474–481. http://doi.org/10.1016/j.chb.2009.12.006

Woods, T. L., Reed, S., Hsi, S., Woods, J. A., & Woods, M. R. (2016). Pilot Study Using the Augmented Reality Sandbox to Teach Topographic Maps and Surficial Processes in Introductory Geology Labs. *Journal of Geoscience Education*, *64*(3), 199–214. http://doi.org/10.5408/15-135.1

Xie, L., Antle, A. N., & Motamedi, N. (2008). Are tangibles more fun? Comparing Children's Enjoyment and Engagement Using Physical, Graphical and Tangible User Interfaces. *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction - TEI '08*, 191–198. http://doi.org/10.1145/1347390.1347433

Yun, L., Yufen, C., & Yingjie, W. (2004). Cognition Theory-Based Research on Adaptive User Interface for Geo-Visualization System, (June 7–9, 2004). In *Proceedings of the 12th International Conference on Geoinformatics* (pp. 424–430). University of Gävle, Sweden.

Appendix A

Cover Letter

Research Study: Comparison of Topographic Map Interfaces to Affect Spatial Thinking

Dear Prospective Research Participant,
We are asking for your help by volunteering to be participant in a research project investigating the use of a variety of interfaces to affect spatial thinking. This research is significant for understanding how the sense of touch may relate to understanding two and three dimensional images. This research will form part of a dissertation for the PhD in Instructional at ISU. Your participation can help to improve spatial thinking skills among college students, which is an important, yet neglected, part of academic training.

The project involves random assignment of participants to use three different topographical map interfaces. No prior knowledge or skill is required. You will perform a series of spatial thinking activities, and receive training on how to use the map interfaces. Your learning will then be measured with a knowledge performance and a mental effort activity assessment. Participation is expected to take approximately one hour. From each trial session, one participant will be randomly selected to participate in a subsequent extended experiment to replicate a topographic map using one of the other map interfaces. Participation in this second experiment optional.

Your decision to participate in this study is completely voluntary. You are not required to participate and declining to participate in no way jeopardizes your academic standing. All responses will be completely anonymous; your results will be associated with an identifying number, not your name, so it will not be possible to match you with your data in any way. Only aggregate (non-individual) results will be reported in a subsequent dissertation and in future academic journal articles.

There are no known economic, legal, physical, psychological, or social risks to participants in either immediate or long-range outcomes. If you agree to participate in one or both experiments, you may choose not to answer any given question, and you may withdraw your consent and discontinue your participation at any time.

An informed consent form is attached. You may complete it and bring it with you to the laboratory (Geosciences Computer lab, room 216) or complete a copy at the time you have selected to participate. If you have any questions about the study, please ask them before you begin.

If you have any questions or concerns about the nature of this research, please contact the primary investigator, Rick Richardson, Department of Organizational Learning & Performance at richric3@isu.edu or by phone (208-705-2019).

If you have any questions about your treatment as a human subject in this study or any questions about your rights as a research participant, you may contact the Idaho State University Human Subjects Committee at (208)-282-2179.

Thank you for considering to help in this valuable research.

Sincerely,

Instructional Design Ph.D. candidate
Department of Organizational Learning & Performance, ISU College of Education

Appendix B

Notice of Informed Consent Waiver

Dear Participant,

We are requesting your participation in an experimental survey of different interfaces for reading topographic maps. No prior knowledge of geography or map reading skill is required. The purpose of this experiment is to explore strategies for enhancing spatial thinking skills among college students. As a volunteer, you will be randomly assigned to one of two experimental groups. After completing this consent waiver, a spatial thinking measurement, and a demographic questionnaire, you will receive training in how to use one of two different map interfaces while wearing an eye tracking headset. Your learning will be measured with a series of learning tasks and a common map skills assessment. You may be randomly selected for a second exercise with the Augmented Reality Sandbox. Any personally-identifiable data will be deleted upon completion of the study. **Participation is completely voluntary and you may withdraw at any time. There is no consequence for withdrawing from the study or electing to not participate.**

For further information regarding this research study, please contact the Principal Researcher, Rick Richardson at (208) 705-2019, email: richric3@isu.edu or Dr. Dotty Sammons (Principal Advisor) at email: sammdott@isu.edu.

If you have any questions about your rights as a research participant you may contact the Idaho State University Human Subjects Committee at (208)-282-2179.

On the following page is a list of experimental trial times. Please select four different dates and times that you are certain that you can attend.

Thank you in advance for your cooperation and support of scientific inquiry at ISU that will support student learning.

Experiment 1: Please indicate your agreement to participate by signing below.

I am 18 years or older and have read and understood this consent form and agree to participate.

Signature: ____

Name: ____ (Please Print)

Date: ____

SESSION ATTENDED: Date_______________
Time________________

TX _______

REGISTERED IN WHICH COURSE? ________________________

Trial Sessions (sign up for a maximum of 4). Each session has this many time slots

Oct 21	Nov 1	Nov 19
Oct 24	Nov 2	Nov 20
Oct 25	Nov 3	Reading Break
Oct 26	Nov 4	Dec 1
Oct 27	Nov 7	Dec 2
Oct 28	Nov 8	Dec 3
Oct 31	Nov 9	Dec 6
	Nov 12	Dec 7
	Nov 13	Dec 8
	Nov 14	Dec 9
	Nov 15	Dec 10
	Nov 16	

8-830am
9-930am
10-1030am
11-1130am
12-1230pm
1-130pm
2-230pm
3-330pm
4-430pm
5-530pm
6-630pm
7-730pm

Appendix C

Demographic Questionnaire

Spatial thinking using topographic maps

Participant Demographic Questionnaire

Please answer each question honestly. All responses are confidential and your identity will be kept anonymous.

* Required

What is your participant id #? *

What is your current academic major?
Check only one option

- Physical Sciences (Chemistry, Physics, Biology)
- Education
- Sports Science
- Computer Science
- Mathematics
- Engineering
- Arts & Letters
- Health Sciences
- Fine Arts
- I have not yet decided on a major

What is your academic standing?
Check only one option

- Freshman
- Sophomore
- Junior
- Senior

- Mature/Returning Student
- Graduate Student

Are you currently enrolled in one of the following course(s)?
Check only one option

- Geology/geoscience
- Education
- I am enrolled in one or more Education AND one or more Geology/geoscience courses
- Graphic design or Visual arts
- Mathematics
- Computer science
- Engineering
- Sport Science/Physical Education
- Mathematics

What is your gender?

- Female
- Male

Topographic maps use contour lines to show elevations and shapes of land forms. How would you rate your knowledge-experience with using topographic maps?
Check only one option

	1	2	3	4	5	
I have never used a topographic map	○	○	○	○	○	I use topographic maps daily

If you have used topographic maps in the past, in what situations did you use them?
Check all that apply

- Hiking/backpacking
- Skiing
- Hunting
- Fishing

- ☐ military
- ☐ pilot
- ☐ orienteering
- ☐ Other:

At what level(s) have you taken a geography course?

Check all that apply

- ☐ I have never taken a geography course
- ☐ elementary school
- ☐ middle school/ junior high
- ☐ high school
- ☐ community college, prior to undergraduate university study
- ☐ during undergraduate university study
- ☐ currently enrolled in a geography course

From the previous question, how many years has it been since your MOST RECENT geography course?

How would you rate your 'comfort level' or experience with using technology?

	1	2	3	4	5	
not comfortable at all	○	○	○	○	○	very comfortable

Do you play action-based video games?

Check only one option (1= "rarely": once per month or less) (2 = "sometimes": twice per month) (3 = "often": once per week) (4 = "regularly": multiple times per week)

	0	1	2	3	4	5	
never	○	○	○	○	○	○	frequently (at least once per day)

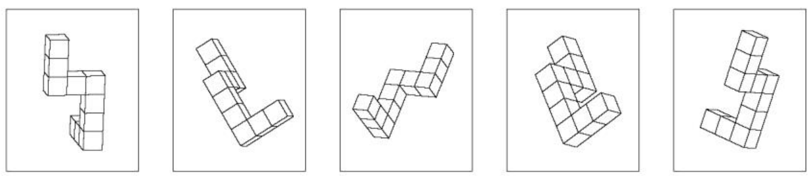
Appendix D

Mental Rotations Test of Spatial Ability

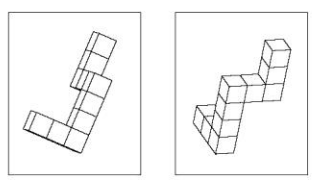
MENTAL ROTATIONS TEST (MRT-A)

This test is composed of the figures provided by Shepard and Metzler (1978), and is, essentially, an Autocad-redrawn version of the Vandenberg & Kuse MRT test.

Please look at these five figures



Note that these are all pictures of the same object which is shown from different angles. Try to imagine moving the object (or yourself with respect to the object), as you look from one drawing to the next.

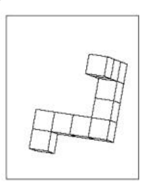


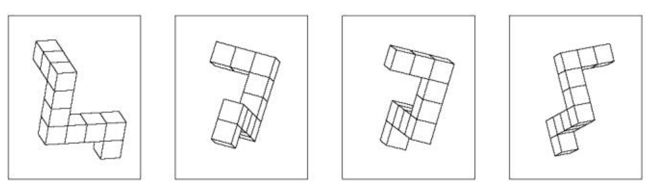
Here are two drawings of a new figure that is different from the one shown in the first 5 drawings. Satisfy yourself that these two drawings show an object that is different and cannot be "rotated" to be identical with the object shown in the first five drawings.

Now look at this object:

Two of these four drawings show the same object. Can you find those two? Put a big X across them.

1.

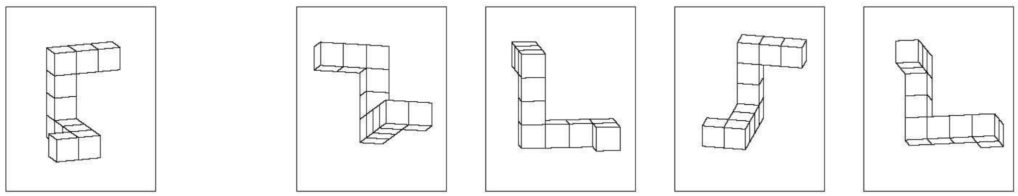




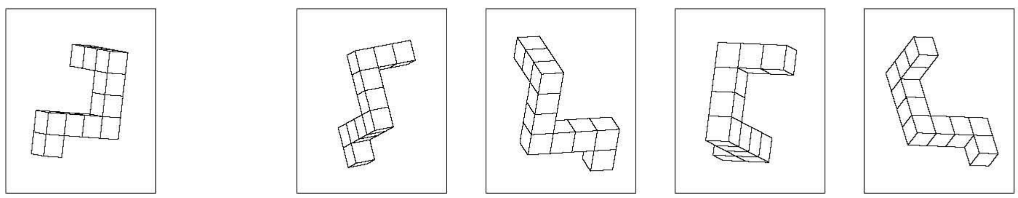
If you marked the first and third drawings, you made the correct choice.

Here are three more problems. Again, the target object is shown twice in each set of four alternatives from which you choose the correct ones.

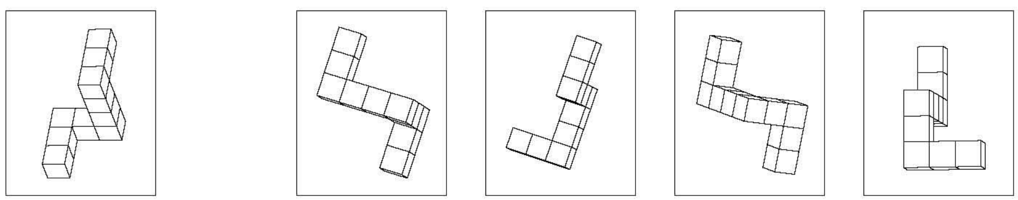
2.a



3.a



4.a



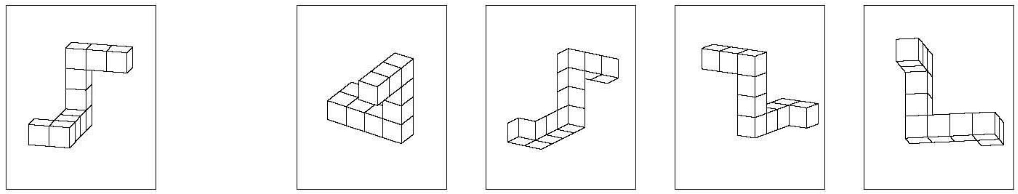
Correct Choice: 2: second and third
3: first and fourth
4: first and third

When you do the test, please remember that for each problem set there are two and only two figures that match the target figure.

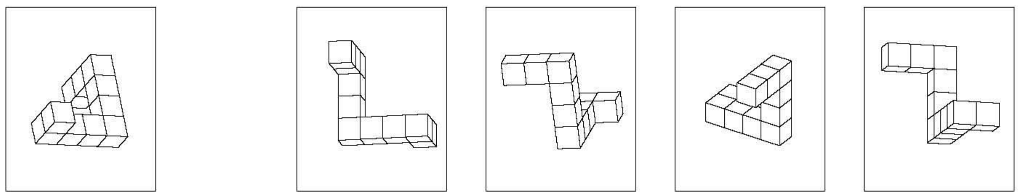
You will only be given a point if you mark off both correct matching figures, marking off only one of these will result in no marks.

3

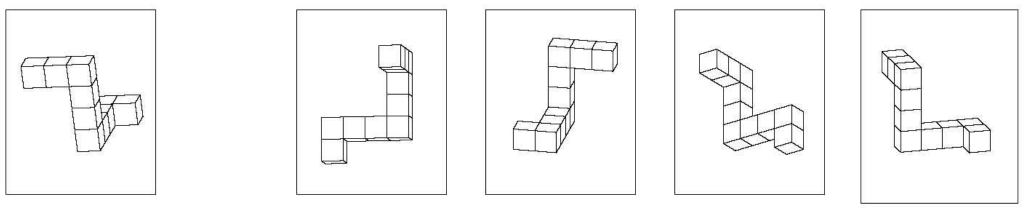
1.a



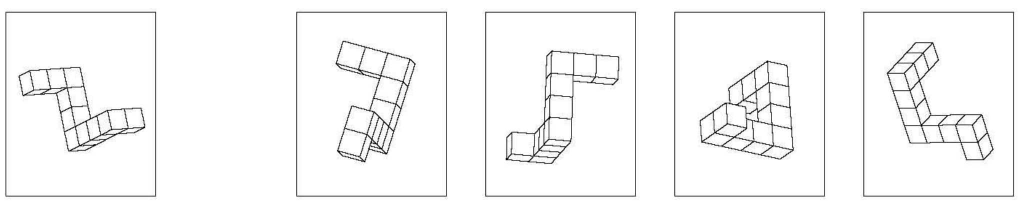
2.a



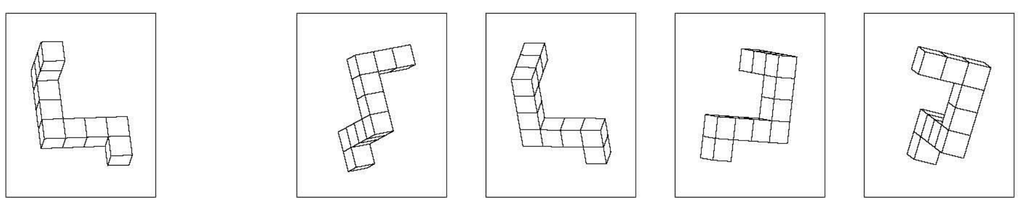
3.a



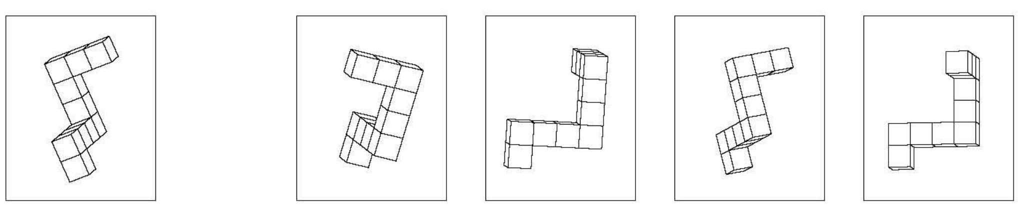
4.a



5.a

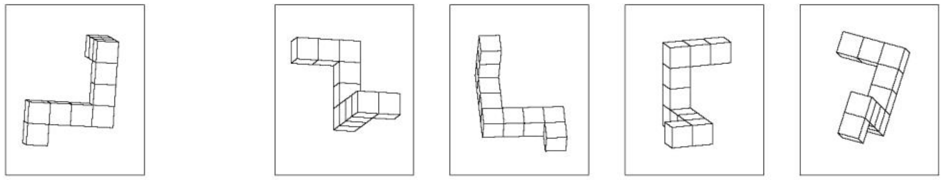


6.a

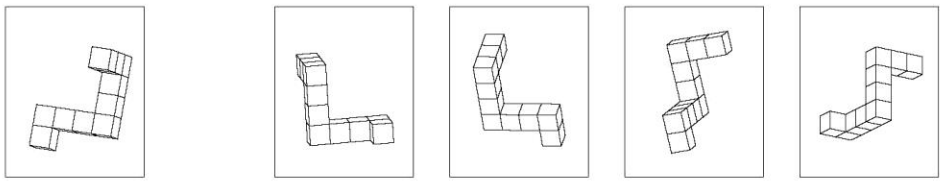


4

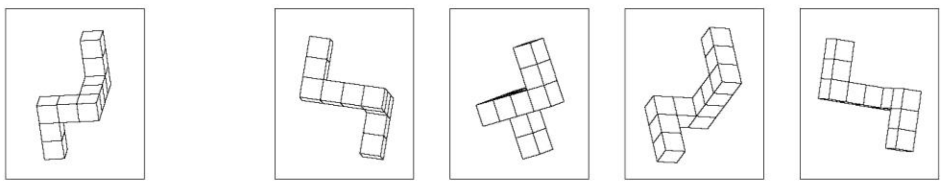
7.a



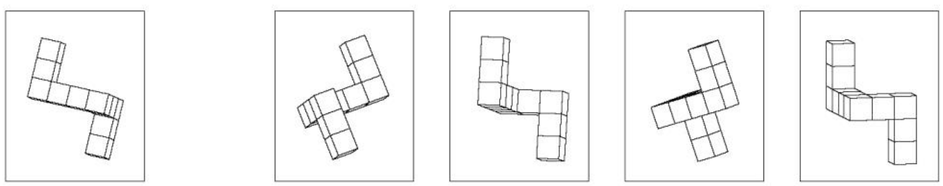
8.a



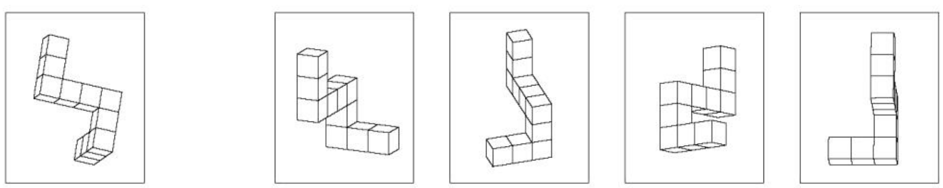
9.a



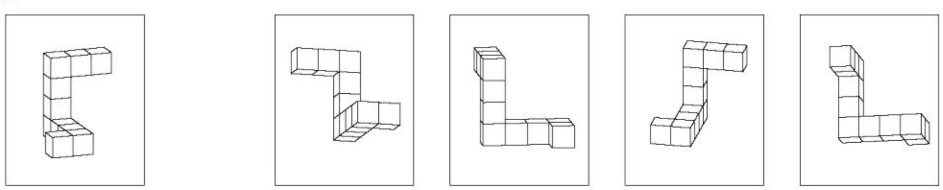
10.a



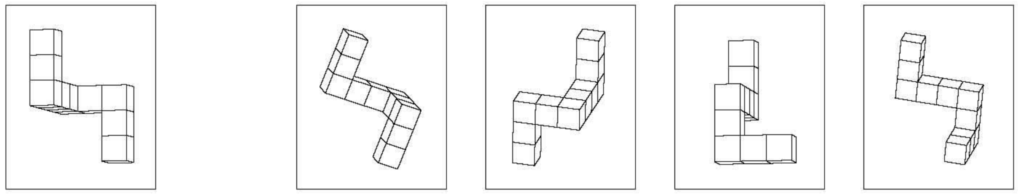
11.a



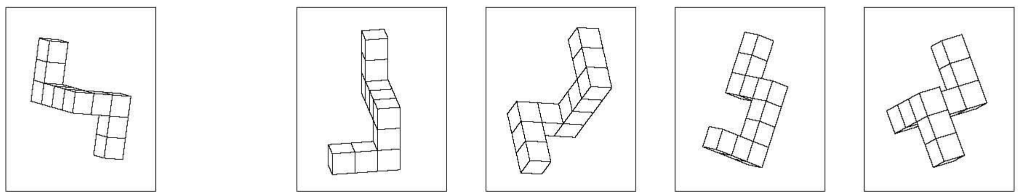
12.a



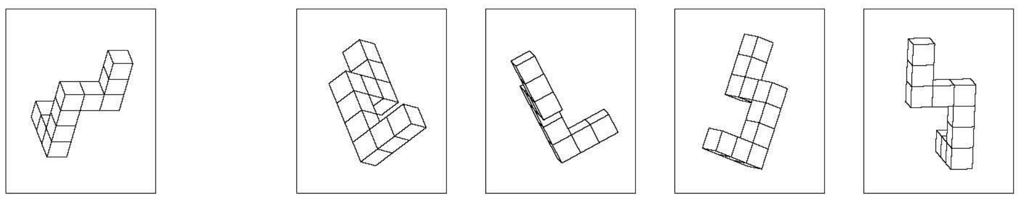
13.a



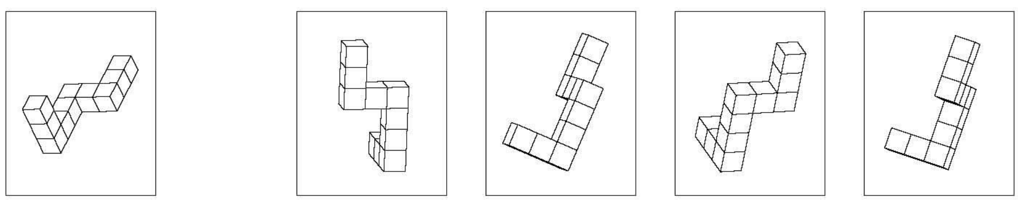
14.a



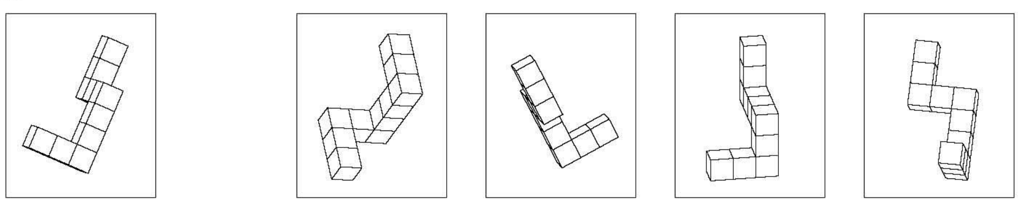
15.a



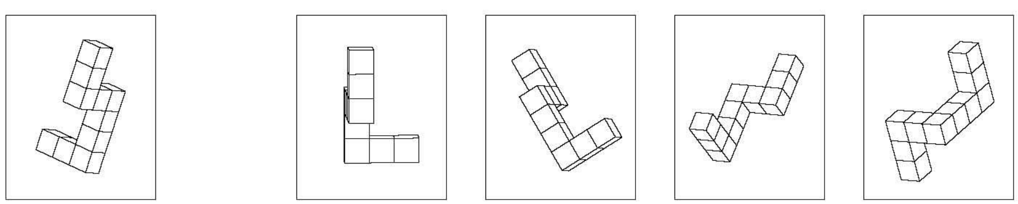
16.a



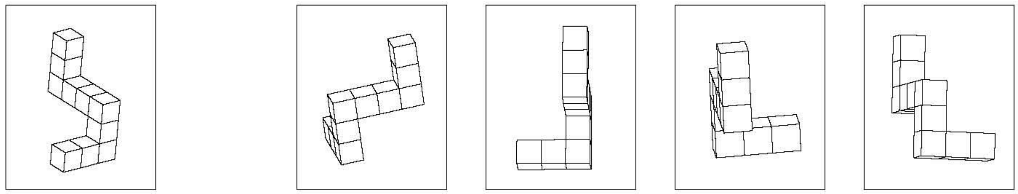
17.a



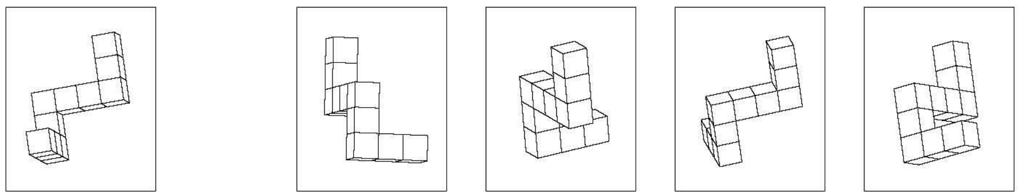
18.a



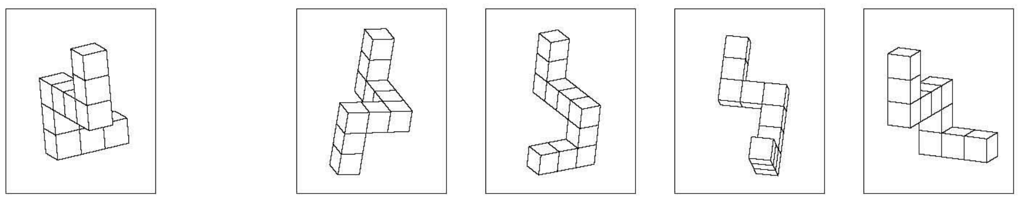
19.a



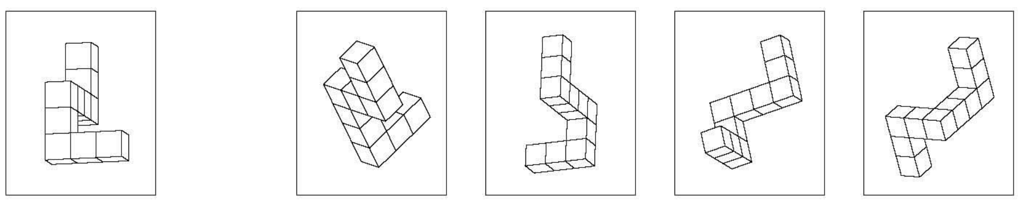
20.a



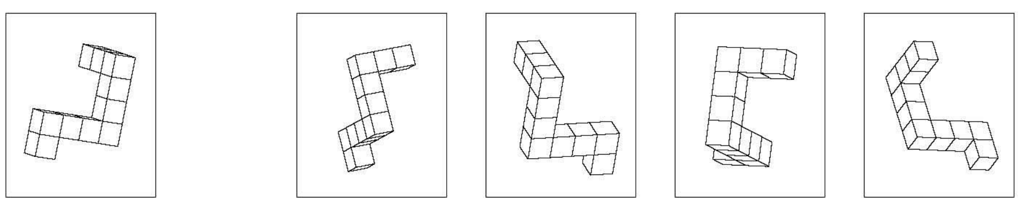
21.a



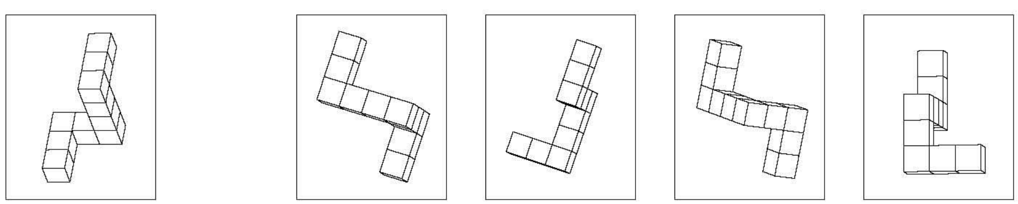
22.a



23.a



24.a



Appendix E

Mental Rotations Test Protocols

Michael Peters (1995). Revised Vandenberg & Kuse Mental Rotations Tests: forms MRT-A to MRT-D. Guelph (ON), Canada: Technical Report, Department of Psychology, University of Guelph.

Revised MRT© When referring to this test in a publication, please cite: Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R. and Richardson, C. (1995). A Redrawn Vandenberg & Kuse Mental Rotations Test: Different Versions and Factors that affect Performance. Brain and Cognition, 28, 39-58.

Four different variations, based on the original Vandenberg & Kuse (1978) Mental Rotations Test figures, which in turn are based on figures provided by Shepard (Shepard & Metzler, 1978), are available:

MRT-A This is the standard set, with stimulus figures redrawn from the original Vandenberg & Kuse set.

MRT-B. This consists of the same items as MRT-A, but in a different mix. This test is meant to be used as alternative test, of similar difficulty as MRT-A. In practice, it seems to have consistently lower means than A but the difference does not reach significance in our samples.

MRT-C While MRT-A and MRT-B involve mental rotation around the vertical axis, this test requires subjects to rotate the figures both around the vertical and the horizontal axis. This renders the test much more difficult.

MRT-D This test is identical to MRT-A, with the exception that all stimulus and target figures are rotated 90 degrees to the left. In order to solve the problems, subjects have to rotate the figures around the horizontal axis. This appears to be more difficult than rotation around the vertical axis.

Comments, scoring and procedures

You will find the V & K test in two forms in the literature, one with 20, and one with 24 problem sets. Our MRT is the 24 problem set. Each problem has a target figure shown on the left and four stimulus figures on the right. Two of these stimulus figures are rotated versions of the target figure, and two of the stimulus figure cannot be matched to the target figure.

Two ways of scoring are encountered in the literature. The first gives one point for each correct answer, and a point is subtracted for each incorrect answer. This would yield a maximum of 48 points. However, we use the other method of scoring, where one and only one point is given if both of the stimulus figures that match the target figure are identified correctly. No credit is given for a single correct answer. This means that the

maximum score obtainable on our test is
24. An excellent discussion of mental rotation performance and procedures can be found in Voyer, Voyer & Bryden (1995), and information specific to the revised V & K tests, as described here, can be found in Peters, Chisholm and Laeng (1995), and Peters et al. (1995).

Whether the test is given to single individuals or for groups, the way in which the experimenter introduces the test is quite important. Ideally, the initial introduction is done in a serious and moderately personal manner, so that subjects take the test situation seriously. At the same time, and this applies especially when the test is given to populations which are not familiar with this sort of testing, the experimenter has to avoid being threatening.

Some administration procedures can have unexpected results. For instance, we find that group testing can lead to better performance than single testing if there is a fair amount of competitiveness in the group tested. This does not create a problem when different groups are tested with similar procedures but when different procedures are used, between-group comparisons are difficult to make.

Giving the Test

Please note that this test is quite sensitive to prior exposure and this is why subjects should be given no more information prior to starting the test than is available on the first two introductory pages (1 and 2).
1. The subjects are asked to look at Page 1. The experimenter points out that the first five figures are all one and the same figure, rotated around the vertical axis. Here, the experimenter can make use of rotating the vertically extended hand in order to illustrate the axis of rotation. Subjects are asked to ascertain that these are all versions of the same figure. Subjects are then asked to proceed to the next set and point out that this set has two figures, both identical but different from the first set of five figures.

2. Proceed to the four problem sets. Now is the time to describe the nature of the problem sets to subjects.

Verbal instructions "You see a target figure on the left, and four stimulus figures on the right. In all problems sets there are two figures on the right which are rotated versions of the target figure, and two figures which cannot be made to match the target figure. In Problem set number 1, try to see which of the two figures are corrected. The answer is given below. The first and the third figures match the target figures. You have to find both of the correct answers to get a point for a problem. A single correct answer or a correct and an incorrect answer do not count."

"Now try the three problems on page 2. The correct answers are given below"

Give sufficient time to subjects to work through these problems, at least 5 minutes for the three problems on page 2.

"Please turn over your test booklet with face down"

The Test: Instructions

"We are ready to start when I say 'begin'. In each problem, remember, there are two and only two correct solutions, and you have to mark these by putting an X across the correct figure ::::experimenter illustrates::::. We do pages 3 and 4 and then we take a little break. You have 3 minutes for the pages 3 and 4. When I say 'stop', turn the test face down immediately, even if you are in the middle of a problem.

Note: you might wish to give 4 minutes, but if you do, you must be sure to mention this in your "method" description.

"Begin" "Stop, please turn your test booklet face down". ::: Couple of minutes rest: "Now we begin. Once again, you have 3 (4) minutes for the two pages. Please, open the test booklet at page 5 and begin the second half". "Begin"
-"Stop, please turn your test booklet face down".

References: Peters, M., Chisholm, P., & Laeng, B. (1995). Spatial ability, student gender, and academic performance. Journal of Engineering Education, 84, 69-73. Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R. and Richardson, C. (1995). A Redrawn Vandenberg & Kuse Mental Rotations Test: Different Versions and Factors that affect Performance. Brain and Cognition, 28, 39-58. Shepard, S., & Metzler, D. (1978). Mental rotation: effects of dimensionality of objects and types of tasks. Journal of Experimental Psychology: Human Perception and Performance, 14, 3-11. Voyer, D, Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of some critical variables. *Psychological Bulletin, 117*, 250-270.

Appendix F

Augmented Reality Sandbox Specifications

(Downloaded from:

http://idav.ucdavis.edu/~okreylos/ResDev/SARndbox/Instructions.html)

Instructions

This page will hold purchase, construction, setup, and calibration instructions for an Augmented Reality (AR) Sandbox.

Hardware Requirements

An AR Sandbox requires the following hardware components:

A computer with a good graphics card, running any version of Linux. The AR Sandbox software, in principle, also runs on Mac OS X, but we advise against it.

A Microsoft Kinect 3D camera. The AR Sandbox software, or rather the underlying Kinect 3D Video Package as of version 2.8, supports all three models of the first-generation Kinect (Kinect-for-Xbox 1414 and 1473 and Kinect for Windows). All three are functionally identical, so get the cheapest model you can find. **Note:** The second-generation Kinect (Kinect for Xbox One or Kinect for Windows v2) is not yet supported by the AR Sandbox software.

A digital data projector with a digital video interface, such as HDMI, DVI, or DisplayPort.

A sandbox with a way to mount the Kinect camera and the projector above the sandbox.

Sand.

Computer

The ideal computer for an AR Sandbox is a dedicated PC running a version of Linux, with a consumer-level, e.g., AMD/ATI Radeon or Nvidia GeForce, graphics card. The PC should have a good CPU, but does not need large RAM (2GB is sufficient to run the AR Sandbox software) or a large hard drive (20GB is sufficient to install Linux and the AR Sandbox software). While the AR Sandbox software runs under Mac OS X, we strongly recommend Linux because Linux-based installations are more stable.

We recommend an Intel Core i5 or Core i7 CPU running at least 3GHz (as of 10/2015), an Nvidia GeForce GTX 970 graphics card (as of 02/2015), and the current release of the 64-bit version of Linux Mint with Mate desktop (17.2 "Rafaela" as of 10/2015). The AR Sandbox requires that the vendor-supplied binary drivers for the graphics card be installed.

The AR Sandbox has two main components: the topographic map renderer, and the water flow simulation. The former is comparatively easy on CPU and graphics card, and works on most current laptops or mid-range PCs. The water simulation, on the other hand, requires a high-end graphics card like the recommended GeForce GTX 970. While the water simulation can be disabled to allow running the AR Sandbox from a lower-end computer, we do not recommend doing so.

The benefit of a dedicated PC is that the AR Sandbox becomes a computational appliance. Since the AR Sandbox software does not require a live Internet connection, it is possible to install the OS and AR Sandbox software, and then disconnect the computer from the Internet and never update the operating system or AR Sandbox software. The AR Sandbox software can be set up to start automatically when the computer boots, in which case it does not even require a monitor, mouse, or keyboard. Should newer versions of the AR Sandbox software become available, they can be uploaded via removable media.

Projector

Ideally, the projector should have a short throw length and a native 4:3 aspect ratio to match the field-of-view of the Kinect camera. The projector's native resolution is secondary; XGA (1024x768 pixels) is sufficient, as the sandbox overall's resolution is limited by the Kinect camera's 640x480 pixels. For practical reasons, short-throw projectors generally project above centerline, i.e., the bottom edge of the projected image appears above an imaginary horizontal plane through the projector's lens. The ideal projector for an AR Sandbox would be a centerline projector, so it could be mounted directly next to the Kinect camera. Since centerline short-throw projectors are rare and typically very expensive, the compromise is to mount an above-centerline projector above the sandbox's rear long edge, while the Kinect is mounted above the sandbox's center (see Figure 1).

We recommend the BenQ MX631ST above-centerline short-throw XGA DLP projector with 13000:1 contrast ratio and 3200 ANSI lumens, which generally sells for around $550. At maximum zoom, it matches the Kinect camera's field-of-view and can thus be mounted at the same height as the Kinect camera, simplifying setup.

We strongly recommend that the projector be connected to the PC's graphics card via a digital video connection, i.e., using an HDMI port on the projector, and an HDMI, DVI, or DisplayPort port on the graphics card. An analog connection, such as using a 15-pin

VGA port on the projector, leads to degraded image quality and can cause misalignment between the projected image and the sand surface.

Sandbox

The sandbox itself should have a 4:3 aspect ratio, to match the fields-of-view of the Kinect camera and the projector. The size of the sandbox is limited by the Kinect camera's minimum and maximum sensing distances, and the desired sandbox resolution. Due to the Kinect camera's approximately 90° field-of-view, the Kinect camera has to be mounted about as high above the sand surface as the sandbox is wide. The Kinect camera should be mounted directly above the sandbox's center point, looking straight down (see Figure 1).

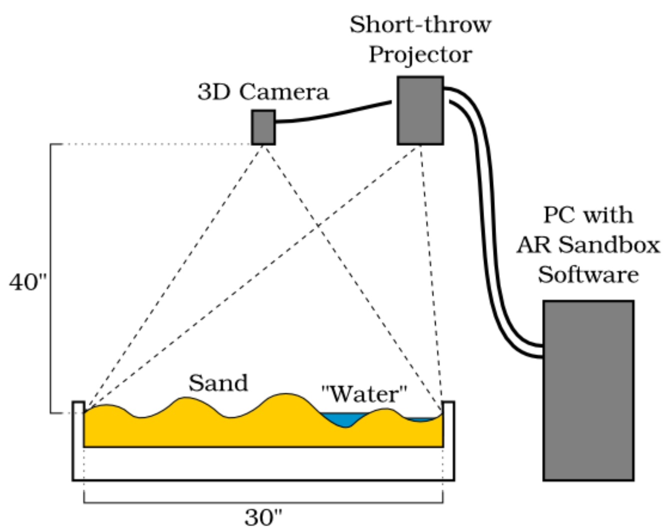


Figure 1: Typical arrangement of projector and Kinect camera above a 40"x30" (1mx0.75m) AR Sandbox. The short-throw projector is mounted at the same height as the Kinect camera, but above to the rear long edge of the sandbox to account for its above-axis projection.

We recommend a sandbox size of 40"x30" or 1mx0.75m, leading to a Kinect camera mounting height of approximately 40" or 1m. At this height, the camera's nominal horizontal resolution is 1.56mm, and its effective horizontal resolution is high enough to resolve features on the order of 5mm. Vertical resolution at the same height is 2.79mm. Increasing the size of the sandbox increases the required height of the camera/projector mount by the same factor, and not only reduces horizontal resolution, but also vertical resolution. In a 2mx1.5m sandbox, for example, nominal horizontal resolution is 3.12mm, and vertical resolution drops to 11.16mm (vertical resolution is roughly proportional to height squared).

Sand

The sandbox should be filled with sand to a depth of around 4" or 10cm. At 40"x30", this totals 2.77 cubic feet or $75dm^3$ or 75l of sand, weighing approximately 198lb or 98kg. We recommend Sandtastik White Play Sand, for between $15 and $25 per 25lb, or $120-$200 total. Sandtastik sand has excellent projection properties, but a shallow angle of repose when dry. We recommend keeping the sand slightly moist to make it moldable. Adding 1 cup or 0.25l of water to 198lb or 98kg of sand and mixing thoroughly is sufficient.

Health Concerns

Regular sand is basically crystalline silica, primarily in the form of quartz. While silica is non-toxic when ingested orally, inhaling very fine silica dust can cause adverse health effects. The recommended Sandtastik play sand does not, according to the manufacturer, contain or release fine silica dust. Neither should alternatives such as Moon Sand or Kinetic Sand, as they are either not made from actual sand, mixed with a binding agent, or surface-treated with a polymer. Regular sand, such as bought in bulk from hardware or home improvement stores, should be washed before use to reduce the amount of fine dust particles contained in it. Here is a how-to guide on washing play sand (step 5, baking, is optional).

Hardware Construction

Building the sandbox itself, and mounting the Kinect camera and projector above it are left as exercises to the reader (see Figure 1 for a rough layout sketch). Our own prototype AR Sandbox is built from wood and some metal. The sandbox proper is plywood on top of a sturdy wheelbase (for mobility), slathered generously with polyurethane to make the sandbox waterproof and rot-resistant (We are using small amounts of water to make the sand moldable). The projector and camera head assembly is made from aluminum slats, and the entire assembly is suspended above the sandbox from a vertical steel pipe. The head assembly offers limited adjustment for camera and projector position and orientation to allow physical alignment between sand surface, camera field-of-view, and projected image. Most importantly, the projector should be oriented such that it projects on-axis onto the "ideal" level sand surface. This minimizes distortion effects and focus problems.

Software Installation

The basic process is to install the Vrui, Kinect, and SARndbox software packages, in that order, on top of a Linux or Mac OSX operating system. See the README file included with the AR Sandbox software package for details.

If you are a first-time Linux user, please follow the detailed step-by-step software installation instructions posted on the AR Sandbox support forum, and/or watch the software installation video.

Kinect Camera and Projector Alignment and Calibration

As mentioned above, the projector should be aligned physically such that it projects orthogonally or on-axis onto the "ideal" level sand surface, and that its projected image exactly fits the sandbox. This usually means that the projector has to be mounted vertically above the back long edge of the sandbox, due to short-throw projectors' usual above-axis projection. The Kinect camera should be aligned such that its field-of-view exactly matches the sandbox as well. This is most easily achieved by mounting the Kinect vertically above the center of the sandbox, orienting it straight down, and then fine-tuning its position and orientation while observing its depth image in the RawKinectViewer application.

Kinect Camera Intrinsic Calibration

Ideally, the Kinect camera should be calibrated intrinsically to capture a proper 1:1 representation of the sand surface. All Kinect cameras are pre-calibrated at the factory, but while that calibration is serviceable, it is not very good. Most importantly, Kinect cameras benefit from per-pixel depth correction; without that, a Kinect will capture a completely flat surface as a bowl-like shape, which noticeably affects elevation contour lines and water flow. Intrinsic calibration is described in detail in the Kinect package's README file, and the following two videos:

Intrinsic Kinect Camera Calibration with Semi-transparent Grid

Intrinsic Kinect Camera Calibration Check

There is no video for the per-pixel depth correction calibration step, which needs to be done first, but it is fairly simple and explained in the README file.

For simplicity, intrinsic calibration should be performed before the Kinect is mounted above the sandbox.

Kinect Camera Extrinsic Calibration

Extrinsic calibration establishes the position and orientation of the Kinect with respect to the sandbox, and the mathematical plane equation of the "ideal" level sand surface and its position and size in 3D space. It is explained in the following two videos:

AR Sandbox Calibration - Step 4

AR Sandbox Calibration - Step 5

Unlike intrinsic calibration, extrinsic calibration has to be performed with the Kinect mounted above the sandbox, and needs to be repeated any time any changes are made to the physical sandbox layout, i.e., any changes to the position or orientation of the projector or camera or the sandbox itself. Also, depending on the sandbox assembly's mechanical stability, it might have to be redone on a regular basis (every month or so). Fortunately, the process only takes a few minutes.

Kinect Camera and Projector Calibration

The final calibration step is to measure the precise alignment of the Kinect camera's field of view and the projector's image. This is done using a calibration prop and a dedicated calibration utility, explained in detail in the following video: New AR Sandbox Calibration Procedure.

Like extrinsic calibration, this procedure has to be performed any time the AR Sandbox' physical layout changes in any way/

Running the Augmented Reality Sandbox

See README file included with the AR Sandbox software package.

Appendix G

Instructional Design Process Model

Instructional Rationale

Educators tend to not be well versed in understanding or teaching the underlying spatial associations that are necessary to translate 2-D information from a map into a 3-D mental schema (Collins, 2014; Earle, 2008). The cognitive aspects of geographic education, in terms of how we conceive of space, is heavily influenced by traditional educational practices. Hence, spatial thinking is rarely addressed within school curricula (Montello, 2009). This is in spite of calls by leading cartography and GIS experts on the need to understand spatial cognition in order to design interfaces that promote critical thinking and problem solving (MacEachren & Kraak, 1997). In spite of an incomplete understanding of spatial cognition, it has not prevented researchers from designing educational programs and software using geographic skills to teach spatial thinking (Bednarz, 2004; Bednarz et al., 2008; Lloyd, 2001a, 2001b; Lynch et al., 2008; Schultz, Kerski, & Patterson, 2008). Many of these programs incorporate GIS or Google Earth software to build maps from layers of data and to query the map to extract information about spatial relationships. Although early GIS software created 2-D maps, the latest versions (ArcScene GIS 3-D Analyst® or CityEngine®) can now create 3-D maps. Recent advances in AR and VR technology have provided new types of interfaces for visualizing maps (Chang et al., 2017; Modjeska & Chignell, 2003; Nam, Li, Yamaguchi, & Smith-Jackson, 2012; Piburn et al., 2002). These new interfaces can enhance students' spatial thinking skills, which have been associated with success in Science, Technology, Engineering, and Math (STEM) education and careers (Clifton et al., 2016; Skulmowski et al., 2016; Stieff & Uttal, 2015; Taylor & Hutton, 2013).

The Instructional Design Process Model

The instructional treatment for each experimental condition was designed using the Kemp Instructional Design Model (Morrison et al., 2011) framework and was further informed by results of a November 2015 pilot study. Instructional design (ID) can be defined as “a systematic process that is employed to develop education and training programs in a consistent and reliable fashion” (Gufstason & Branch, 2007, p. 11). Morrison et al. (2011) expanded this definition by emphasizing application of evidence-based research on learning and educational theories to address a known performance problem.

The Kemp model (Figure 10) consists of a nine phase process for designing instruction: (1) defining Instructional Problems; (2) identifying Learner Characteristics; (3) creating a Task Analysis; (4) crafting Instructional Objectives; (5) arranging instructional delivery in a meaningful order through Content Sequencing; (6) selecting appropriate Instructional Strategies; (7) determining how to deliver instruction by Designing the Message; (8) creating the instructional resources during the Development of Instruction phase; and (9) constructing Evaluation Instruments. These nine phases are located within the inner circle of the model. There are two sets of outer rings represent ongoing factors that impact all nine phases. Within the inner ring Revision, Formative Evaluation, Summative Evaluation, and Confirmative Evaluation emphasize the importance of iteratively revisiting each of the nine inner phases to refine and improve the design of each and, hence, the effectiveness of the overall instructional process. The outer ring lists factors that may be somewhat external to the design process (Planning, Implementation, Project Management, and Support Services) which, nonetheless, influence the design outcome as the entire process from planning to implementation is considered.

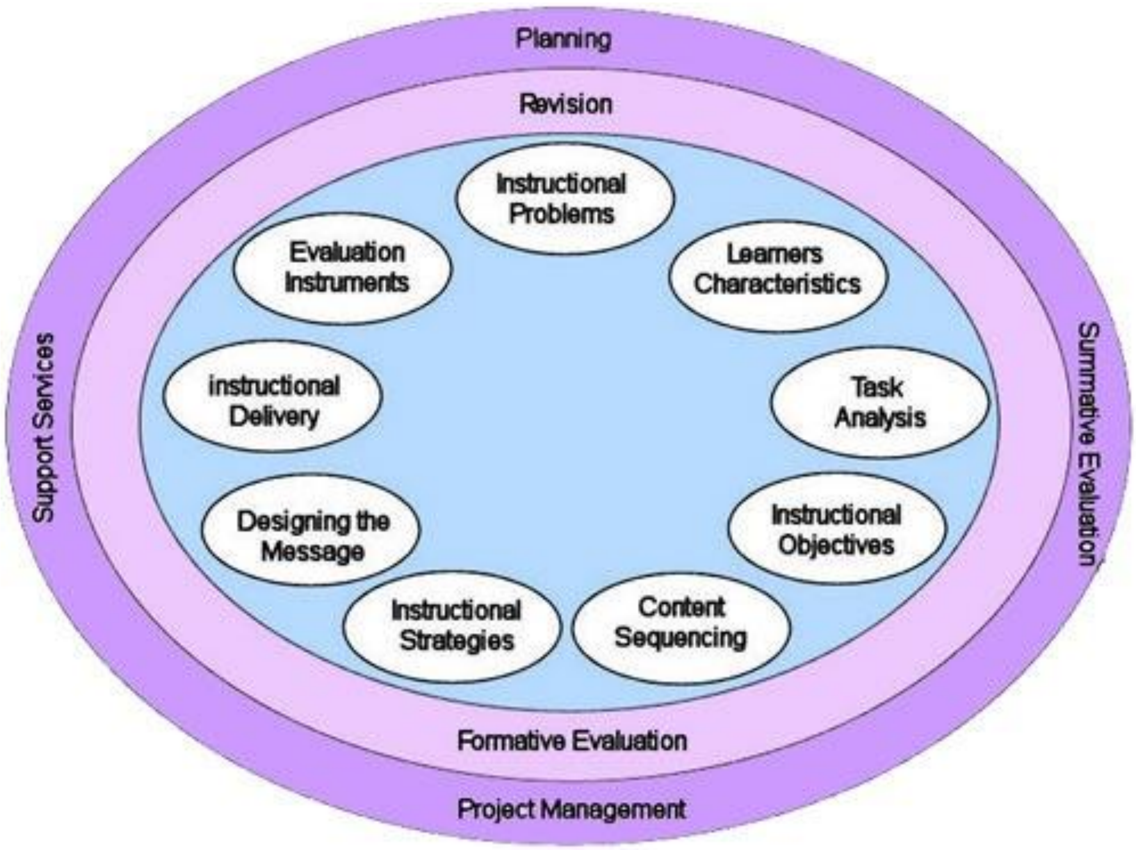


Figure 10. The Kemp Instructional Design Model (Morrison, Ross, Kalman, & Kemp, 2011, p. 12)

The Kemp model was chosen over other ID models for a number of reasons. Firstly, each element within the process stand as independent, yet flexible components. The circular nature of the model allows the designer to begin at any step, hence its flexibility. This circularity emphasizes the iterative, non-hierarchical, and interconnected nature of the ID process. Other ID models such as ADDIE and ICARE (Hoffman & Ritchie, 1998) are more linear in nature, which is not truly reflective of the need for frequent design revision. The Rapid Prototyping Model, (Tripp & Bichelmeyer, 1990) was considered due to its non-hierarchical and iterative structure, like the Kemp model, and its affinity for computer-based instruction. It was rejected because the technology used in this study was already in situ and instruction and assessment was not solely managed by computer.

Like other ID models, the Kemp model emphasizes the importance of understanding the learner's needs and adjusting instruction to suitable learning outcomes. Yet the Kemp model has an explicit focus on evaluation: formative evaluation that influences the emerging design, summative evaluation that gauges the overall learning effectiveness of the design, and confirmative evaluation that assesses the long-term learning retention of participants. Its emphasis on formative and summative assessments, both of which were employed in this study, made it the most appropriate choice for guiding the design of multimedia instruction, learning practice, and learning retention activities.

This study will follow the stages of the Kemp Model in a clockwise order, beginning with an Instructional Problem. However, the Evaluation Instruments phase will precede Content Sequencing. This is a deliberate decision made by the researcher, who supports the Understanding by Design ID model's claim that assessments should be designed before instructional materials are created (Wiggins & McTighe, 2005). This ensures that the assessments match the objectives.

Instructional problem. The ID process begins by determining if training is appropriate to remediate the problem (Morrison et al., 2011). The following items establish a learning deficiency associated with reading topographic maps that can be remediated through training and practice:

- Ability to read topographic maps is an important skill for geoscientists
- Topographic maps are difficult to interpret
- Lack of evidence-based best practices and of a coherent set of instructional resources for teaching topographic map interpretation
- Novice geoscientists at the study site do not receive training in reading topographic maps, yet it is regarded as an important skill for advanced coursework
- Learning to read topographic maps also teaches spatial thinking skills, which contribute to success in STEM fields, such as geoscience

Learner characteristics. For sampling purposes, the student population is restricted to students enrolled in introductory geoscience and psychology courses. There was no assumption that this sample is representative of the university population, however. Most students will have used thematic maps, but few will have knowledge of or skill using topographic maps. Therefore, it was assumed that there was a diversity of familiarity/experience with topographic maps, which was indicated in the demographic survey: the median value of prior experience using topographic maps was two out of a Likert scale of 5, indicating limited exposure to this type of map.

Task analysis. Each instructional outcome (see Table 12) and each TMA posttest question was examined to determine what type of learning task was required:

either rule using (performing a task), concrete concept (identification of key features/details) and verbal information (recall of facts) to (Davidson-Shivers & Rasmussen, 2006). The emphasis was on finding instructional and assessment tasks that required the participant to answer questions during multimedia instruction and particularly during learning practice where hands-on practice was paramount to reinforcing learning.

There are several instructional techniques used to teach topographic map skills to novices. Almost all are based on visual instruction using examples. Typically, leaners are provided with a series of contour lines that correspond to a topographic feature (mountain, ridge, valley, basin, etc.). User interactivity is limited; the most common learning practice activity is drawing topographic profiles, which involve translating a cross-section of a 2D topographic feature to a scale drawing of its topographic contours. Although drawing or tracing on maps is sometimes used as a learning activity, drawing individual topographic features from memory is atypical. This assessment was based on a review of college-level contemporary geography, geology, and mapping textbooks (Christopherson, 2010; Dorling & Fairbairn, 1997; Levin, 1986; Lounsbury & Aldrich, 1986; Petersen et al., 2011; Selby, 1985; Strahler, 1987).

Instructional objectives. The primary instructional objective was to increase students' ability to read topographic contour maps, thus enhancing their ability to think spatially within this context. As a secondary objective, this learning may benefit them in their current geoscience class or in future STEM-related courses. Note that instructional objectives were distinct from the research objectives of this study. Based on a review of contemporary geography, geology, and mapping texts, as noted above, 17 learning objectives or outcomes were identified and summarized in Table 14.

Evaluation instruments. Both the learning practice activity questions and the mTMA posttest was aligned to match the 17 instructional objectives, as shown in Table 12. The learning practice questions were designed by the researcher and examined for face validity by two geoscience experts. The TMA has been used in previous research (Newcombe et al., 2015) and has good reliability ($\alpha = .76$). The MEMS scale (Paas & Van Merriënboer, 1993) is highly reliable ($\alpha = .90$).

Content sequencing. The nature of the research design provided the overall structure from instruction in basic topographic map reading skills, to a formative assessment learning practice activity, to a posttest measure of learning retention from practice. With regard to the multimedia slideshow (see Appendix H), presentation of terms and concepts were structured from easy to more difficult. The basic structure of the multimedia slideshow was as follows: (1) introductory sequence that established the purpose of instruction (slide 1), an advance organizer to provide learners with a sense of flow of instructional (slide 2), a basic definition of a topographic map with several examples (slides 3-5); (2) a definition and practice with interpreting contour lines—a key concept; (3) a series of examples of topographic features, starting with simple features (slopes) and progressing to increasingly more complex features, culminating in spurs, which were a combination of hills and valleys (slides 8-18); and (4) a series of slides that required learners to apply their knowledge of the topographic features covered in slides eight to 18 when viewed from 2D and 3D perspectives (slides 19-31).

Table 14

Learning Outcomes for Multimedia Instruction, Learning Practice Activities, and the Modified Topographic Map Assessment Posttest by Assessment Question

Learning Outcome	**Multimedia Instruction Slide #**	**Learning practice question #**	**mTMA posttest question #**
1. Learners will correctly identify topographic contour lines on a map (Core competency)	3	1-20	1-25
2. Learners will correctly identify topographic contours as a series of concentric circles, with the same elevation at each point on the contour line (Core competency)	4,6	1-20	1-25
3. Learners will calculate elevations using contour line data	6	5,6	4,6
4. Learners will indicate lowest and highest points on a topographic map		1,2,8,13	4
5. Learners will determine elevation differences between one or more features		3,6,9	5,7,9
6. Learners will differentiate between low angle and steep slopes and cliffs		9	5,9
7. Learners will correctly identify geographical features based on an understanding of shapes and spacing of contour lines: (a) low angle vs. steep slopes; (b) hill vs. mountain; (c) ridge and ridgeline; (d) saddle; (e) mesa vs. butte; (f) valley; (g) spur; (h) depression/basin	8-18, 27	3,4,7-14	1-7
8. Learners will draw (non-AR condition) or sculpt (AR condition) each of these features during learning practice activity only	N/A	3,4,7-14	N/A
9. Determine if one location can be seen from another based on either (a) obstructing, higher elevation feature, or (b) clear view (no obstruction)	20	15-17	3,8

Learning Outcome	Multimedia Instruction Slide #	Learning practice question #	mTMA posttest question #
10. Learners will draw or trace direction water will flow in valleys and drainages and where it will settle to form a basin	16,17	20 (a-e)	2,10,11,12
11. Learners will draw or trace a path from one point to another on a topographic map to indicate the least steep (easiest) line of travel	21	18,19	1, 17
12. Learners will correctly match a cross-section on a 2D map to the correct elevation profile	5		13,14
13. Learners will correctly identify 3D views of a given landscape when given a 2D plan view from a sample topographic map	27		15
14. Learners will correctly match a sample 2D plan view to the corresponding 3D landscape view	23		18
15. Learners will correctly match a sample 3D landscape view plan view to the corresponding 2D plan view	28-30		8
16. Learners will correctly identify the direction they are facing on a 2D map when given a 3D scene	23		15,16
17. Learners will correctly identify the direction they are facing on a 3D scene when given a 2D map	20,23		17

Note. Multimedia instruction and Learning practice activities cover all learning outcomes necessary to complete the mTMA posttest questions. Question 8 is specific to the map interactions during the learning practice activity, which reinforced learning outcome 7, but could not be reproduced for assessment in the mTMA posttest.

Instructional strategies. The design of the multimedia slideshow and the learning practice activity were informed by multimedia learning theory because of its propensity to enhance learning and reduce cognitive load in digital environments. Novel visual data interacts with prior knowledge in WM before entering long LTM. Yet, limits on WM make it difficult to process multiple graphical elements simultaneously (Harrower, 2007; Sweller, Van Merriënboer, & Paas, 1998). Consequently, the goal of multimedia instruction was to maximize germane cognitive load (memory associated with effective mental processing) and minimize extraneous (distractive) cognitive load (Ayres & Paas, 2012; Chandler & Sweller, 1991; Clark & Mayer, 2011; Mayer, 2005; Paas, Renkl, & Sweller, 2003) by attending to multimedia and cognitive load theory recommendations. Otherwise, poorly-designed user interfaces may actually increase the latter and thus inhibit learning (Skulmowski et al., 2016; Song et al., 2014). Managing cognitive load by providing user controls or structuring animations via pre-training, segmenting, and narration are important features of optimal user interfaces (Clark & Mayer, 2011). Harrower (2007) claimed that it is essential to allow users to replay and vary speed and manipulate geovisual animations since user control increases germane cognitive load. Pause and playback options were available on each slide in the instructional slideshow, thereby allowing participants to review previous slides. The ARS accomplishes Harrower’s suggestion of interactive animation with a moldable surface that promotes active engagement through touch. Yet if the actions required to control the ARS are not automatic (either on the part of the user or the software) and there is no pre-training, a distractive split attention effect is possible, because the user must focus on manipulation while simultaneously examining the image. Spatial and temporal contiguity thus become important design considerations (Crooks et al., 2008). The ARS avoids this problem by displaying changes in the topographic surface image in real time based on changes in sand level. Intuitive controls that give users free reign to explore TUI maps demonstrate the potential to maximize germane cognitive load—a pre-conditions for maximum learning.

Designing the message. Text was kept to a minimum, mostly as labels, with instruction primarily delivered through narration from the researcher (thereby satisfying the multimedia principle) using non-technical language who appeared in a video clip in the top right corner while speaking (fulfilling the personalization principle with a pedagogical agent) (Clark & Mayer, 2011). There was a deliberate emphasis on using a consistent background with a graphic as the dominant visual feature in each slide. To facilitate comparison of abstract topographic features in plan (overhead) view as shown on a topographic map with contour lines alone, an adjacent drawing of the same landscape as it would appear to a viewer. This comparative method of instruction was based on a similar instructional technique used in a research study by Potash et al. (1978) for the source images. Use of simple line drawings placed side by side meets Clark and Mayer's (2011) coherence and contiguity principles respectively. Sample images from the First Volunteer U.S. Cavalry Regiment (Map Reading, n.d., Retrieved from: http://www.1stusvcav.com/Techniques/Move/map_reading.html) were used to compare 2D topographic map renderings to landscape illustrations. Placing the two images side-by-side aligned with the principle of contiguity for multimedia learning (Clark & Mayer, 2011; Mayer, 2005). This juxtaposition was intended to provide a cognitive scaffold to facilitate mental abstraction from a 2D topographic representation to a realistic, 3D landscape view and vice versa.

Instructional delivery. All participants, regardless of any prior experience with topographic maps, watched the 20 minute slideshow to provide a common level of understanding. Since this study examined the effects of interaction on learning, a series of 20 learning activity questions were designed to reinforce the concepts and skills covered in the multimedia slideshow with hands-on practice prior to a posttest assessment of learning retention. The sequencing of practice questions were aligned with the learning outcomes but were not delivered in the same order. Some of the activities were not covered in the multimedia instruction; they were deliberately included to foster high-order learning through application of the basic concepts introduced in the slideshow. For example, participants were asked to trace or draw the lowest angle route from one location to another. This required an understanding of relative slope angle based on contour line spacing. Another question extended the concept of valleys as drainage basins by having participants trace or draw the path rain would follow during a storm.

Opportunities were provided for learner practice, followed by feedback from the narrator. In some cases the correct answer was narrated, in other instances, the narrator drew on the slide to identify key features. The mental effort rating slides interspersed with slides 19-31 were not instructional elements, but rather used as part of the research study to gauge mental effort required to translate 2D and 3D map views. The results from this analysis were not addressed in this study as they are intended for a subsequent and separate publication.

.

Appendix H

Multimedia Instructional Slideshow

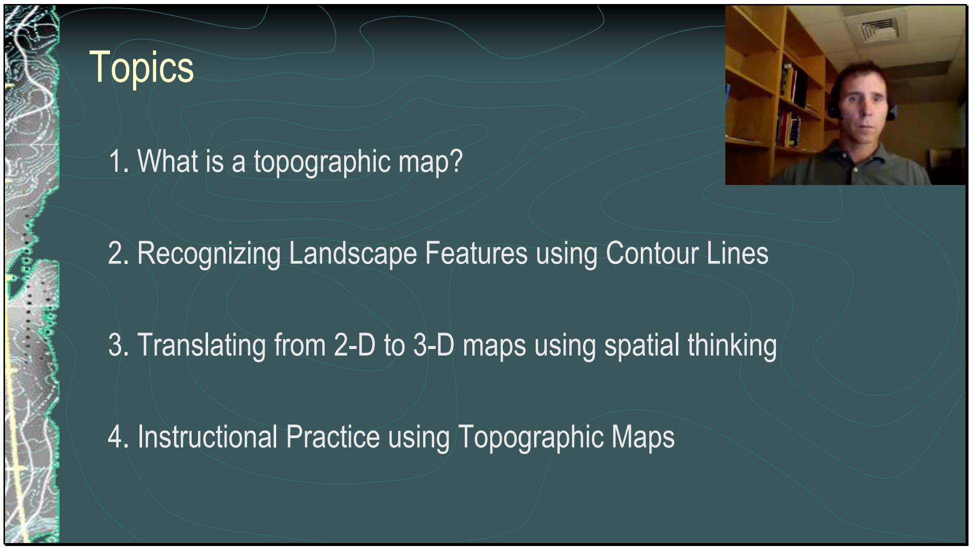
Slide 1

Have you ever looked at a map and noticed how it doesn't do a particularly good job of showing you what the landscape really looks like?

This is because maps summarize a lot of spatial information into a small space. Topographic maps, for example, have to show three dimensional features of the landscape on a two dimensional sheet of paper. This requires users to transfer information from a two dimensional perspective to a three dimensional perspective and vice versa. Not an easy task for many people!

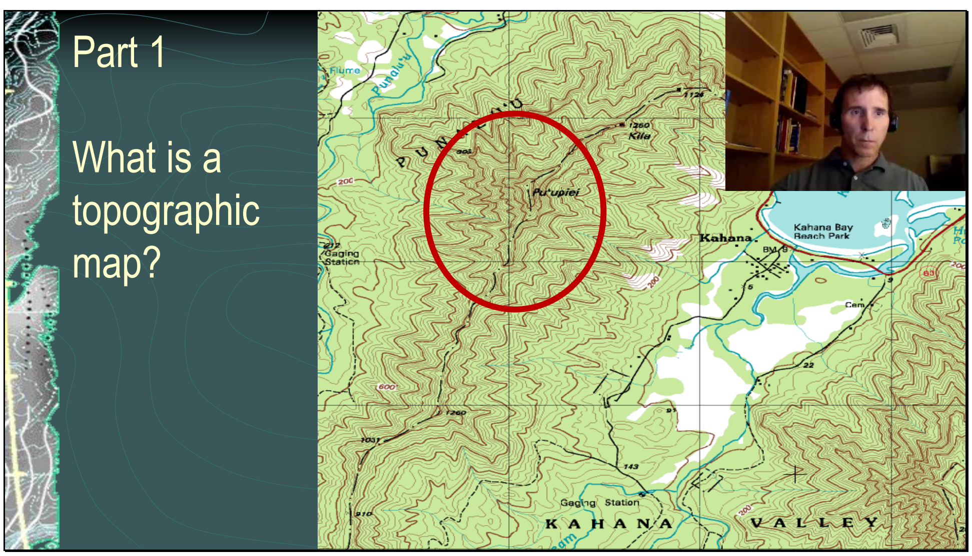
To be proficient in interpreting topographic maps, we must use abstract spatial thinking skills. The purpose of this experiment is to measure performance and mental effort when learning new ways to read topographic maps.

Slide 2



This instructional module will cover these four major topics

Slide 3



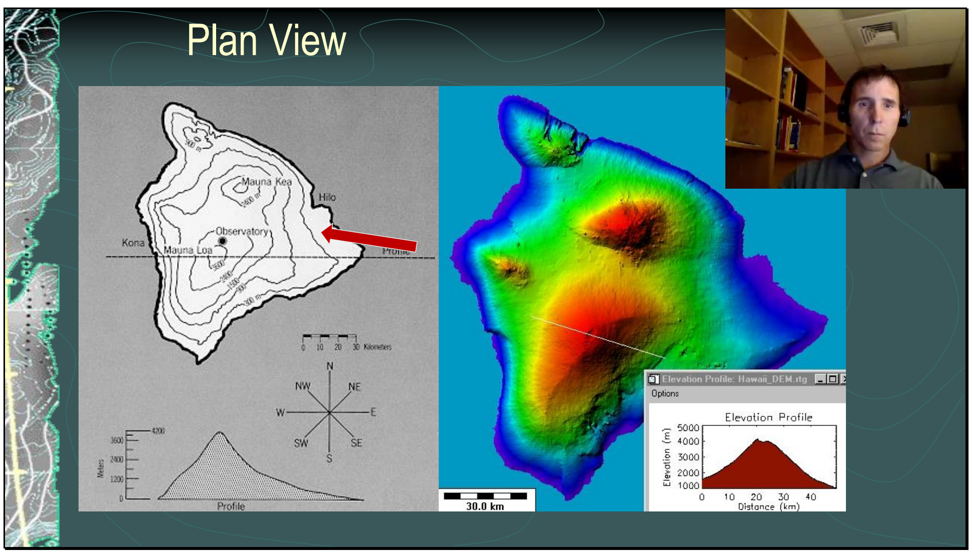
*A **topographic map** is a graphic representation of the earth's surface. It displays spatial information such as location, elevation, distance, & direction.*

Topographic maps show the shape of landforms and elevations above sea level in either meters or feet

- *The defining characteristic that makes any map a topographic map are **<u>contour lines. They are shown here (T) and represent a tall mountain</u>***

- ***<u>Contours</u>** are imaginary lines that connect points of **<u>equal</u>** **<u>elevation and form concentric circles</u>***

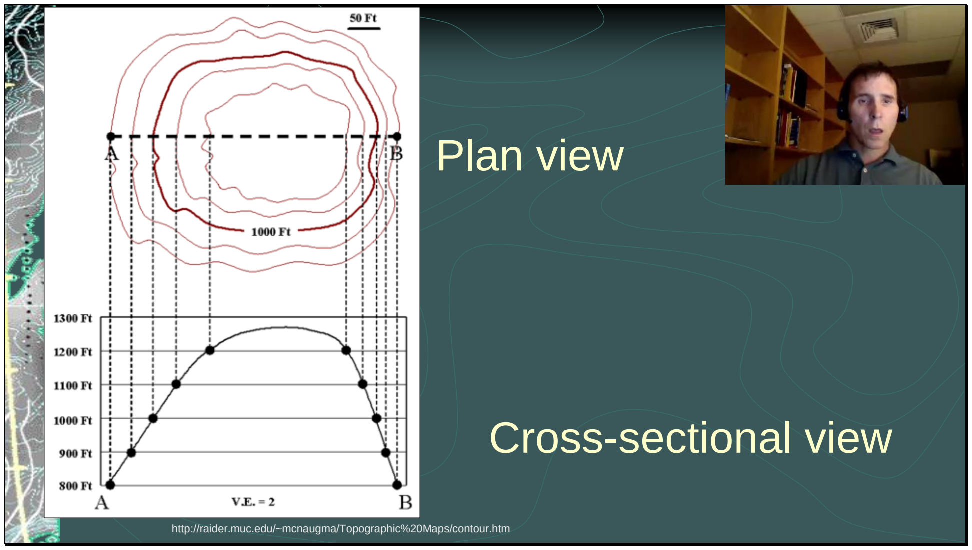
Slide 4



Most topographic maps are shown from directly overhead, what is referred to as a plan view. Color is sometimes used as a visual aid to understanding relative differences in elevation.

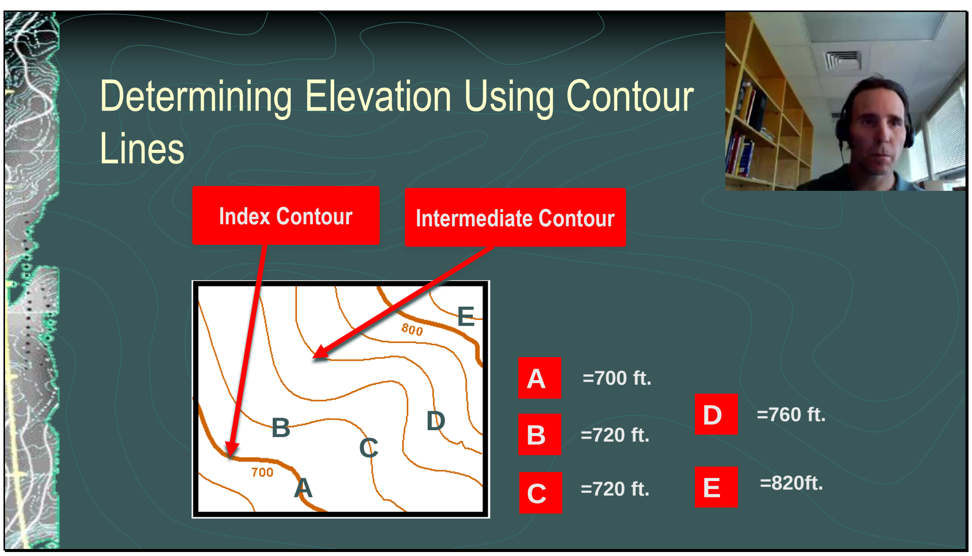
Note that as elevation increases, the concentric circles become smaller. This is an important concept to remember

Slide 5



Another way to understand how two dimensional surfaces on a topographic map represent a three dimensional object, is to create a cross section. Note how each contour line from the plan view is associated with a particular elevation in the cross-section view. Creating a cross-sectional profile indicates how elevation changes with horizontal distance.

Slide 6

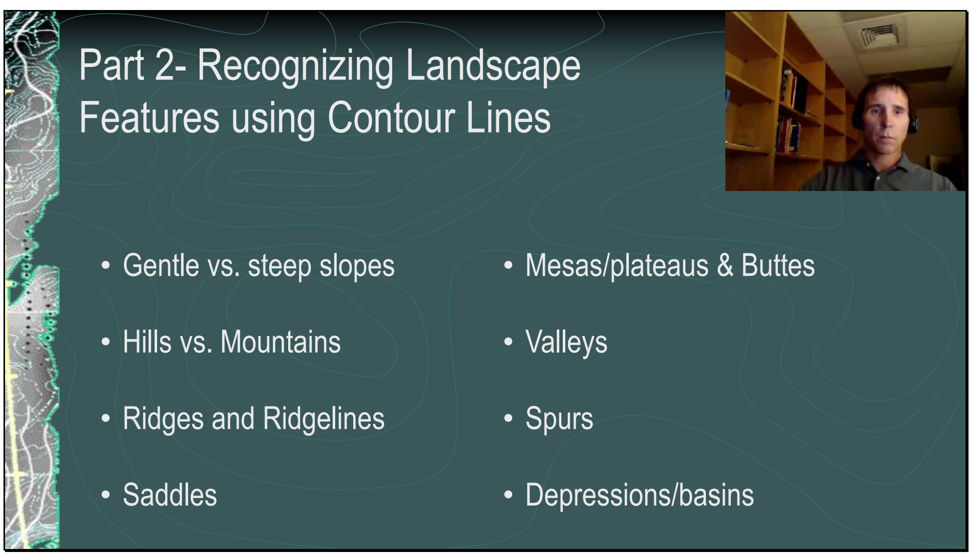


We can use contour lines to indicate elevation, which is the height in feet or meters above sea level. Topographic maps use two types of contour lines. Index contours are bold lines with the elevation listed, as shown for line A. (T). Lines that do not have elevations listed are called intermediate contours (T).
Topo maps are labelled with a contour interval, which is the difference in elevation from one contour to the next. You can use this value to calculate the elevation of an intermediate contour by counting from the nearest index contour. At any point along a contour line the elevation is the same.

So, what would be the elevation of point B? (pause) (T). At point C? (pause) (T). Since B and C are on the same contour line, they are at the same elevation.

What would the elevation be for contour lines D and E? (pause). The correct response is 760 feet (T) for point D and 820 feet (T) for point E.

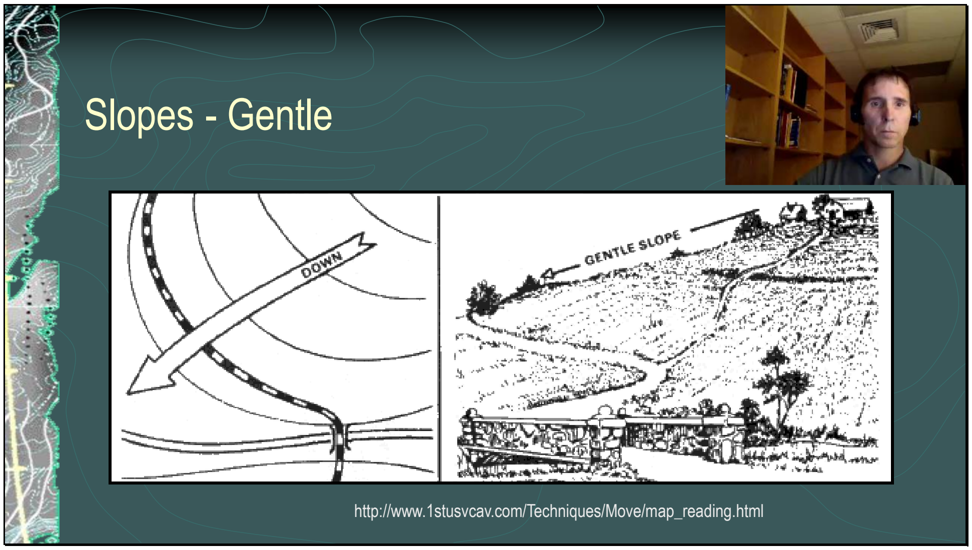
Slide 7



Now we'll turn to how to recognize landscape features based on the shape of contour lines.

We'll examine these common landforms by comparing a landscape view to how they are represented with topographic contour lines

Slide 8

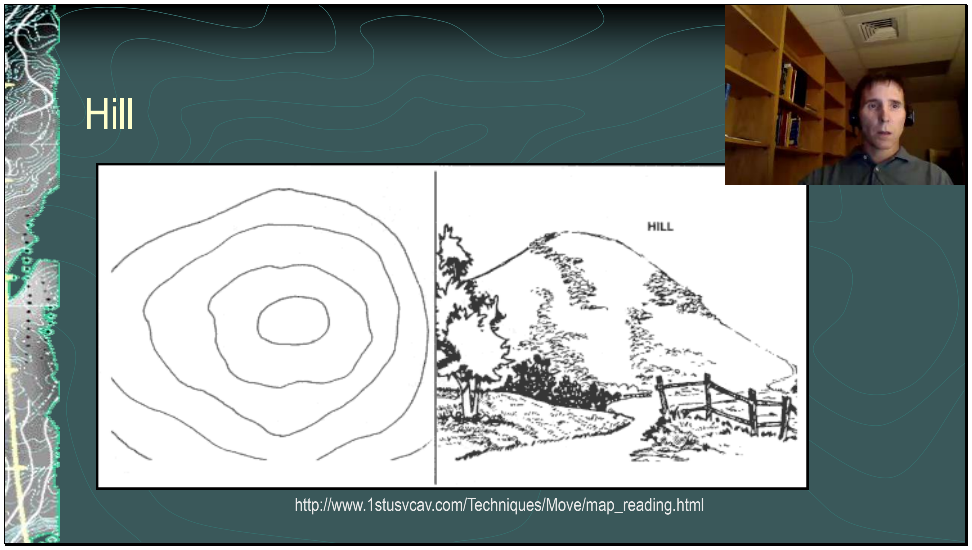


Let's begin by looking at the steepness of slopes. Examine the landscape illustration on the right and compare it to how it would appear in the topographic map on the left. Gentle slopes have widely spaced contour lines

Slide 9

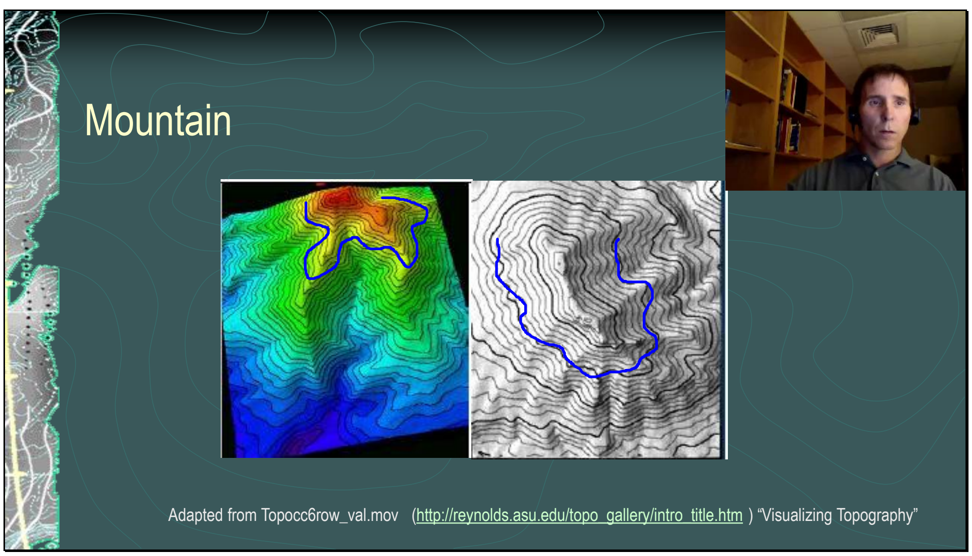
Steep slopes are the exact opposite of gentle slopes; they have closely-spaced contour lines.

Slide 10



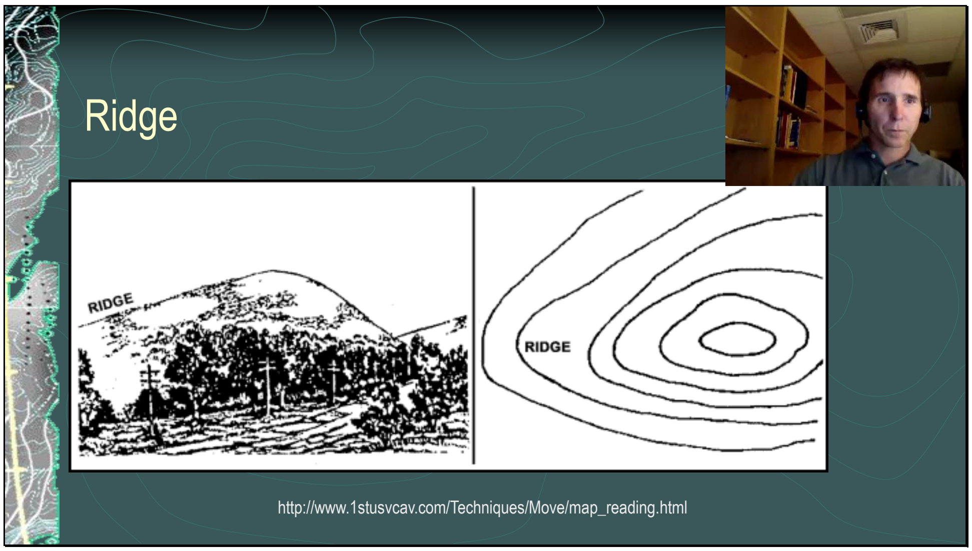
Hills and Mountains often have similar shapes. For our purposes, we define a hill as a rounded feature, with regularly spaced contour lines as shown. A hill is lower elevation than a mountain.

Slide 11



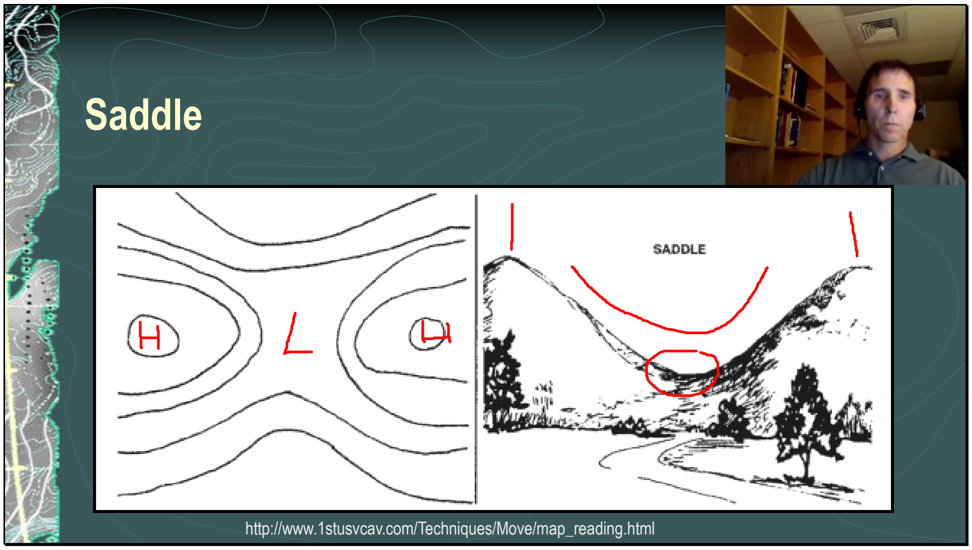
Mountains are always higher in elevation than hills, have steeper slopes, and are complex in shape. Mountains can have multiple landscape features such as cliffs, valleys, plateaus, etc. Click on the image on the left to view the mountain from different 3D perspectives. Compare it to the plan view on the right.

Slide 12



Ridges are similar to hills and mountains but their defining characteristic is that they are elongated, often forming a line of elevated terrain. Ridges can be at low elevations and may connect hills or at high elevations and connect mountains.

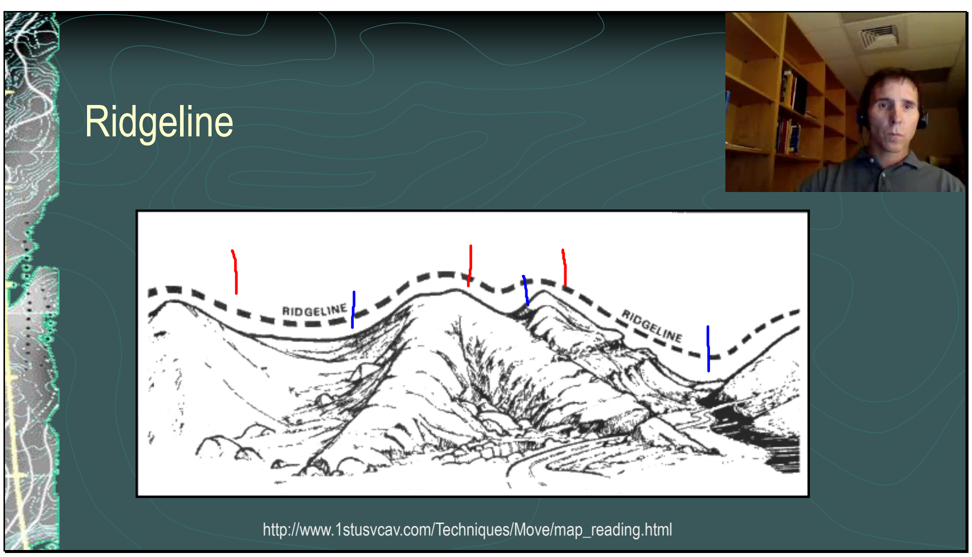
Slide 13



A Saddle is a ***low point separating two high points, usually a ridge top or two mountain tops***
Study the landscape and topographic map profile carefully. Notice the characteristic 'saddle shape' with a low elevation area separated by two high points.

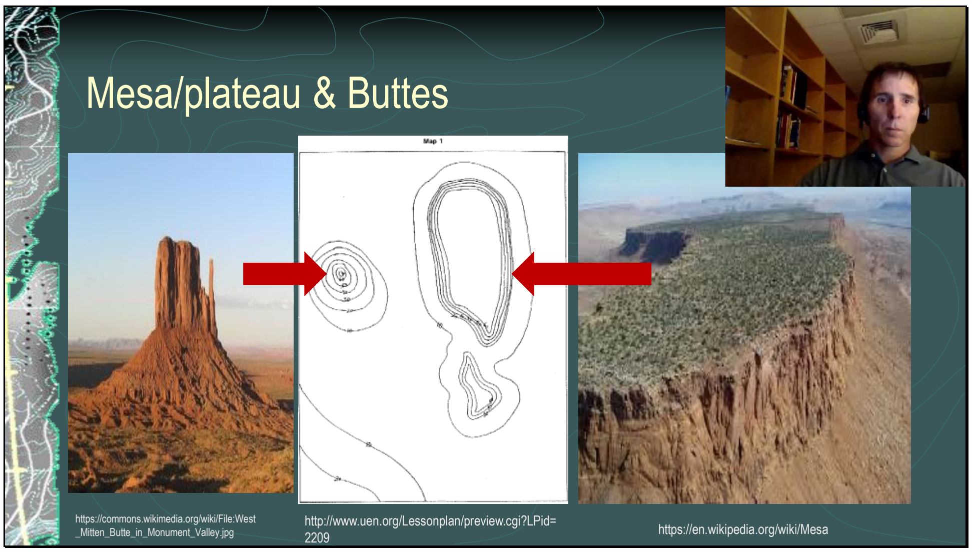
Many ridges contain at least one saddle.

Slide 14



A ridgeline consists of multiple ridges connected together. Notice the high points (these could be hilltops, if at a low elevation, or mountain peaks, if high elevation) separated by lower points, each of which is a saddle.

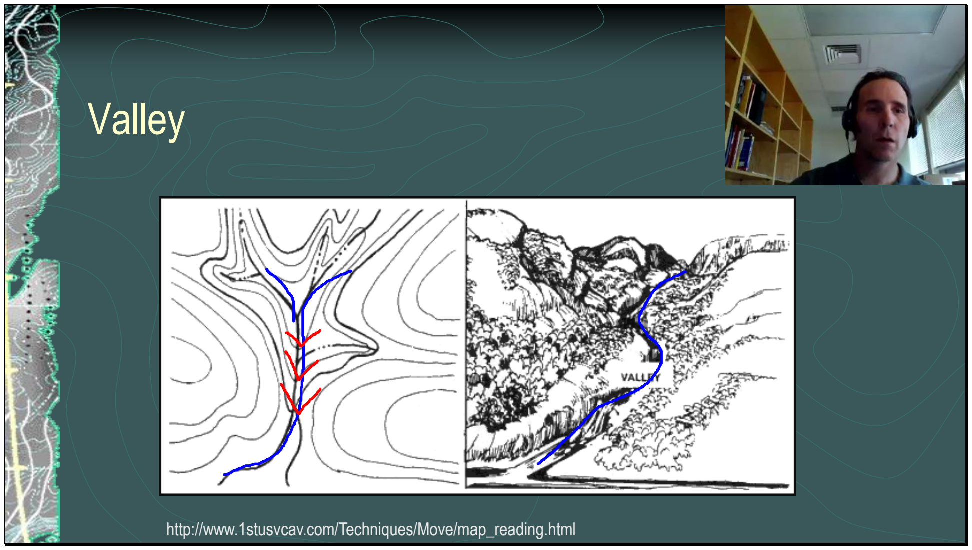
Slide 15



An image of a mesa is shown on the right. Mesa are essentially flat-topped hills. The flat area on top can also be referred to as a plateau. The key defining feature of mesas is that it is surrounded by very steep slopes, usually cliffs. You will notice the close spacing of topographic lines at the edges of the mesa.

An example of a butte is shown in the left image. A butte is essentially the highly-eroded remnant of a mesa. It is also surrounded by cliffs but has a very small plateau at the top. In this example, it is barely visible. Compare the similar shapes yet different sizes of a mesa and butte in the topographic examples to the accompanying photos.

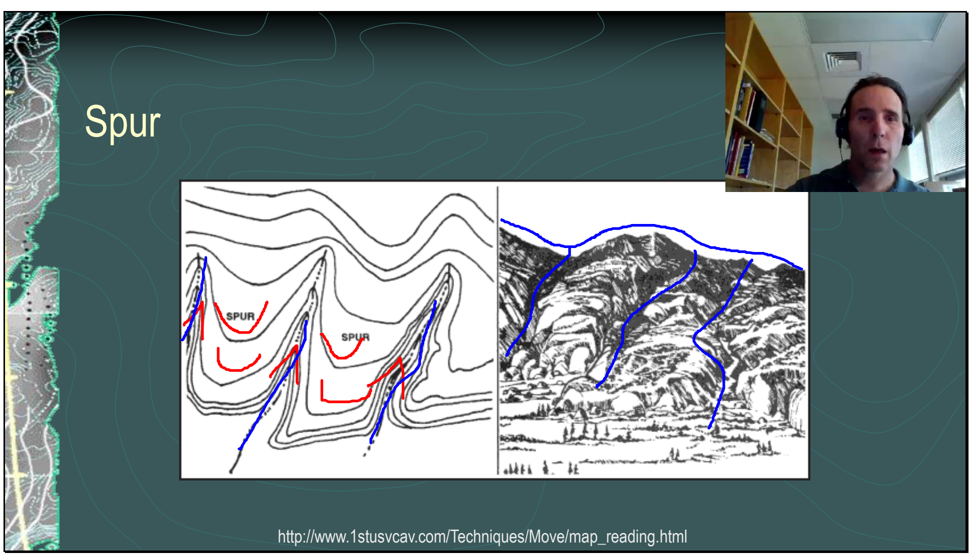
Slide 16



Now let's turn to landscape features that are created by erosion.

Valleys are recognized by a v-shaped pattern, as shown above. The direction of water flowing downhill/downstream is opposite the way the v-points, as shown here.

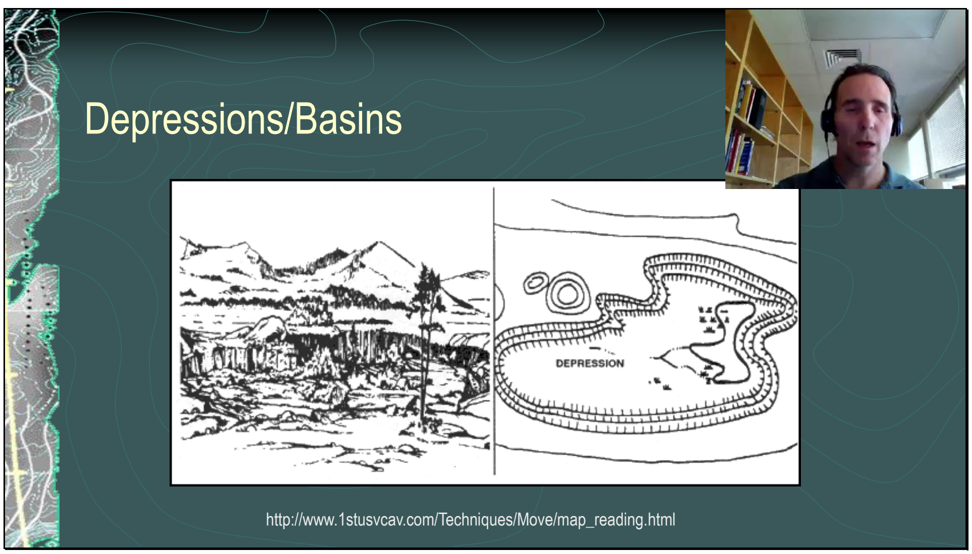
Slide 17



A spur is a combination of two features: a valley and a ridge. Notice the pattern of the topographic lines and compare them to the landscape view. What do you notice? Spurs and valleys occur in a repeating fashion. They are common features of ridges and mountains. The topographic lines associated with the valleys point uphill, as discussed in the previous slide, so water will flow downstream in the opposite direction of the V's. The topographic lines that indicates spurs point in the opposite direction of valleys, they point downhill. Also note that running water tends to form steep valleys between the more gentle, rounded slopes of spurs.

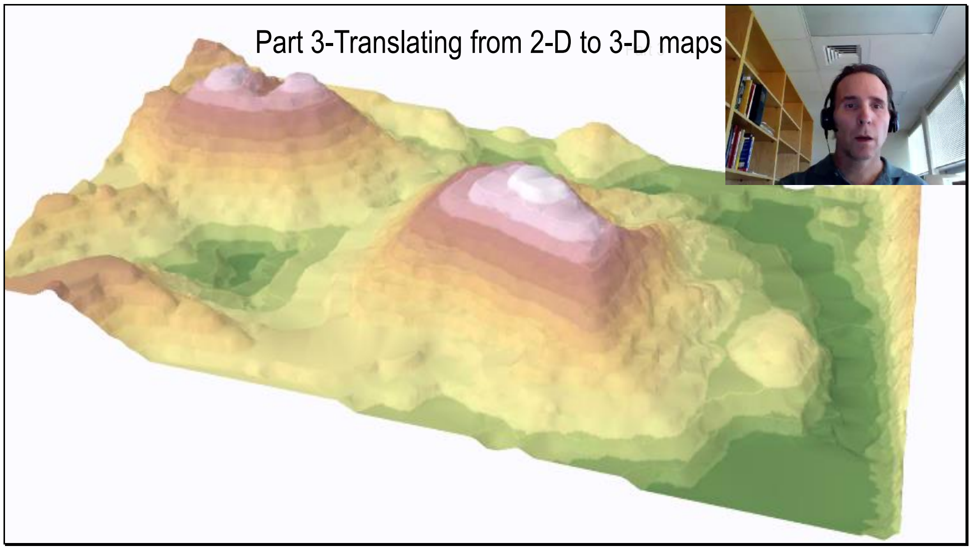
A useful tip to remembering the difference between a spur and a valley is that valleys have an upslope pointing V-shape and spurs a downslope pointing u-shape

Slide 18



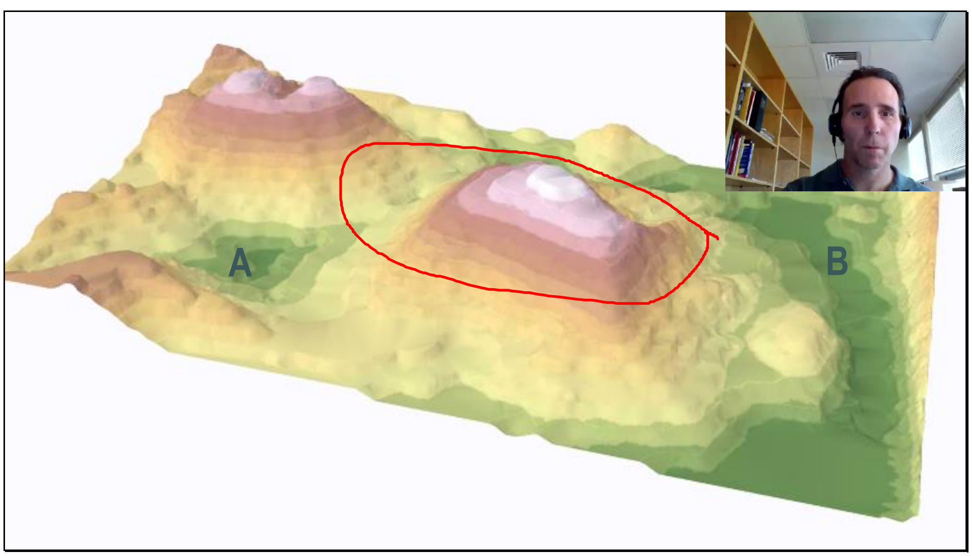
Depressions or basins are areas that are below the average level of the landscape. These are locations where water tends to collect from rainfall or streamflow. Hachure marks, as shown in the topographic example may be used to indicate a depression. In the maps we will use color as an aid to determining relative elevation; darker colors are associated with higher altitudes.

Slide 19



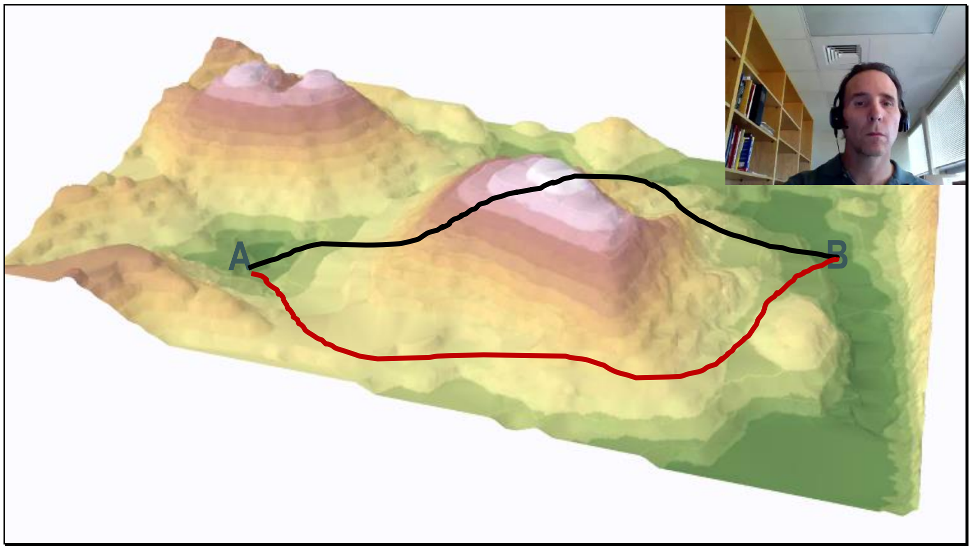
Now we're going to visualize the landscape features we've identified using this imaginary 3-D oblique view landscape. Note that color is used to show relative elevation. Green is the lowest elevation and white is the highest in this example.

Slide 20



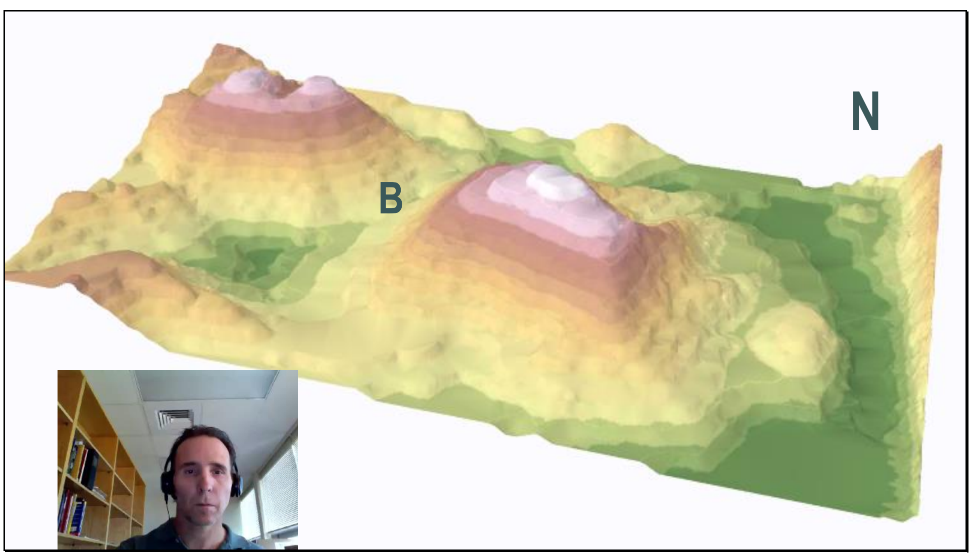
(T) If you were standing at point A, would you be able to see point B? ***Answer****: No*

Slide 21



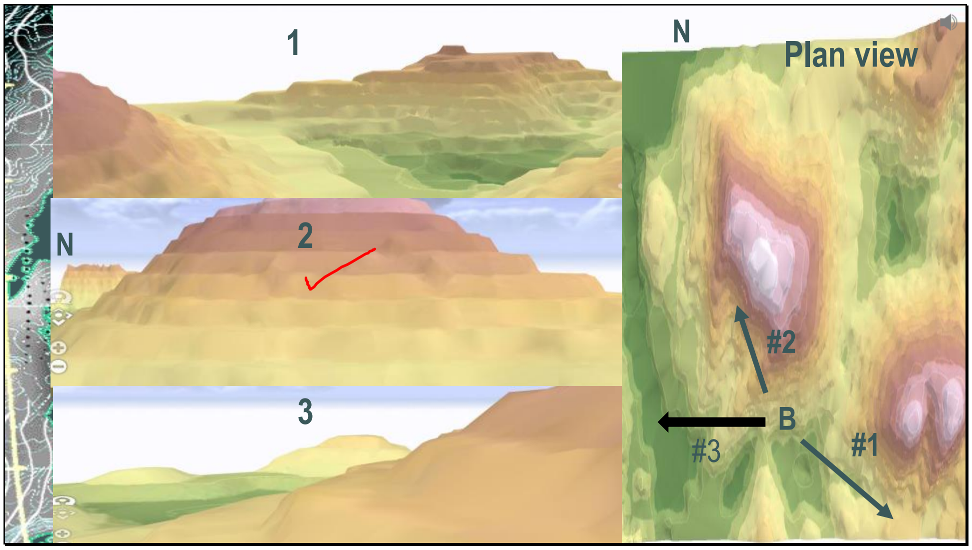
The route marked in red is easier, even though it is longer because you travel on the lower angle slopes on the side of the mountain, as opposed to the black route, which is shorter, but more strenuous climbing up and over the summit of the mountain

Slide 22



North is shown in the upper left hand corner of this map. Which of the views on the next slide would have you looking north from the saddle at Point B?

Slide 23



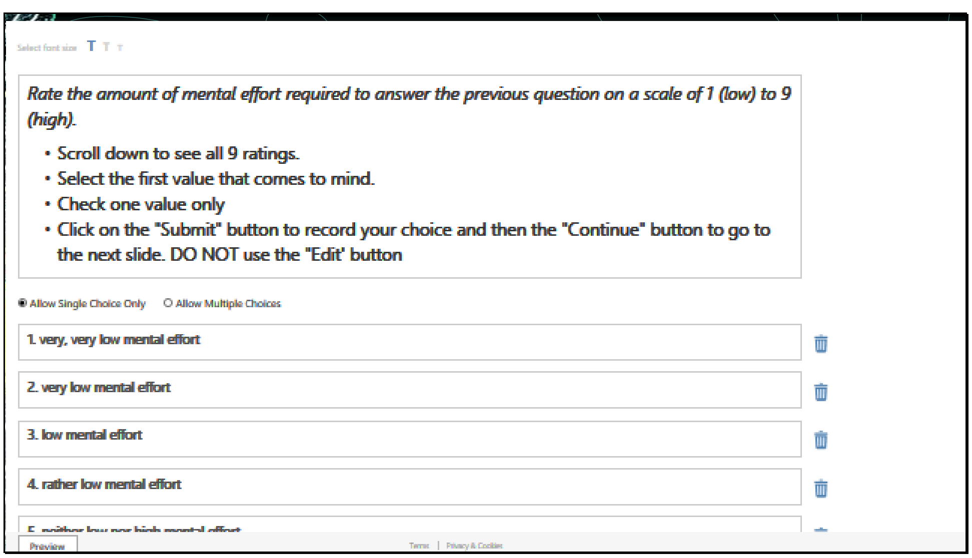
Which view has you looking N if you're standing at point B?

2 is the correct answer (T check mark and arrow)

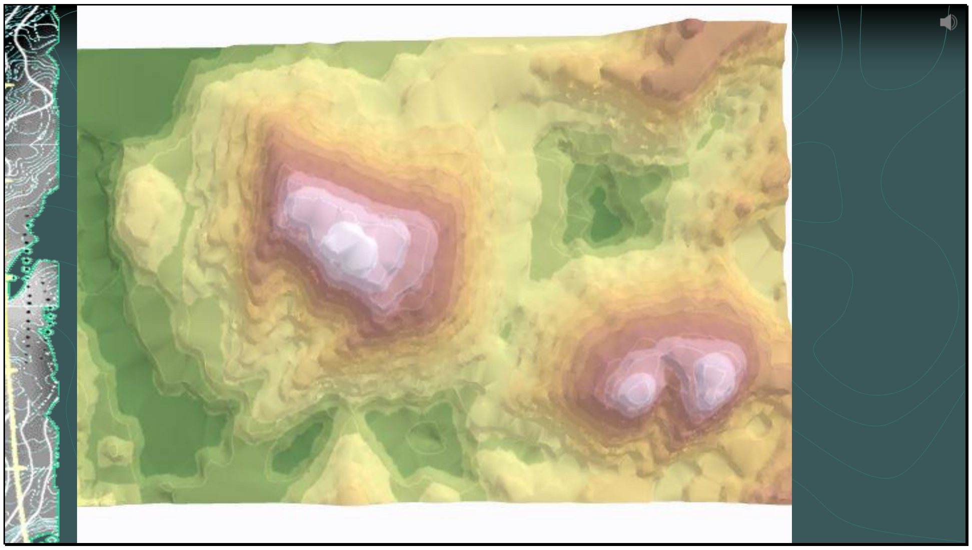
1 is looking South (T)

3 is facing W

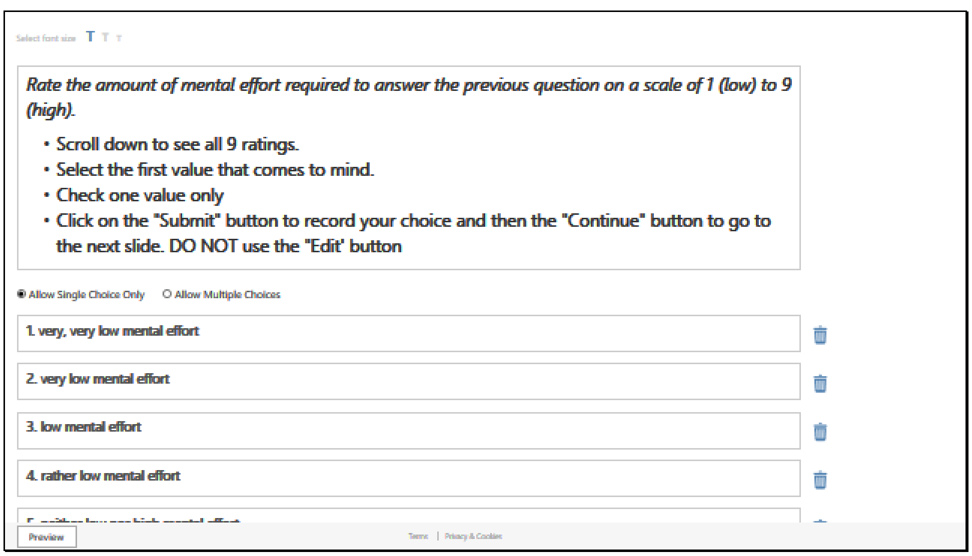
Slide 24



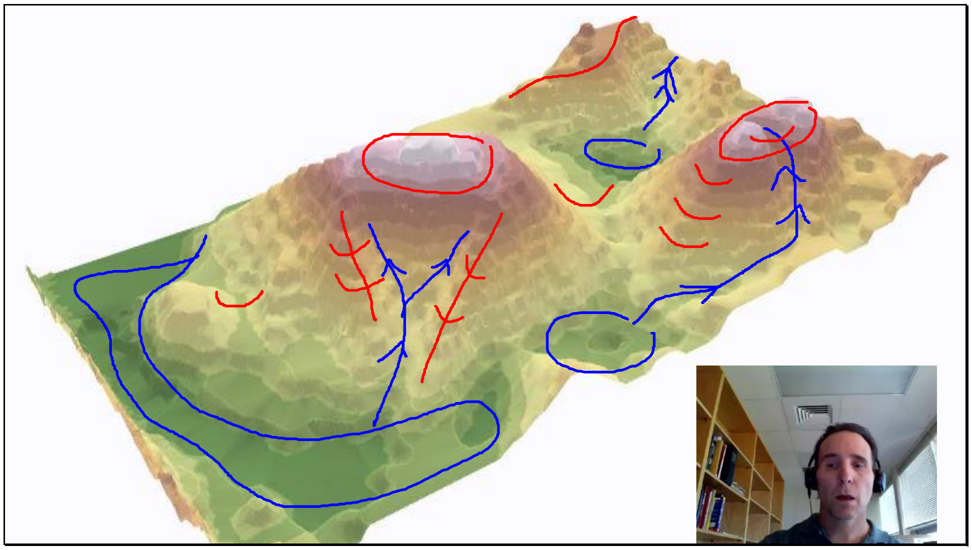
Slide 25



Slide 26

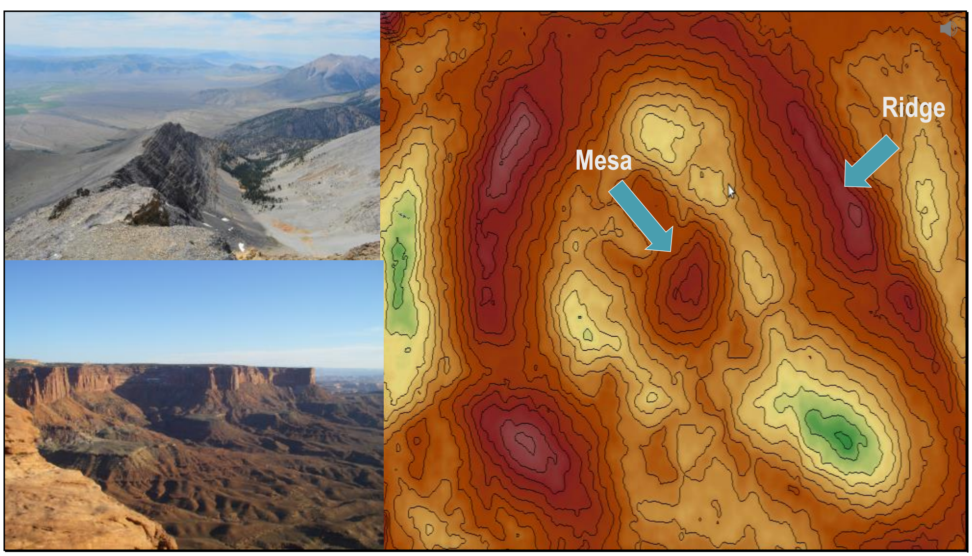


Slide 27

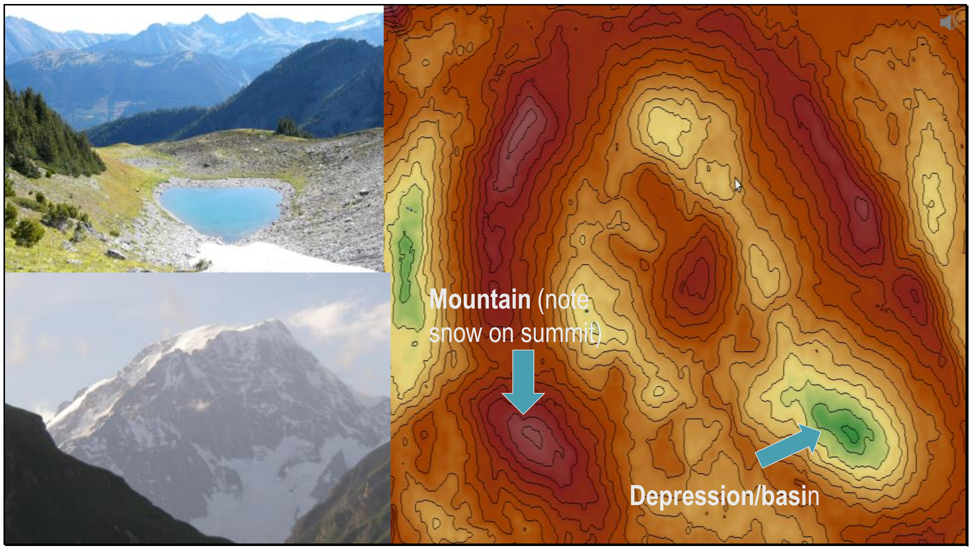


Let's visualize at least one example of each of the landscapes that we covered in part 2, on this map. In the next slide you will identify them from a 2-D plan view perspective.

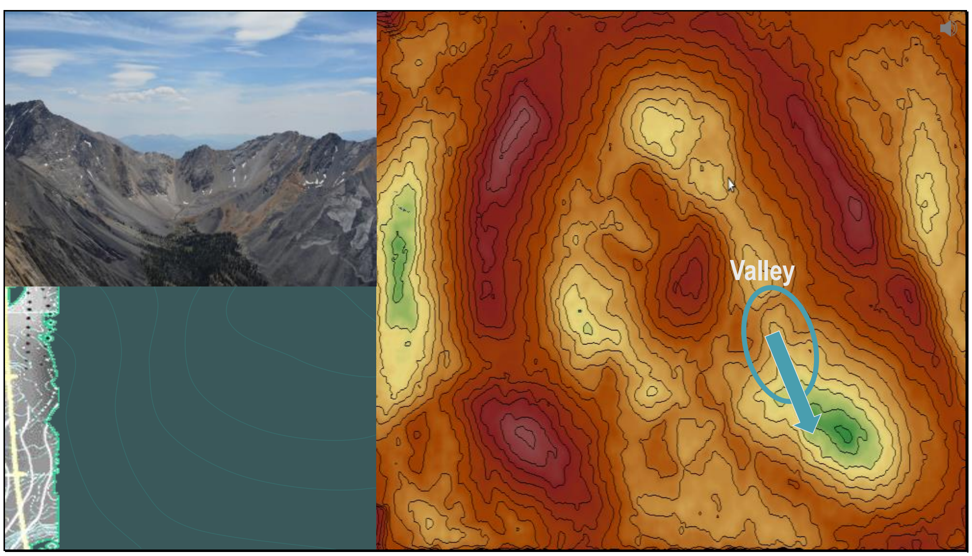
Slide 28



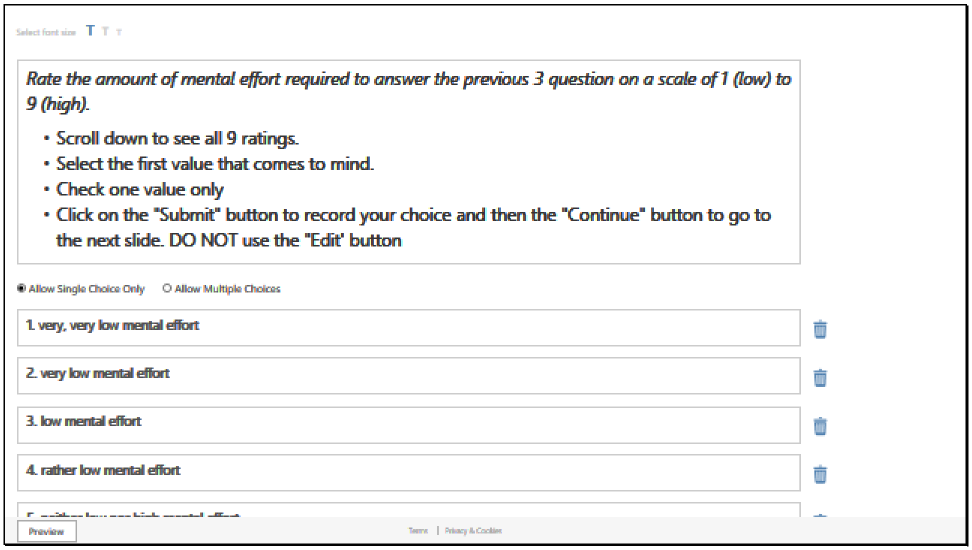
Slide 29



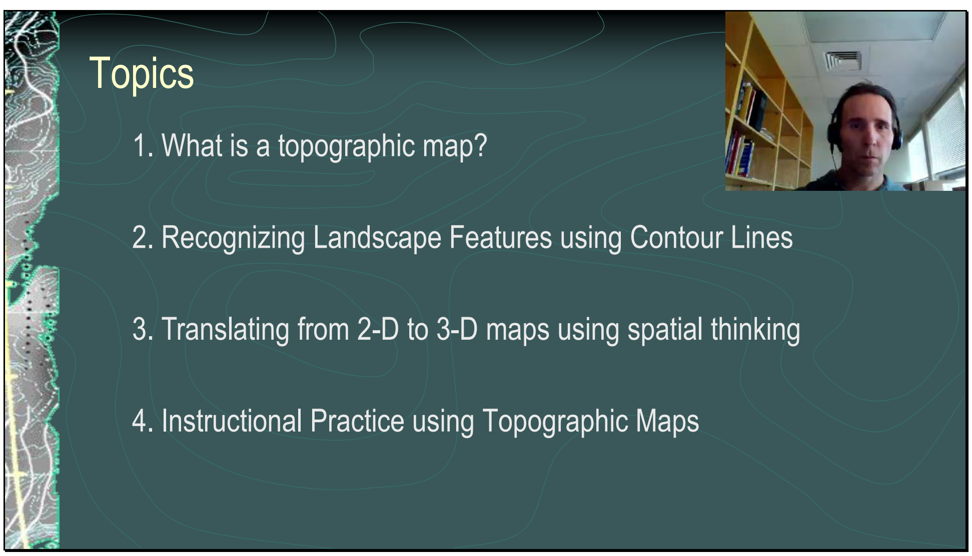
Slide 30



Slide 31



Slide 32



This concludes the instructional video. We will now spend time interacting with one of two types of maps where we will apply the knowledge we have gained. We hope you have enjoyed this experience in learning how to read topographic maps. Thank you for participating in this experiment.

Appendix I

Learning Practice Activity Questions

Learning Tasks- ARS Map & 2D Topo Map groups Participant id# ______

Protocols

ANNOUNCE TO BOTH GROUPS. *"Your map uses colors to indicate relative elevation. Water and snow feature on certain parts of the map"*

ARS group (Tx. 1)

- *Use the interaction space to* BUILD *topographic features (show location and demo)*
- *Drag a finger across the sand to trace a route*
- *Touch the sand with a finger to mark a point*
- *Demonstrate raincloud feature and then use 'D' key to dry landscape.*
- *Participant may use rainfall feature at any time. Add a note next to any tasks when they use it.*
- *Use 'D' key frequently to keep the landscape dry.*

2D Top map group (Tx.2)

- *Use paper provided to DRAW topographic lines (keep paper on map to align with map calibration)*
- *Draw/trace directly on the map using dry erase markers. Erase markings after each task*

1. **Touch/label the highest point on the map** ____/1

2. **Touch/label the lowest point on the map** ____/1

a. (**MEMS**___________)

3. **Build/draw topographic lines to show a gentle slope with the direction from higher to lower elevation** _________(Correct/incorrect)

a. **Trace at least 2 areas on the map with gentle slopes** ____ /1

b. (**MEMS**__________)

4. **Build/draw topographic lines to show <u>a steep slope with the direction from higher to lower elevation</u>** _________(Correct/incorrect)

a. **Trace at least 2 areas on the map with steep slopes** ____/1

b. (**MEMS**__________)

5. **If the elevation at point A1 is 300', and the contour interval is 50', what is the elevation at point A2?** (150') ___/1 (***NOTE****: values may change slightly on rebuilds of ARS map so re-calculate and score as necessary)*

a. (**MEMS** ________)

6. **At point A3?** (700')____/1 *(may need to recalculate with ARS)*

a. **Which point is higher, point A3 or point A4?(**A4) ___/1

b. (**MEMS** ________)

7. **Build/draw topographic lines to show a <u>hill</u>** _________(Correct/incorrect)

a. **Touch/label any hills on the map** ______ /1

b. **(MEMS**__________)

8. **Build/draw topographic lines to show a <u>mountain</u>**

_________(Correct/incorrect)

a. **Touch/label any mountains on the map** ______ /1

b. (**MEMS**__________)

9. **Build/draw a <u>mesa</u>** ________(Correct/incorrect)

a. **Touch/label any mesas on the map** ______ /1

b. **Touch/label any <u>buttes</u> on the map** ____ (**there are none**)

c. (**MEMS**__________)

10. **Build/draw topographic lines to show a <u>saddle</u>** _________(Correct/incorrect)

a. **Touch/label any saddles on the map** ______ /1

b. (**MEMS**__________)

11. **Build/draw topographic lines to show a <u>ridge/ridgeline</u>**

_______(Correct/incorrect)

a. **Trace/draw a line to show the path of the ridge on the map** ______ /1

b. (**MEMS**__________)

12. Build/draw topographic lines to show a <u>valley</u>

a. **Trace/draw a line to show the <u>path of at least 2 valleys</u>** ______ /1

b. (**MEMS**__________)

13. **How many <u>basins/depressions</u> can you touch/label on the map?** ___/1

a. (**MEMS**__________)

14. **Build/draw topographic lines to show <u>2 or more spurs</u>**

________(Correct/incorrect)

a. **Touch/label any spurs on the map(must explain which are valleys and which are spurs)** ______ /2

b. (**MEMS**__________)

15. **Can you see Point B from Point A?** NO ____/1

a. (**MEMS**__________)

16. **Can you see Point D from Point C?** YES ____/1

a. (**MEMS**__________)

17. **From Point E, can you see the lake?** NO ___/1

a. (**MEMS__________**)

18. **If you wanted to walk from point 1 to point 2, what is the easiest, lowest angle route you would take? Trace/draw your route on the map**. ___/2

a. **(MEMS__________**)

19. **If you wanted to walk from point 3 to point 2, what is the easiest, lowest angle route you would take? Trace/draw your route on the map**. ___/2

a. (**MEMS__________**)

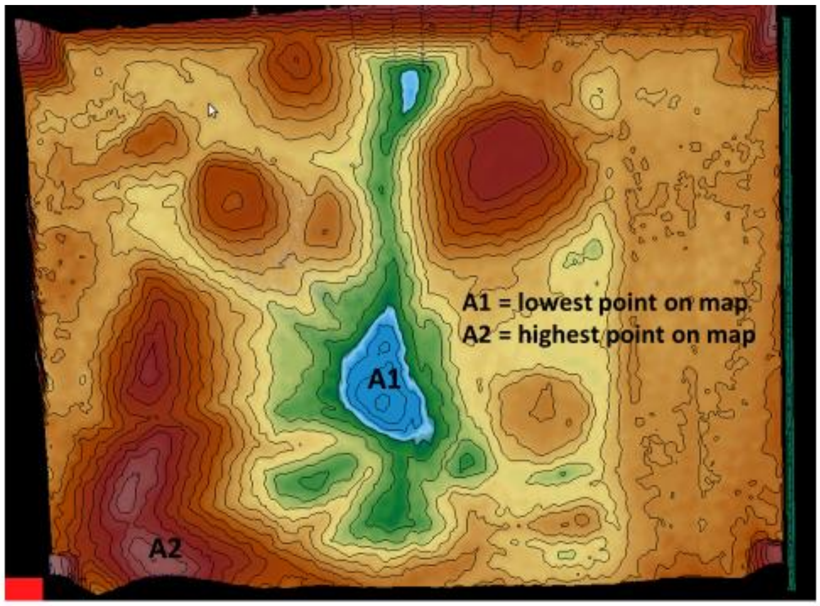
20. **If it was to rain at points A, B, C, D & E, trace/draw the route the water would flow** ___/5 (ARS may use rainfall feature. Record any participants who do not use it)

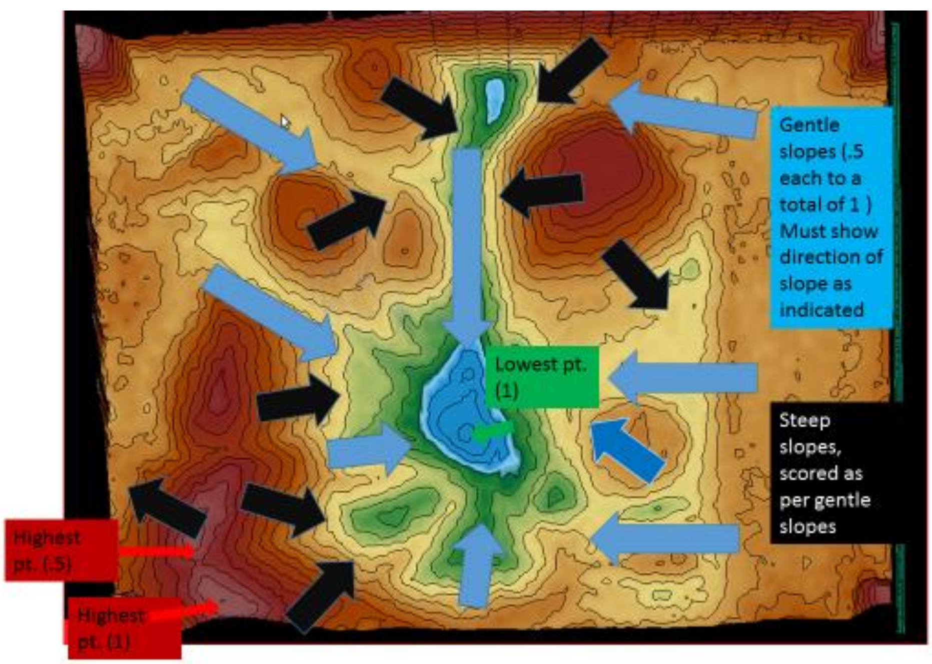
a. (**MEMS__________**)

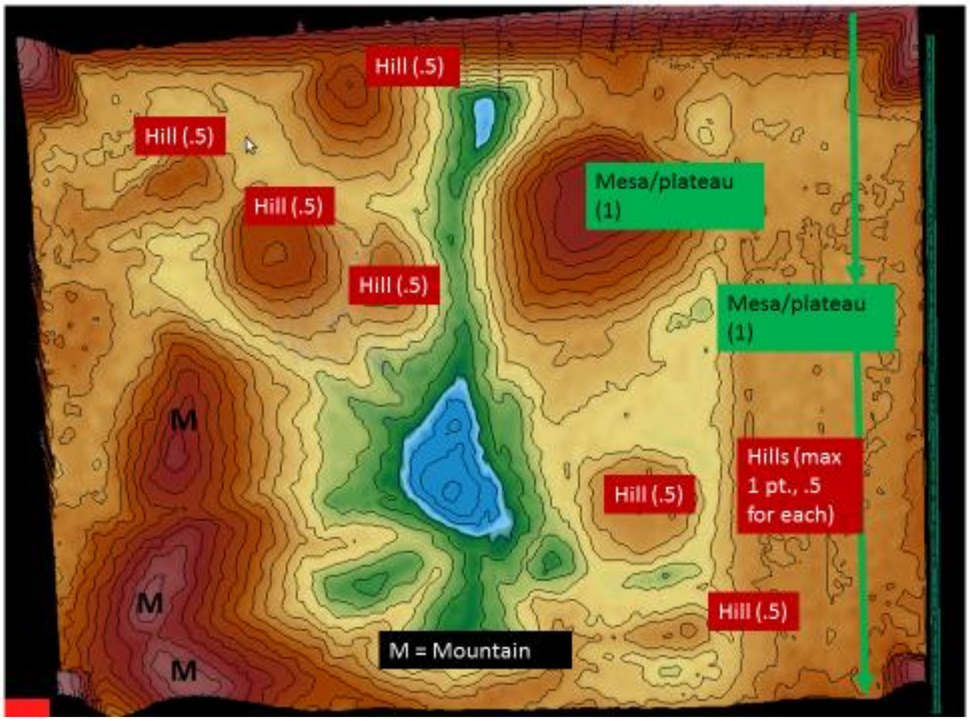
TOTAL _____/28

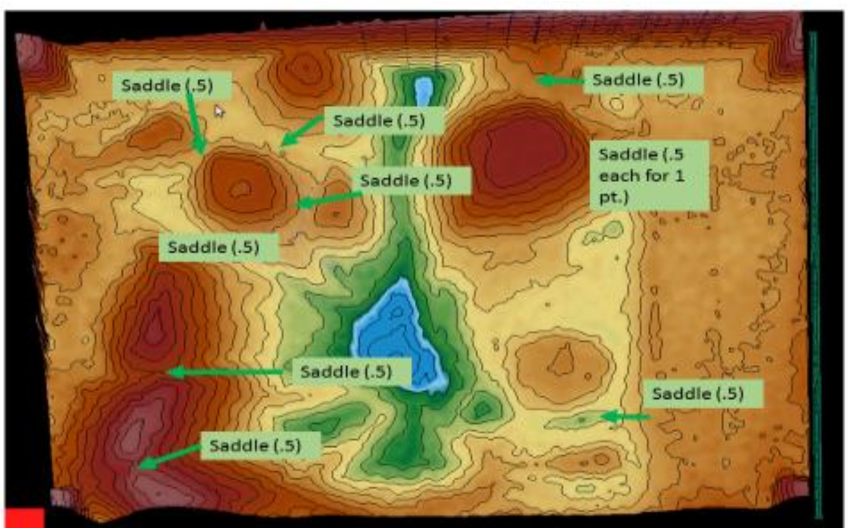
Appendix J

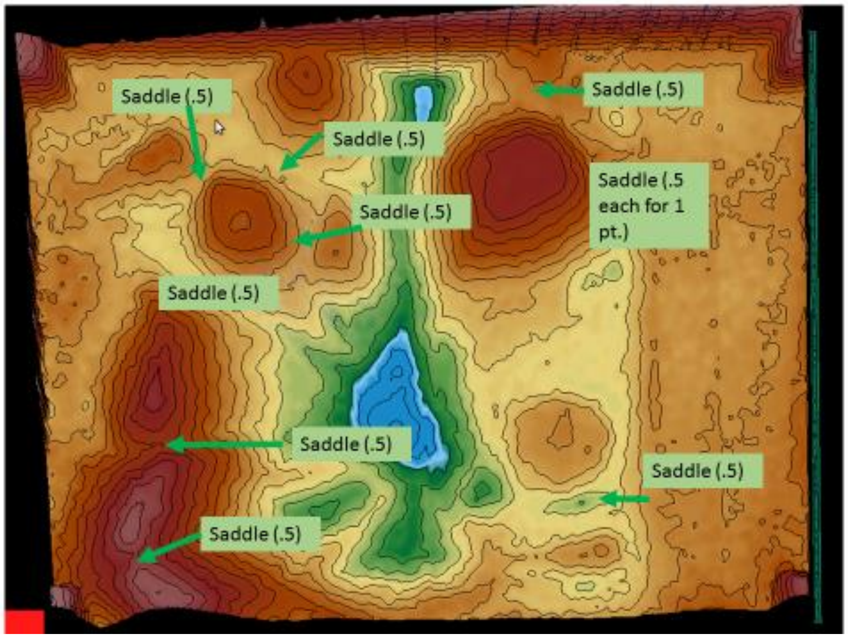
Learning Practice Activity Answer Key

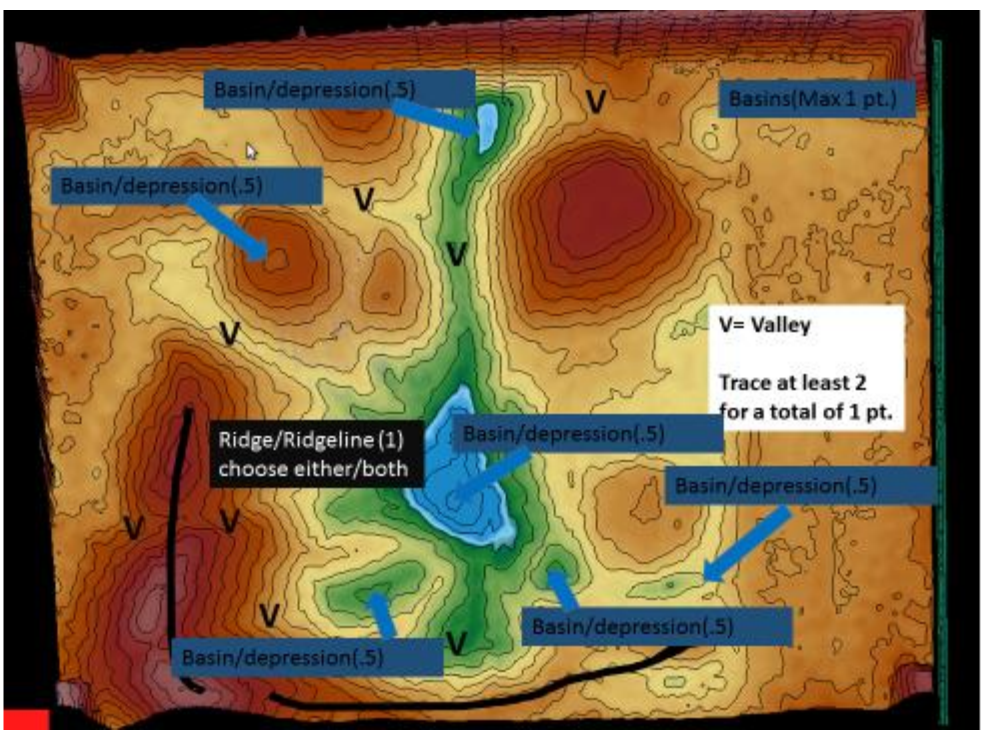


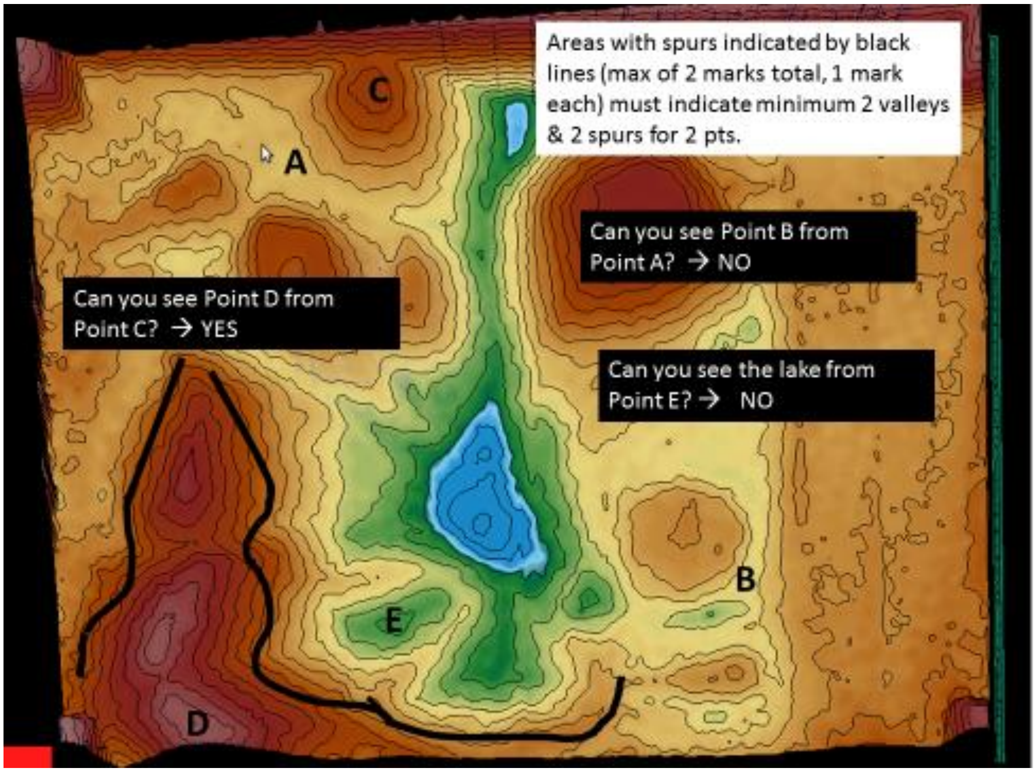


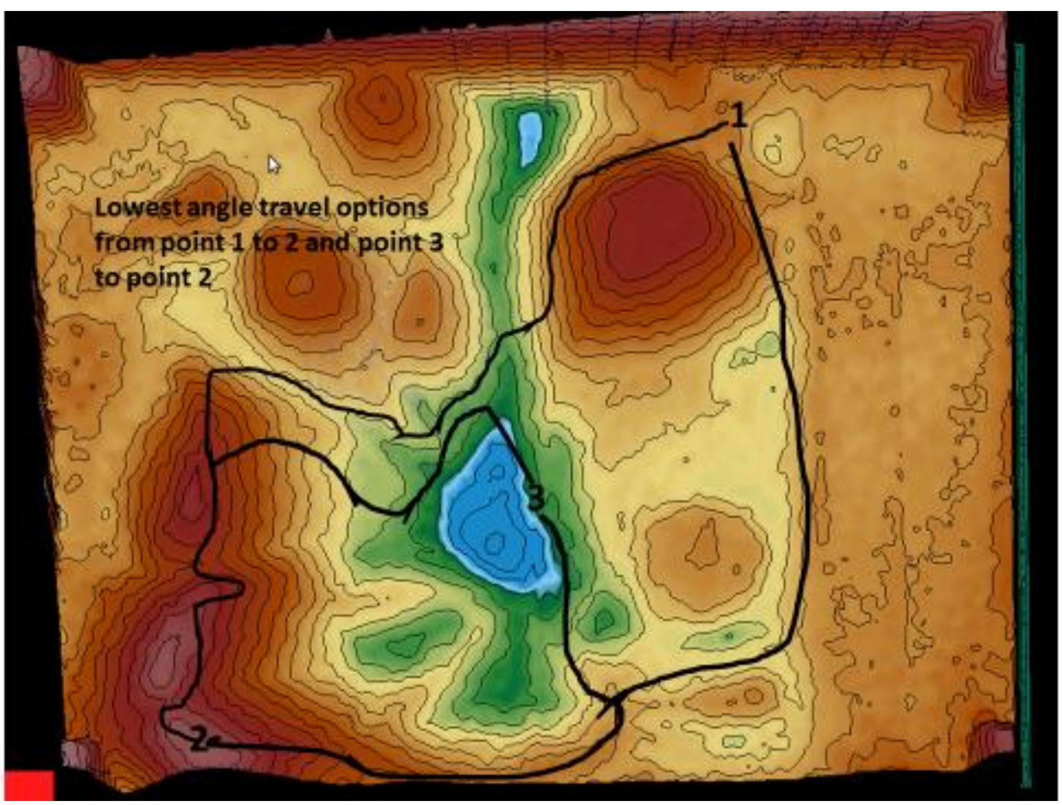

Hill (.5)
Hill (.5)
Hill (.5)
Hill (.5)
Mesa/plateau (1)
Mesa/plateau (1)
M
M
M
Hills (max 1 pt., .5 for each)
Hill (.5)
Hill (.5)
M = Mountain

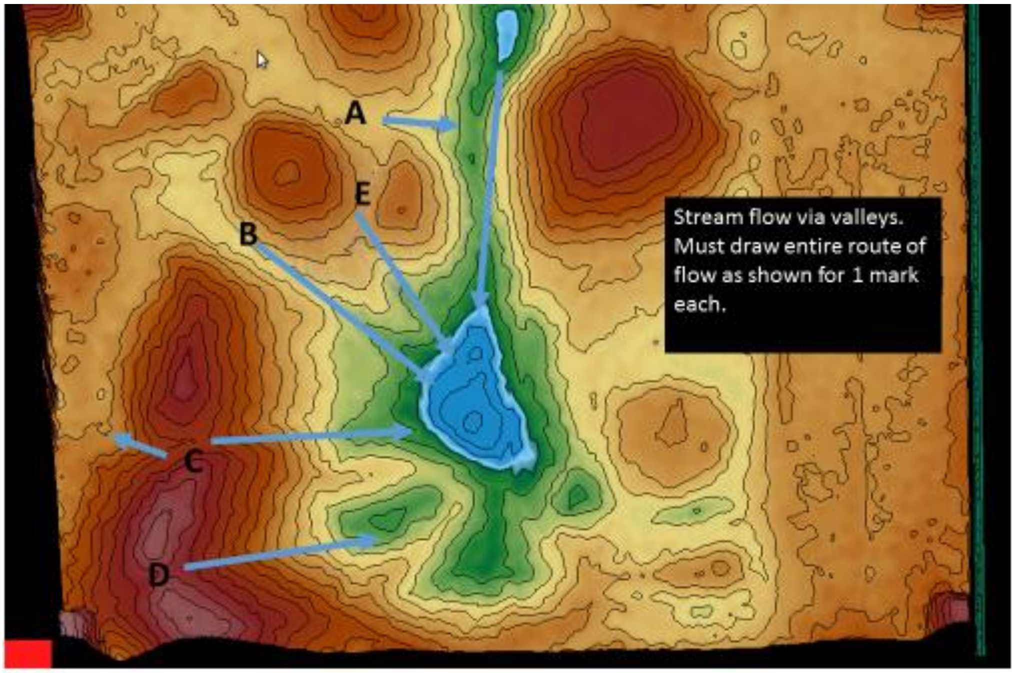

Saddle (.5)
Saddle (.5)
Saddle (.5)
Saddle (.5 each for 1 pt.)
Saddle (.5)
Saddle (.5)
Saddle (.5)
Saddle (.5)
Saddle (.5)


Saddle (.5)
Saddle (.5)
Saddle (.5)
Saddle (.5
each for 1
pt.)
Saddle (.5)
Saddle (.5)
Saddle (.5)
Saddle (.5)
Saddle (.5)


Basin/depression(.5)
Basins(Max 1 pt.)
Basin/depression(.5)
V
V
V
V
V= Valley
Trace at least 2
for a total of 1 pt.
Ridge/Ridgeline (1)
choose either/both
Basin/depression(.5)
Basin/depression(.5)
V
V
V
V
Basin/depression(.5)
Basin/depression(.5)


Areas with spurs indicated by black lines (max of 2 marks total, 1 mark each) must indicate minimum 2 valleys & 2 spurs for 2 pts.
C
A
Can you see Point B from Point A? → NO
Can you see Point D from Point C? → YES
Can you see the lake from Point E? → NO
B
E
D


1
Lowest angle travel options from point 1 to 2 and point 3 to point 2
3
2


A
E
B
Stream flow via valleys. Must draw entire route of flow as shown for 1 mark each.
C
D

Appendix K

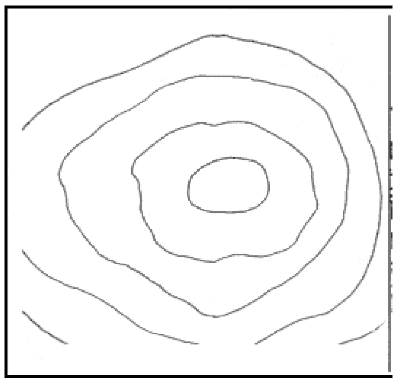
mTMA posttest

The TMA posttest assessment (Newcombe et al., 2015) comprised 18 questions. To correspond with the 17 learning outcomes for novice level topographic map reading competency identified during the instructional design process (particularly outcomes seven and eight, which the TMA does not assess), two additional pages were created by the researcher and added to the beginning of the TMA to account for outcomes seven and eight. Hence the third page of the posttest delineates the beginning of the TMA. It was left in situ to preserve the fidelity of the original TMA format. Permission was sought and gained from the developers of the TMA to add these seven additional questions plus a space for participants to score their mental effort from one to nine using MEMS ratings (Paas & Van Merriënboer, 1993) after completing each question. These changes to the original TMA are reflected by re-labelling the version of the test used in this study as the modified Topographical Map Assessment (mTMA) with 25 questions.

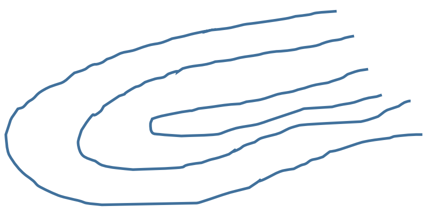
Participant id# ____________________

1. **Match the correct topographic feature to its label using the correct labels from the word bank**

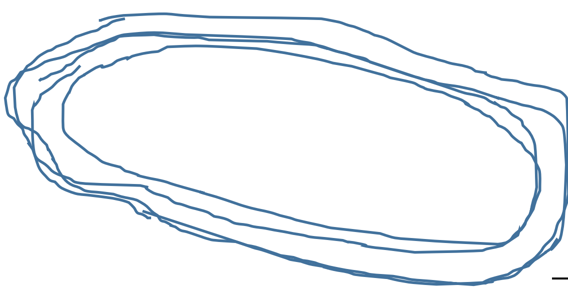
WORD BANK: mountain, hill, ridge, spur, valley, butte, spur, basin, saddle, moraine, mesa



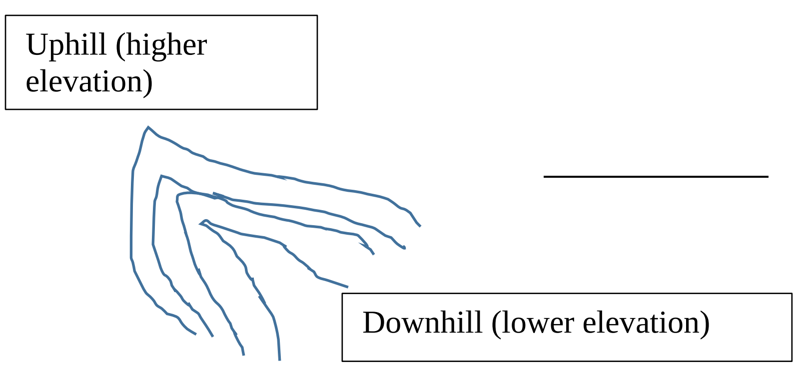
_________________ Mental Effort (1-9) _____



_________________ Mental Effort (1-9) _____



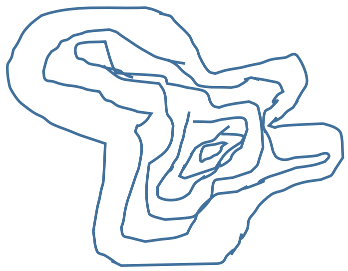
_______________ Mental Effort (1-9) _____



_______________ Mental Effort (1-9) _____

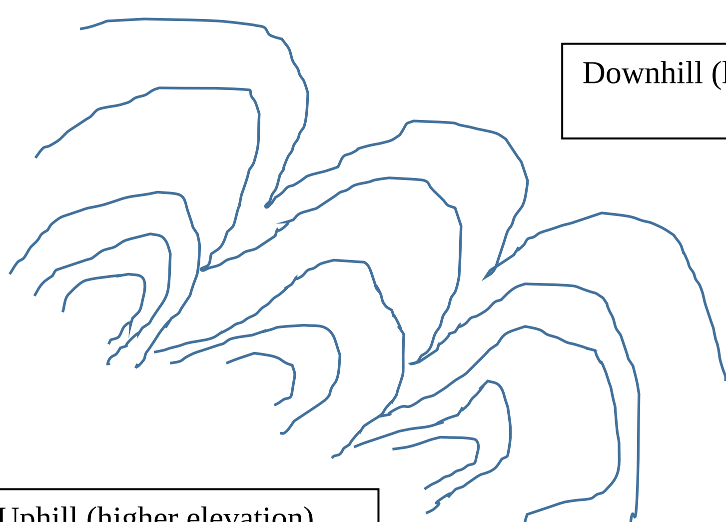
2. **Match the correct topographic feature to its label using the correct labels from the word bank**

WORD BANK: mountain, hill, ridge, spur, valley, butte, spur, basin, saddle, moraine, mesa



__________________ Mental Effort (1-9) ______

____________________ Mental Effort (1-9) ____



Downhill (lower elevation

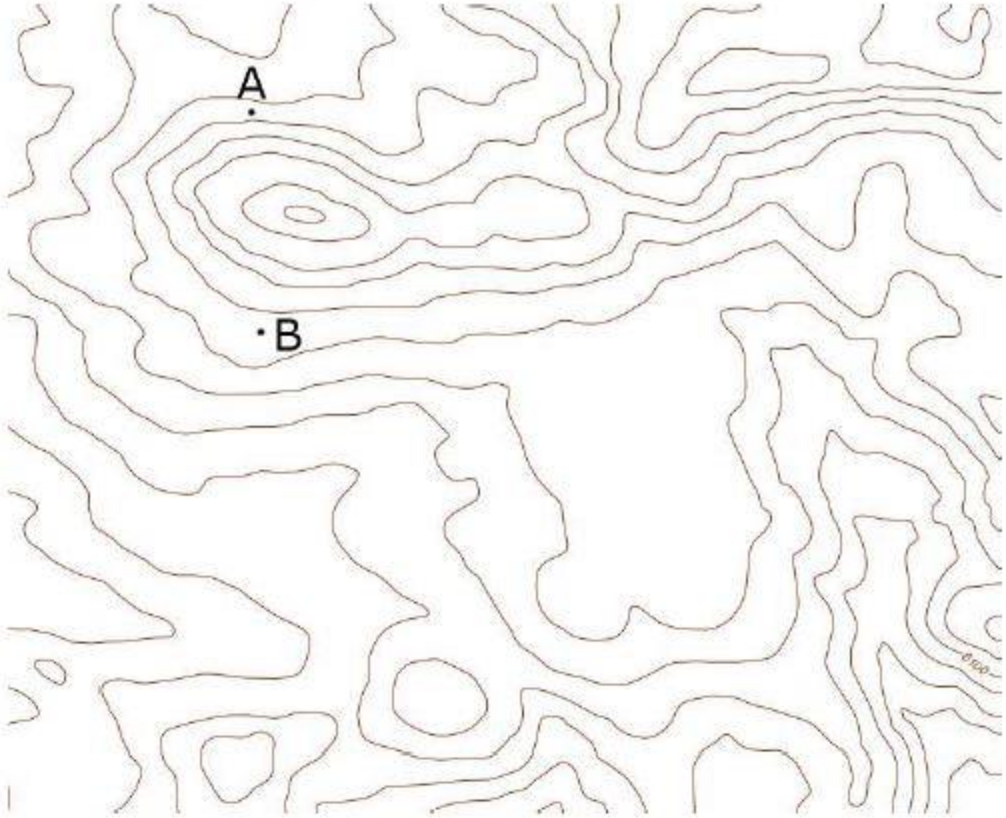
Uphill (higher elevation)

__________________ Mental Effort (1-9) _____

Topographic Map Assessment

Developed by Matt Jacovina, Carol Ormand, Thomas F. Shipley, Steven Weisberg

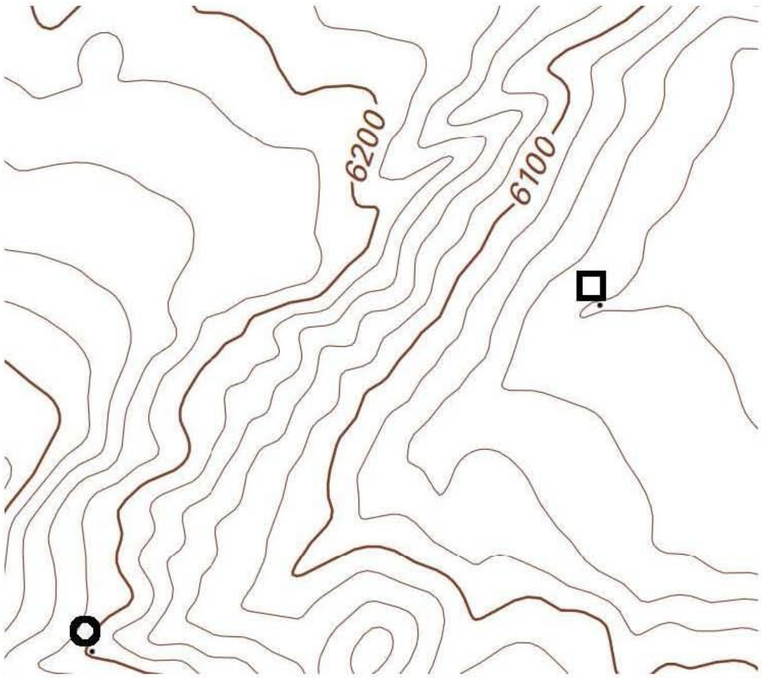
Please complete this 18-item assessment. The assessment is not timed. Try to answer each item to the best of your ability.



1. The contour interval for this map is 100 ft.

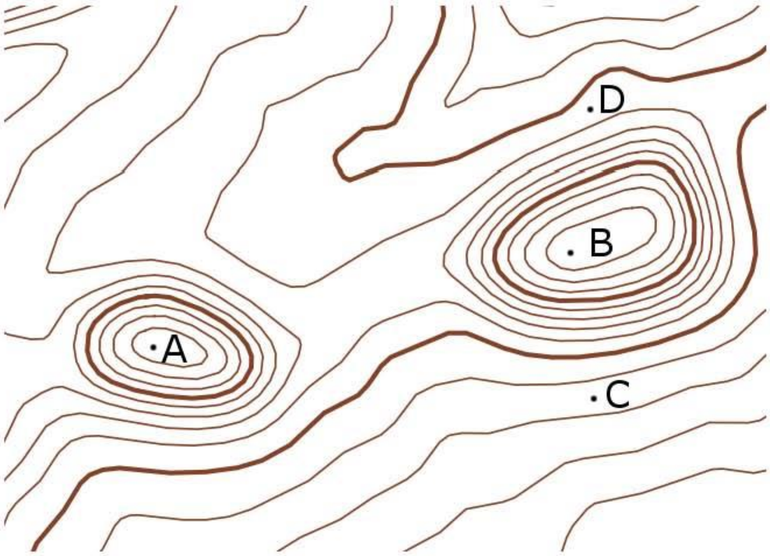
Imagine you had to walk to get from point A to point B, and wanted to do so as easily as possible. Sketch the route you would build, and explain why you chose that particular path.

Mental Effort (1-9) _________



- 2. Imagine there is a stream that connects the circle and the square. In which direction would the water flow? Why? Please draw the path the stream would take.

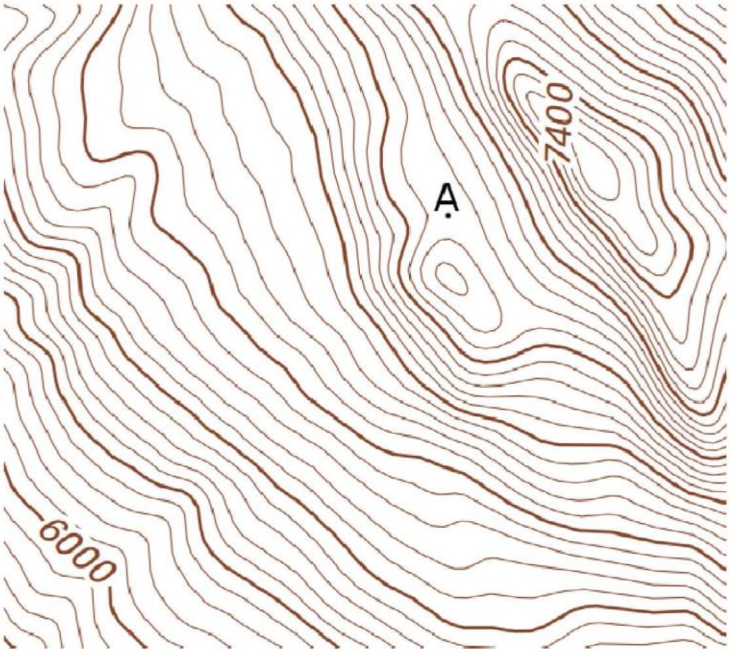
Mental Effort (1-9) _________



3. The contour interval on this map is 20 feet. One person is standing at each point on the map. Please answer (Y/N) the following questions about whether the people standing at two points can see each other. Assume they are able to use binoculars. Also assume there is no vegetation.

- A and B.
- A and D.
- B and C.
- C and D.
- B and D.

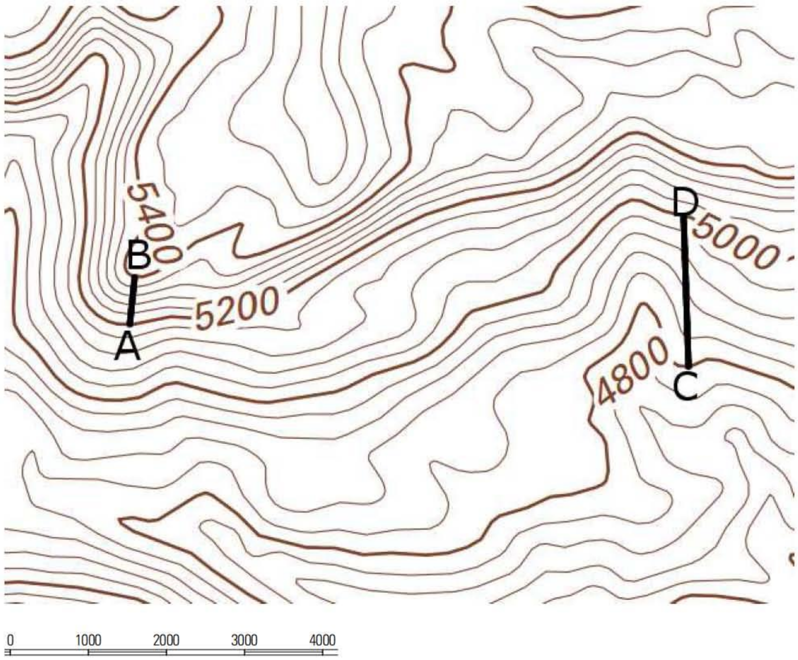
Mental Effort (1-9) _________



4. The contour interval for this map is 40 feet.

What is the elevation at point A?

Mental Effort (1-9) _________

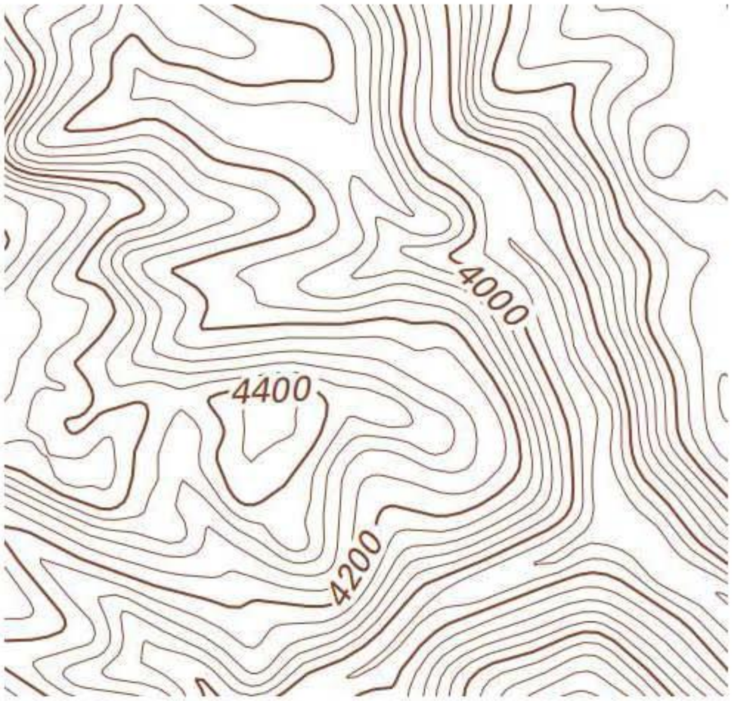


5. Imagine Josh traveled on foot from point A to point B, and Amy traveled on foot from point C to point D.

Who walked up a steeper slope?

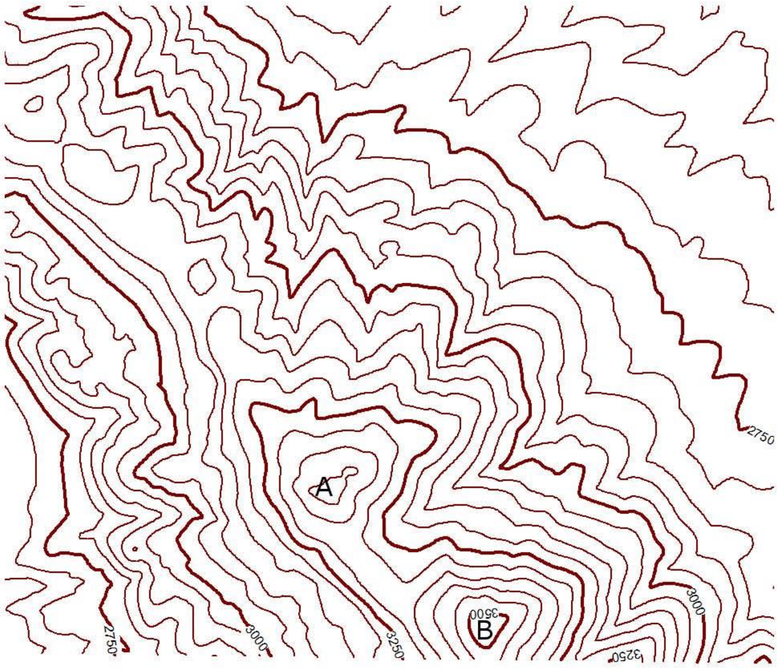
Why did you choose the answer you did?

Mental Effort (1-9) _________



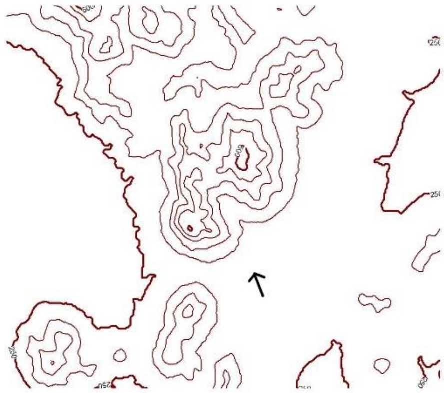
6. What is the contour interval on this map? That is, how much does elevation change moving from one line to another?

Mental Effort (1-9) ________

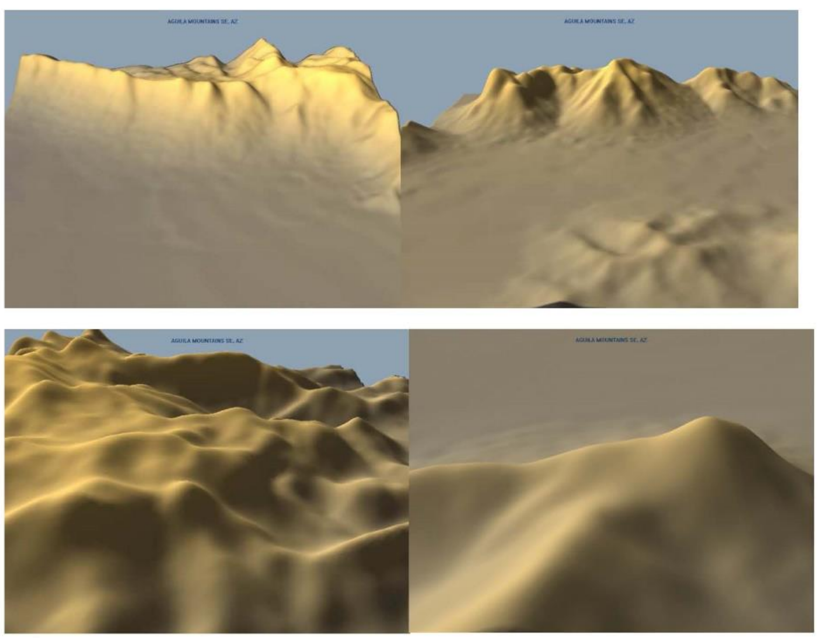


7. Which hill is higher: A or B?

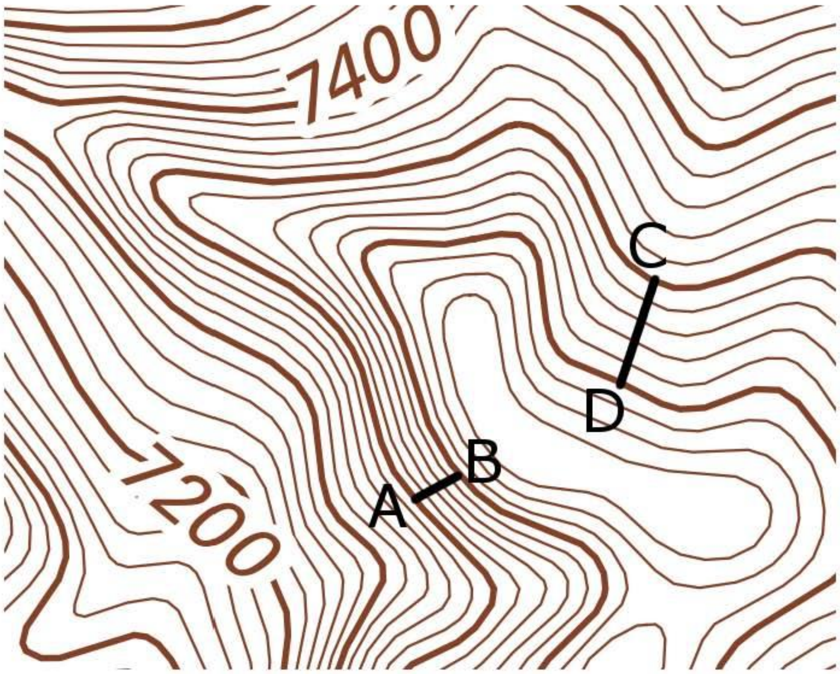
Mental Effort (1-9) _________



8. Which of the following views best represents what someone would see standing at the start of the arrow and facing in that direction?



Mental Effort (1-9) _________

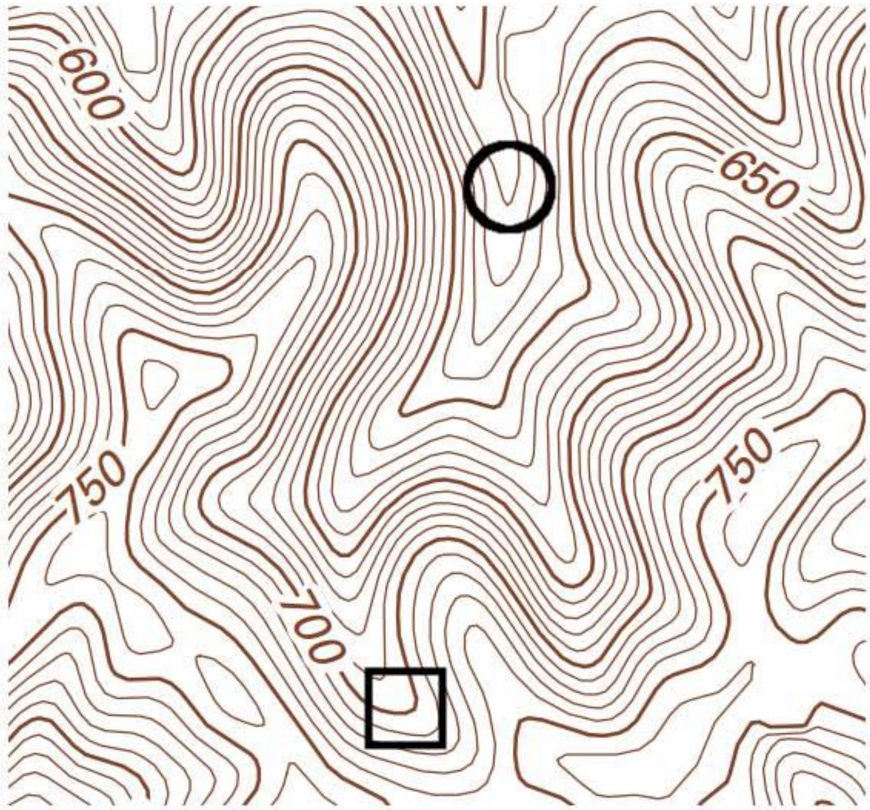


9. Imagine Josh traveled on foot from point A to point B, and Amy traveled on foot from point C to point D.

- Who walked up a steeper slope? How can you tell?

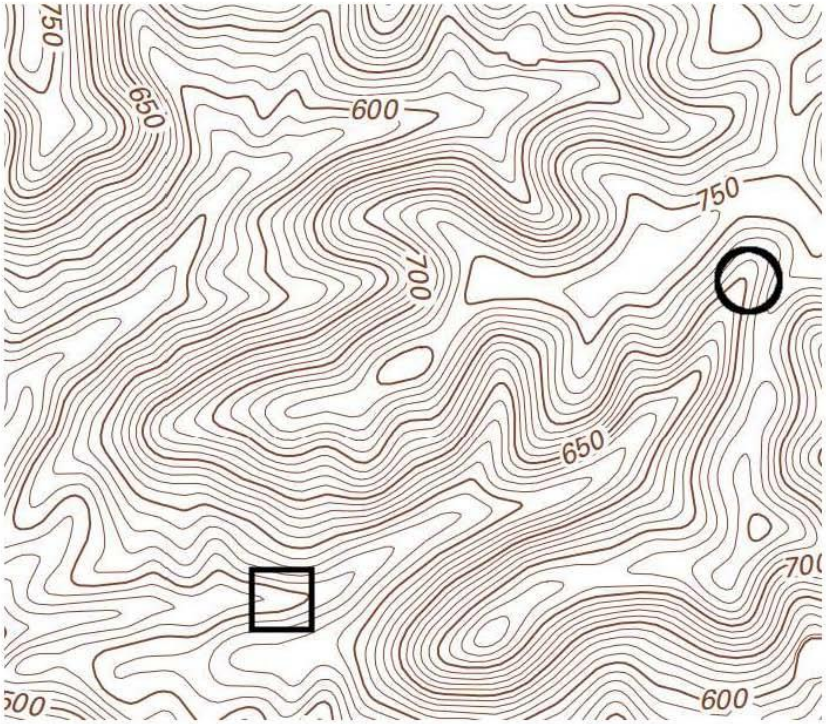
- Who traveled a greater vertical distance? How can you tell?

Mental Effort (1-9) _________



10. Imagine there is a stream that connects the circle and the square. Please draw the path you believe the stream would follow. In addition, clearly mark the direction you believe the water would flow, and why.

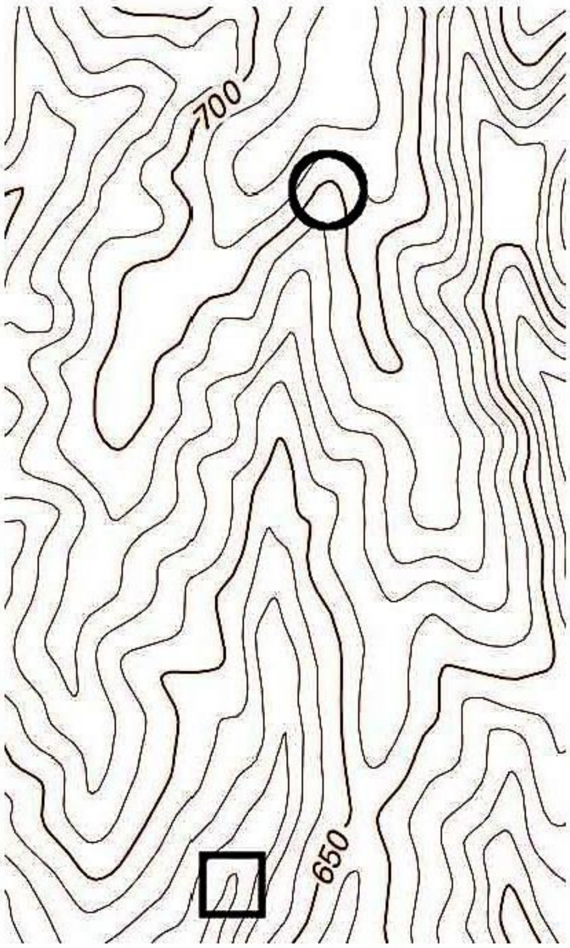
Mental Effort (1-9) _________



11. Imagine there is a stream that connects the circle and the square. Please draw the path you believe the stream would follow. In addition, clearly mark the direction you believe the water would flow, and why.

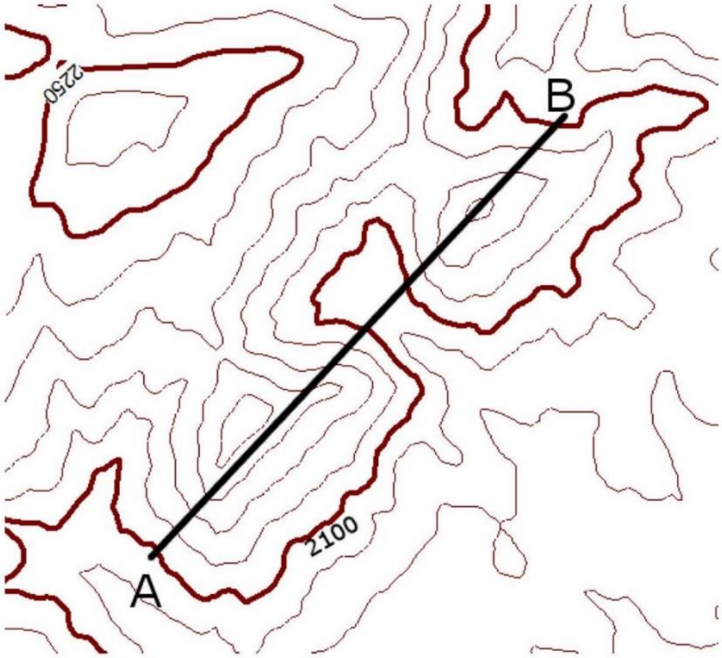
Finally, do you think the water would flow faster near the circle, or near the square? Why?

Mental Effort (1-9) _________

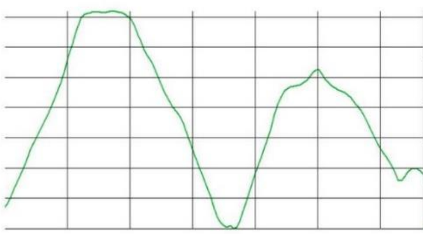
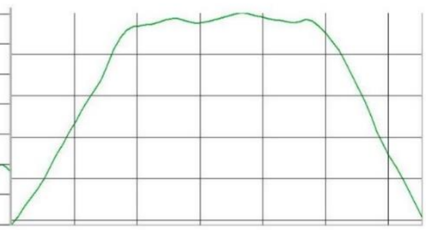


12. Imagine there is a stream that connects the circle and the square. Please draw the path you believe the stream would follow. In addition, clearly mark the direction you believe the water would flow, and why.

Mental Effort (1-9) _________



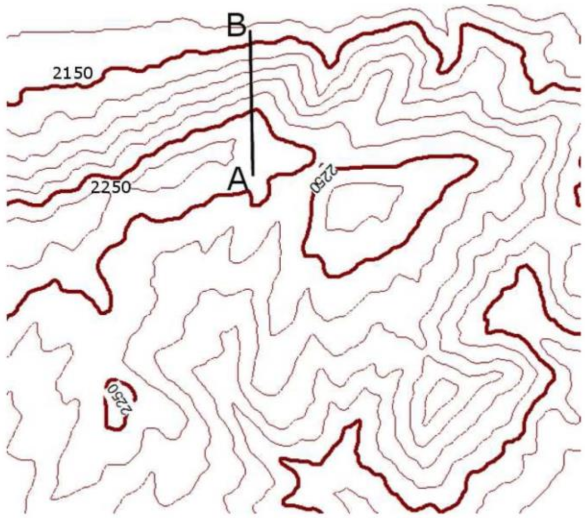
13. Which elevation profile (below) matches the cross-section of the line AB (above)?

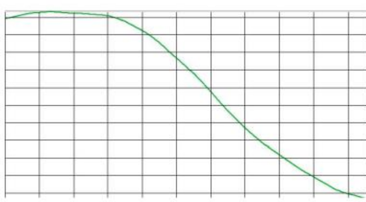
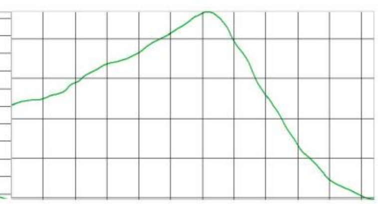
A B A B


A B A B

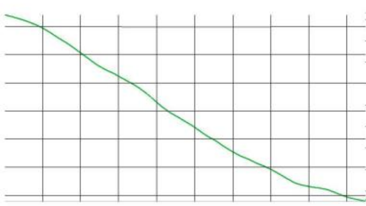
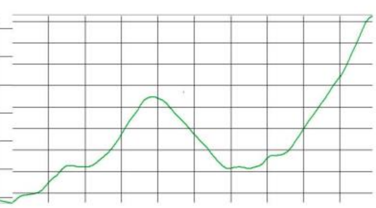
Mental Effort (1-9) _________



14. Which elevation profile (below) matches the cross-section of the line AB (above)?

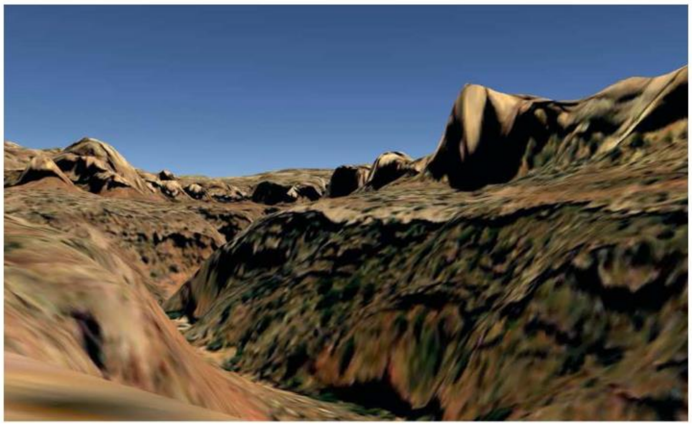



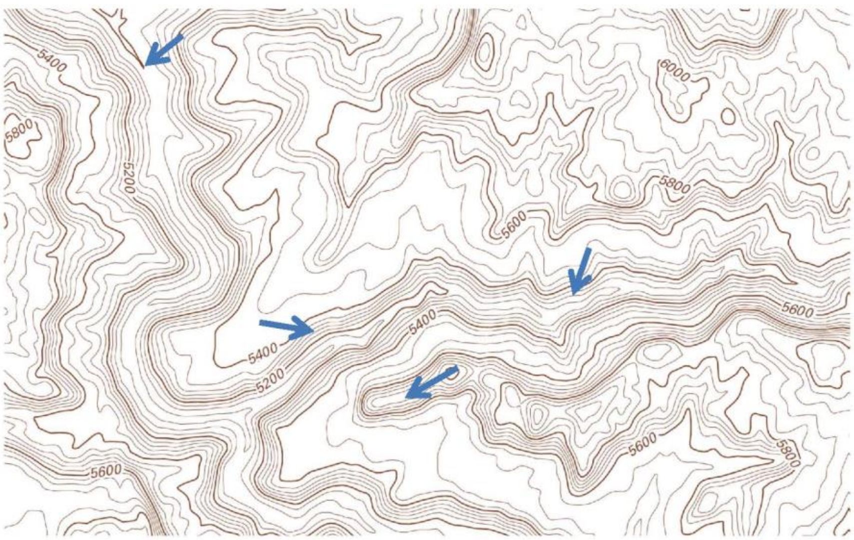
A B A B

A B A B

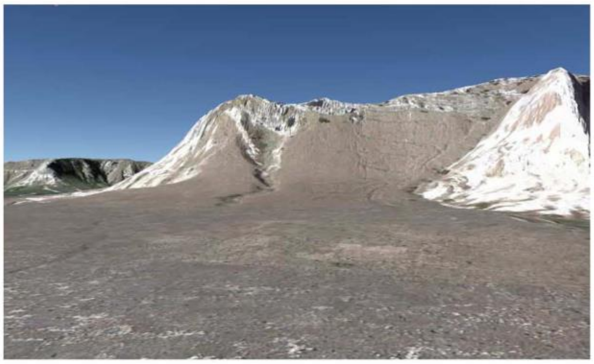
Mental Effort (1-9) _________

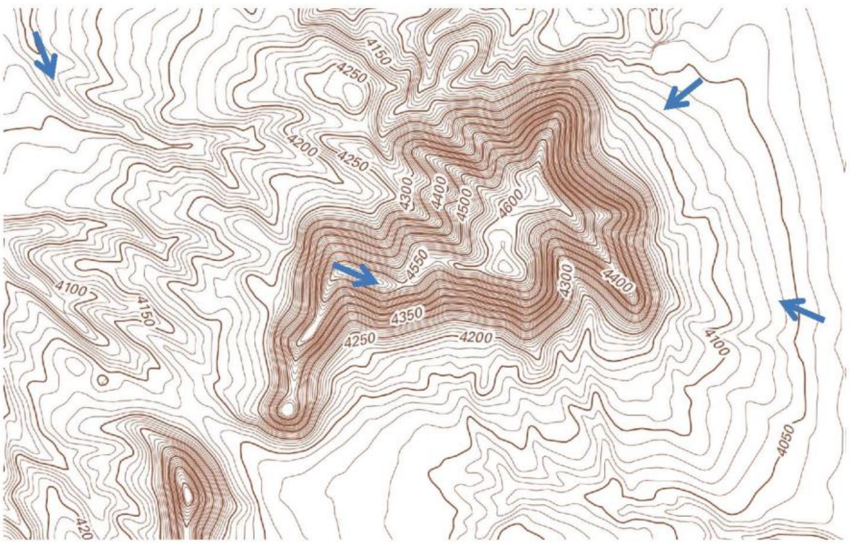




15. Imagine you see the view of the picture above. Circle the arrow on the map that indicates where and which direction you think you are facing.

Mental Effort (1-9) _________



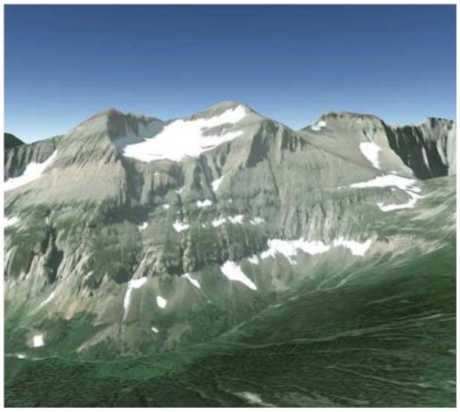


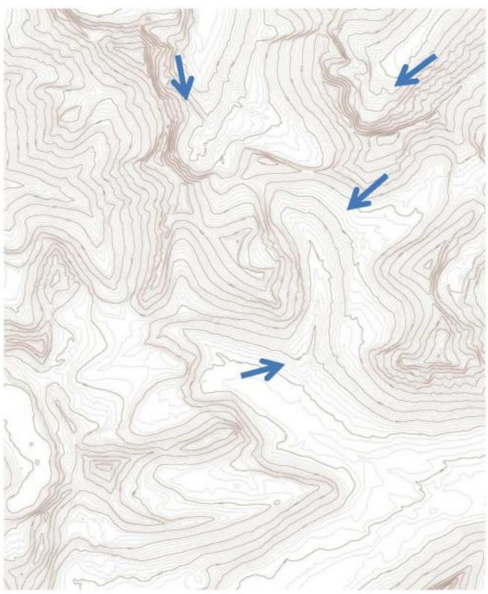
16. Imagine you see the view of the picture above. Circle the arrow on the map that indicates where and which direction you think you are facing.

Mental Effort (1-9) _________

17. You are standing at the square, but you want to get to a place (on the map) where you would be able to see a small lake at the circle. Assume there is no vegetation. Please draw a line from the square to another place on the map that indicates the route you would take to a spot where you can see the circle. Explain below, why you chose the spot as well as the route to get there:

Mental Effort (1-9) _________





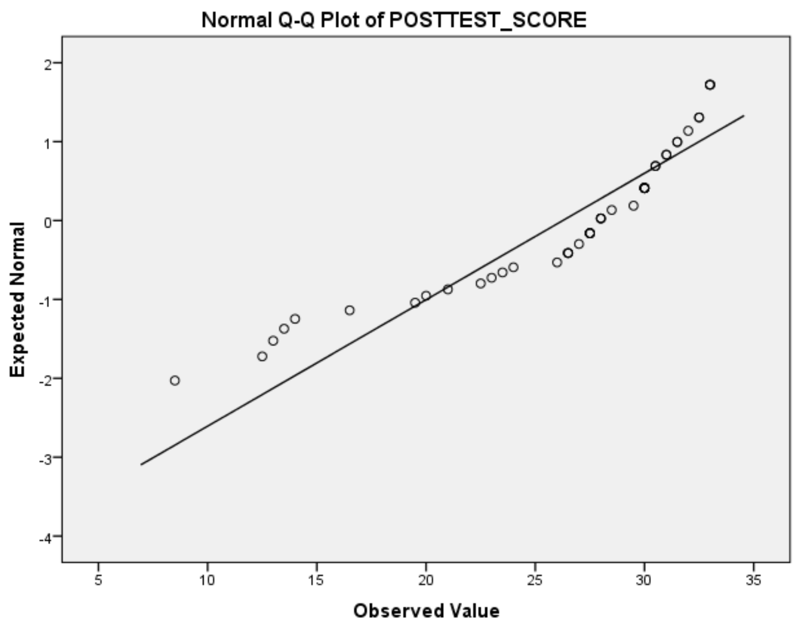
18. Imagine you see the view of the picture at top. Circle the arrow on the map that indicates where and which direction you think you are facing.

Mental Effort (1-9) _________

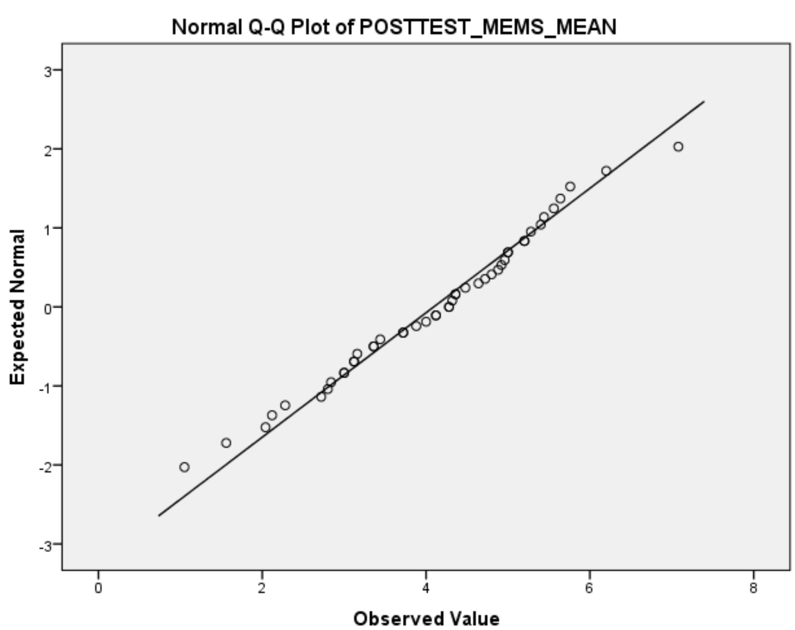
Appendix L

Normality Plots for Two-Way ANOVA

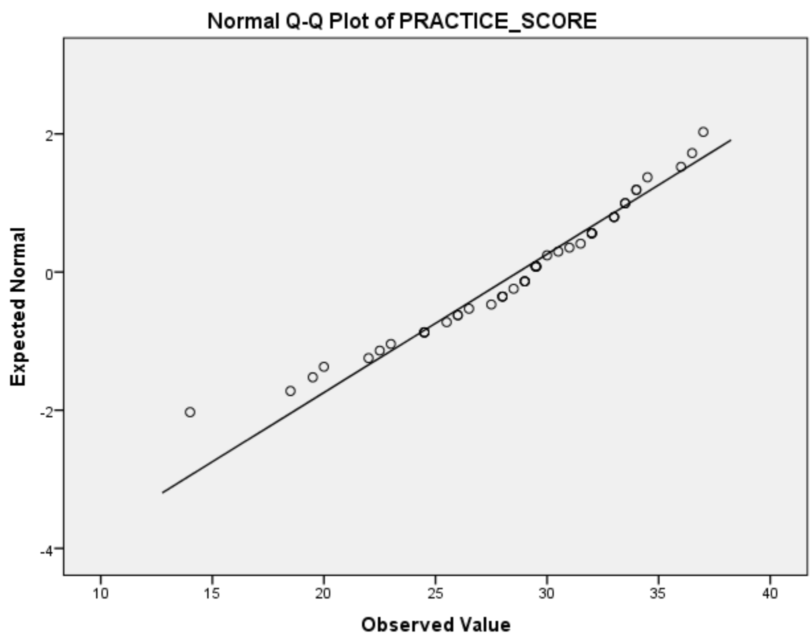
mTMA posttest scores distribution



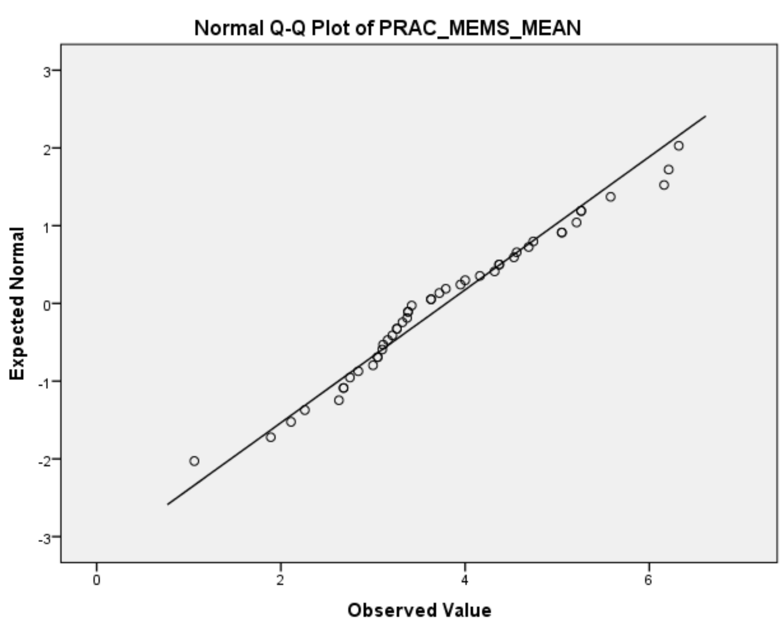
Posttest mean MEMS values distribution



Learning practice activity scores distribution



Learning practice activity mean MEMS values distribution



Appendix M

Two-Way ANOVA Summary Tables

Dependent variable: **mTMA posttest score**						
Source	***SS***	***df***	***MS***	***F***	***p***	**Partial η^2**
Intercept	31646.088	1	31646.088	763.319	.000	.948
Prior Spatial Ability	1.218	1	1.218	.029	.865	.001
Instruction Method	1.850	1	1.850	.045	.834	.001
Prior Spatial Ability * Instruction Method	6.720	1	6.720	.162	.689	.001
Error	1741.258	42	41.459			
Total	33474.500	46				

Dependent variable: **Mean Posttest Mental Effort (MEMS) value**						
Source	***SS***	***df***	***MS***	***F***	***p***	**Partial η^2**
Intercept	772.312	1	772.312	516.251	.000	.925
Prior Spatial Ability	1.595	1	1.595	1.066	.308	.025
Treatment Condition	2.435	1	2.435	1.628	.209	.037
Prior Spatial Ability * Treatment Condition	6.016	1	6.016	4.021	.051	4.021
Error	62.832	42	1.496			
Total	843.280	46				

Dependent variable: **Learning practice activity score**						
Source	***SS***	***df***	***MS***	***F***	***p***	***Partial η^2***
Intercept	37760.238	1	37760.238	2189.045	.000	.981
Prior Spatial Ability	19.716	1	19.716	1.143	.291	.026
Treatment Condition	332.082	1	332.082	19.252	.000	.314
Prior Spatial Ability * Treatment Condition	56.169	1	56.169	3.256	.078	.072
Error	724.485	42	17.250			
Total	39028.250	46				

Dependent variable: **Learning practice activity mean mental effort (MEMS) value**						
Source	***SS***	***df***	***MS***	***F***	***p***	***Partial η^2***
Intercept	666.246	1	666.246	699.445	.000	.943
Prior Spatial Ability	5.616	1	5.616	5.896	.020	.123
Treatment Condition	9.457	1	9.457	9.928	.003	.191
Prior Spatial Ability * Treatment Condition	7.107	1	7.107	7.461	.009	.151
Error	40.007	42	953			
Total	723.693	46				