2015

The development and use of a web-based assessment to measure students' use of representations in general chemistry

Jack Polifka
Iowa State University

The development and use of a web-based assessment to measure students' use of representations in general chemistry

by

Jack Polifka

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Human Computer Interaction

Program of Study Committee:
Thomas Holme, Major Professor
Jared Danielson
Joseph Burnett

Iowa State University

Ames, Iowa

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First, I would like to thank Dr. Holme for giving me the opportunity to work with him. In addition, for his guidance and support while being his student.

Next, I want to thank Dr. Danielson and Dr. Burnett for being on my committee and for the numerous suggestions which enhanced my research.

In addition, I would like to thank Dr. Appy, Dr. Bonaccorsi, Dr. Carlson, Dr. Meyer, Dr. Miller, Dr. Murphy, Dr. Petrich, Dr. Schewe-Millar, and Dr. Windus for helping me with my work and their quality feedback.

I would also like to thank the Holme research group, John Balayut, Allie Brandreit-Harshman, Jenny Brigham, Chamila DeSilva, Sanjay Hari, Cindy Luxford, Anna Prisacari, and Jessica Reed for their inspiration and insight.

I want to thank the following individuals for helping me for my research with anything ranging from encouragement to technical issues to writing: Chris Campbell, Anthony Ellertson, Matthew England, Chris Groves, Sarah Huffman, TJ Rakitan, Jeff Sorenson, and Rose Toppar.

Lastly, I would like to thank the National Science Foundation for supporting this project with funding from DUE-1245536.

ABSTRACT

Being able to use representations is a fundamental skill in chemistry. It is important for instructors to make sure that students are proficient in using them in the classroom. Until now, research related to students' use of representations as mainly focused on the knowledge differences between experts vs novices and what misconceptions students have while using representations. This study sought to contribute to that research by investigating which types of representations students use by the way of a web-based assessment which was able to track student representation use, including time spent accessing specific representations within a set of five possibilities.

A two phase research study was used to complete this work. The first phase consisted of developing the initial build of the web-based assessment tool then refining it using usability testing. Twenty-three chemistry students were interviewed regarding the usability of the web-based assessment which led to seven major modifications of the assessment's interface. During the second phase, the web-based assessment was given to 626 general chemistry students across two large Midwestern universities. Results showed the web-based assessment to be an effective method of measurement and that chemistry students seem to be narrow in their choice of representations. Specifically, students used the Lewis structure for the majority of the assessment. This observation points to the importance of the chemistry context in driving student representation use.

In conclusion, this interdisciplinary work offers a new method with which to measure students' use of representations in chemistry while highlighting general chemistry students' tendency to use a limited choice of representations in an assessment setting.

CHAPTER 1
INTRODUCTION

Educational assessment can be defined as the process of gathering information about what students know in order for individuals, such as instructors and policy makers, to make an evaluation about the current state of education (Pellegrino, Chudowsky, & Glaser, 2002). This type of assessment generally has two outcomes: accountability and instructional feedback. Accountability is the act of ensuring that students are learning what has been deemed necessary to learn (Kubiszyn & Borich, 1990). Instructional feedback is the information gathered by assessment, which can be used by instructors to adjust their teaching as needed (Sadler, 1989).

Due in part to the increased awareness of the need to gather such information, the demand for assessment efforts in university courses, including chemistry, has increased in recent years. Determining the extent to which new assessment demands have influenced instructor thinking presents an important need for those interested in educational reform. To better understand the level of this demand, the American Chemical Society (ACS) – Exams Institute (ACS-EI) conducted a national survey in 2010 asking chemistry college faculty around the United States if their departments were expected to enhance assessment efforts or prepare reports related to assessment (Emenike, Schroeder, Murphy, & Holme, 2011). Of the 1546 participants, 72% responded that they were aware of such endeavors related to assessment in their department (Emenike et al., 2011).

Partly in response to these departmental endeavors, the ACS-EI has completed several projects to enhance the assessment efforts of chemistry instructors. These projects include the following: an anchoring concepts map for general chemistry to help instructors better report assessment outcomes in their classrooms (Murphy, Holme, Zenisky, Caruthers, & Knaus, 2012), a national survey to determine what visual skills are most important to biochemistry instructors (Linenberger & Holme, 2014), and an analysis tool called the Exam Data Analysis Spreadsheet which allows instructors to preform customizable analyses using their students' ACS exams and to compare the results with national results (Brandriet & Holme, 2015). Respectively these projects were completed to help general chemistry instructors better understand the content areas they were measuring, to better understand biochemistry instructors' view regarding the visualization of molecules, and to help instructors better interpret data from their measurements.

One area of chemistry that would benefit from such projects would be the assessment of representations in general chemistry. Representations are a correct component in chemistry and given the ACS-EI's work related to molecular visualizations, it would seem to be a logically next step. Representations act as a type of common language in chemistry for discussing scientific phenomena (Hoffman & Laszlo, 1991). This is done by the way representations complement and constrain information (Ainsworth, 1999, 2006). Confirming that students have representational competence, or the ability to understand and use a variety of representations in chemistry is an important task (Russell & Kozma, 1997; Stieff, Hegarty & Deslongchamps, 2011). While a considerable amount of research has been conducted to investigate how students use representations in chemistry (Cooper, Underwood, & Hilley, 2012; Hinze, Rapp, Williamson, Shultz, Deslongchamps, & Williamson, 2013; Luxford & Bretz, 2013; Russell & Kozma, 1997; Williamson, Hegarty, Deslongchamps, Williamson, & Shultz, 2013), there is no research related to the how students' comfort with representation usage influences measurements made during assessment.

Chemical education research has recently started to quantitatively measure how students view chemistry representations as a means to understand how students use representations (Hinze et al., 2013; Williams et al., 2013). This research suggests a unique direction that could be used for the development of assessment materials of chemistry representations.

Purpose of Study

This research study had two primary goals. First, to develop and to understand how users think and interact with a method capable of quantitatively measuring how general chemistry college students use representations. Second, develop an online assessment using that method of measurement. The assessment can then be used by general chemistry college instructors for collecting instructional feedback on their students' ability to use chemical representations.

To address these goals, the study was conducted in two phases. During the first phase, a method of measurement was developed. It was called the Variable Representation Assessment (VRA). The VRA is an online assessment application which records how long students use different representation images while solving general chemistry problems as a means to measure how they use representations. The representations chosen for the VRA, at least for general chemistry, were the following: chemical formula, Lewis structure, dash-and-wedge models, ball-and-stick models, and electrostatic potential maps. These representations have been shown to be common in university level chemistry textbooks (Bergqvist, Drechsler, De Jong, & Rundgren, 2013). Once the initial programing for the VRA was complete, the interface was enhanced through a series of qualitative interviews and quantitative methods measuring its usability.

The second phase used the VRA to understand how general chemistry students use chemistry representations and to use that information to enhance the assessment materials.

During this phase, the VRA was administered in two sessions to students from two large Midwestern universities. The results of the assessment along with the results of how students viewed the different chemistry representations reveal how students used the representations answering items that mimic chemistry test questions and provide direction for improving the assessment.

To guide the efforts of the study, the following research questions were used:

1. *How is the usability of the VRA perceived by users?*
2. *What does the data produced by the VRA tell us about how general chemistry students use representations?*

Thesis Outline

The remaining sections of this thesis are organized as follows: Chapter 2 provides a review of the relevant research literature focusing on what is a representation, their role in chemistry, how their usage has been measured, and what are the benefits and obstacles of web-based measurements. Chapter 3 presents the development of the VRA and chemistry faculty members' opinions of it. Chapter 4 covers two experiments performed using the VRA and the results of each experiment. Chapter 5 concludes with general results of the study, limitations, and directions for future research.

CHAPTER 2

REVIEW OF RELEVANT LITERATURE

This chapter presents the literature reviewed for this study. Discussion in this chapter will start by answering the following questions: what is a representation and what is the role of representations in chemistry? The discussion then transitions to answering: how is the use of representations measured in chemistry and what are the benefits and obstacles of web-based assessments?

Representations

The use of representations is present in many disciplines including: biology (Tsui & Treagust, 2003), mathematics (Arcavi, 2003; Brenner, Mayer, Moseley, Brar, Duran, Reed, & Webb, 1997), and physics (Fredlund, Airey, & Linder, 2012; Tang, Tan, & Yeo, 2011). These disciplines require students to understand concepts to which they cannot directly interact with. To aid students in these disciplines, instructors use representations to help them understand these concepts. Representations act as intermediaries between abstract concepts and students, which they can use to learn those concepts (Jong, Ainsworth, Dobson, Hulst, Levonen, Reiman, Sime, Someren, Spada, & Swakk, 1998). The use of representations, especially multiple representations, helps students to understand different concepts by three primary methods: complementing the information each depicts, constraining each other, or a combination of the previous two (Ainsworth, 1999, 2006).

Representations in Chemistry

Chemistry is another discipline that uses many representations. Representations act as a common language for chemists of various expertise to understand and discuss chemical phenomenon (Hoffman & Laszlo, 1991). The use of representations is a skill individuals need to have in order to be successful in chemistry (Treagust & Chittleborough, 2001). There are multiple ways that an individual can express proficiency using representations; two of these include the ability to traverse Johnstone's triangle (Johnstone, 1982, 1993, 2000, 2010) and representational competence (Russell & Kozma, 1997).

Johnstone's triangle is the idea that knowledge in chemistry has three aspects of knowledge and representation: the macroscopic, the microscopic, and the symbolic (Johnstone, 1982, 1993, 2000, 2010). Coal can be used to illustrate these three aspects. The macroscopic or observable properties of coal would be its black and rocky exterior. Coal's microscopic attributes, details that cannot be seen, are the atoms that bond together to form a piece of coal. The chemical formula that represents coal would be an example of its symbolic domain. Being able to transition from one aspect to another is a key in being successful in chemistry (Johnstone 1993, 2000, 2010). Figure 2.1 shows Johnstone's triangle with various coal illustrations.

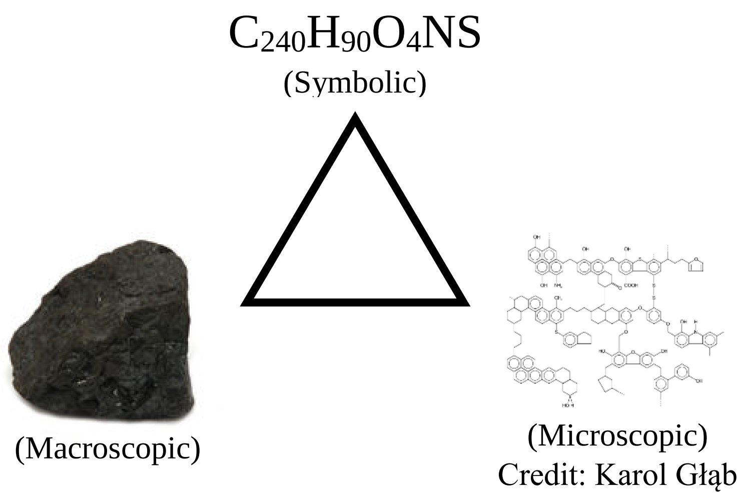


Figure 2.1: Johnstone's triangle of chemistry

Representational competence in chemistry is the ability to understand and use a variety of representations (Russell & Kozma, 1997; Stieff, Hegarty & Deslongchamps, 2011). When Kozma and Russell (1997) introduced the term, they defined individuals as having the ability to "…see expressions with different surface features as all representing the same principle, concept, or chemical situation, and they can transform the expression of a chemical concept or situation in one form to a different form."

The Measurements of Representation Use in Chemistry

Because the use of representations is important in chemistry, many studies have been conducted to investigate how students use them. Many of these studies can be classified into one of three categories: new techniques to support the teaching of representations (teaching), measuring misconceptions related to representations (misconceptions), and the general measurement of representation use (measurement).

The following section will focus solely on misconceptions and measurement research as those studies better align with the study presented in this thesis. Table 2.1 highlights a number of studies that measured how students used representations and specifically lists, which category of representation research it fits, the authors of the study, and the method of measurement in the study.

Table 2.1: Studies Measuring Individuals Use of Representations

Type of Study	Authors	Method of Measurement
Misconceptions & Measurement	Luxford & Bretz (2013, 2014)	Interviews and Concept Inventory
Measurement	Williamson, Hegarty, Deslongchamps, Williamson, & Shultz (2013)	Eye-tracking
Measurement	Hinze, Rapp, Williamson, Shultz, Deslongchamps, & Williamson (2013)	Eye-tracking
Misconceptions & Measurement	Cooper, Underwood, & Hilley (2012)	Interviews and Concept Inventory
Measurement	Kozma & Russell (1997)	Interviews

Measuring Bonding Misconceptions by Luxford & Bretz (2013, 2014)

Luxford & Bretz (2013, 2014) investigated students' understandings of ionic and covalent bonding. For this purpose, they wanted to develop a concept inventory from a single context. Concept inventories at the time related to students' understandings of bonding were developed from various expert knowledge contexts. The context Luxford & Bretz chose was students' ideas of bonding. To understand these ideas, 24 interviews were conducted with high school and first year college chemistry students where they created multiple representations using colored pens, paper, Playdoh®, and toothpicks followed by discussion of expert generated representations. Quotes and themes that emerged from the interviews were used for the development of a 32 item pilot concept inventory called the Bonding Concept Inventory (BRI).

In a pilot study, the BRI was administered to 247 second-semester general chemistry students. The final version of the BRI was revised to 23 questions based on the results of the pilot study and was sent to 1072 high school and college students across the United States. Analysis of the students' results showed that the BRI could measure students' bonding misconceptions in a valid and reliable manner. In addition, it highlighted the potential of mixed-method research and revealed some common misconceptions students held as related to bonding.

Using Eye Tracking to Measure Students' Choice of Ball-and-stick versus Electrostatic Potential Maps by Williams et al. (2013)

Williams et al. (2013) explored the possibility of using eye-tracking in chemical education research as they investigated students' eye movements and focus on ball-and-stick images versus electrostatic potential map images. Williams et al. conducted a pilot study with 9 organic chemistry students that asked students to solve 40 randomized questions. Twenty-four of those questions involved two static side-by-side images: one was a ball-and-stick representation and the other was an electrostatic potential map representation. The 24 questions were comprised of 4 common questions repeated with 6 sets of different molecules. Analysis of students' eye movements revealed that the amount of time spent on each representation was dependent on the questions being answered. Specifically, analyses showed that the amount of time students focused on ball-and-stick images compared to the electrostatic potential map images was statistically different for most questions. Interestingly though, when the relationships between item correctness and time spent on ball-and-stick images were compared, a negative or neutral relationship was found for all questions. When the same relationship was calculated with electrostatic potential maps, a positive relationship was found for all questions. The results of the study showed eye-tracking software to be an effective tool in chemistry education research and that an individual's focus was an important qualitative measurement.

Further Studies with Eye Tracking by Hinze et al. (2013)

Hinze et al. (2013) sought to further investigate students' choice of ball-and-stick representations vs electrostatic potential map representations. They used eye-tracking in a similar fashion when compared to the Williams et al. (2013) study. Hinze et al. (2013) focused on how prior knowledge and question type affected students' choice of ball-and-stick representations and electrostatic potential map representations. A total of 225 organic chemistry students from Texas A&M University were asked to complete an 18 question in-class quiz, which was used to determine high prior knowledge (HPK) (top 1/3 of scores) and low prior knowledge (LPK) (bottom 1/3 of scores) participants. Based the results, 18 HPK and 12 LPK students were recruited for the study. The students recruited attended eye-tracking sessions where they answered 24 questions involving two static side-by-side images: a ball-and-stick representation and an electrostatic potential map representation. Five seconds prior to each question, participants could freely view each image. This pause served as an opportunity to measure if any bias was present with each set of images. Analyses did not find any statistically significant differences from these moments of viewing. Each question in the study was defined as being an identification or inference. Analyses revealed that HPK students scored statistically higher than LPK students on the activity overall. Specifically, HPK students performed statistically higher on both identification questions (p-value = 0.05) and inference questions (p-value = 0.003). When students' scores were analyzed in the context of the eye tracking measurements, they found statistical significance between the amount of time HPK and LPK students viewed electrostatic potential maps. The study by Hinze et al. (2013) forth enforced eye-tracking as a potential research tool for chemical education.

The Implicit Information from Lewis Structures Instrument by Cooper et al. (2012)

Cooper et al. (2012) had the goal of developing a method that could measure students' misconceptions related to Lewis structures. They accomplished this by developing a concept inventory grounded in student understanding. Their initial step consisted of asking 32 second-semester general chemistry students, 134 second semester organic chemistry students, and 10 graduate students the following question: "What information can be obtained from a Lewis structure?" Based on the responses, two researchers found eleven topics regarding Lewis structures. A random subset of students' responses were selected and coded according to those topics. Inter-rater reliabilities ranging from 0.83 to 1 were reached by researchers. The coded responses were then used to develop the preliminary version of the Implicit Information from Lewis structures (IILSI) concept inventory. The IILSI was reworded for clarity purposes over three semesters after being administered to 648 general chemistry students and 2,336 organic chemistry students. The final version of the IILSI was administered to 2,349 students. In order to determine if the IILSI produced reliable measurements, two sets of general chemistry students from the academic years of 2010 and 2011 who had the same instructor were compared with a Chi-square analysis. This analysis determined that there was no significant difference between the two sets of students' answers. Overall, the IILSI provides instructors with a tool that they can use to easily assess what their students know and do not know about Lewis structures.

Expert vs Novice Representation Use by Kozma & Russell (1997)

Kozma & Russell (1997) explored the differences between experts' and novices' use of representations with the hope of discovering methods that could be used to develop effective multimedia techniques for teaching representations. They conducted two think-aloud experiments to investigate these differences. The first study had 11 experts and 10 novices who were given a stack of fourteen 4-by-6 inch cards with different chemical images on them. The images belonged to one of four categories: diagrams, graphs, animations, and video. All images, excluding diagrams, had a corresponding 10 to 40 second movie clip. Participants were asked to construct card groupings after they had viewed all of the cards. Card groupings could be of any size and created for any reason, such as but not limited to, all of the cards had a common method of portraying information or all of the cards related to a specific chemical concept. Participants were asked to think aloud as they created their groupings. Analysis of the groupings showed that experts, when compared to novices, created groupings that were larger, more diverse in the number of different types of representations used, and more focused on conceptual than surface details.

The second experiment had the same participants complete a transforming task. Participants were shown 15 different representations and then asked to select an equivalent representation from a stack of 15 cards or to generate their own representation if necessary. Each transformation had a single correct answer. Experts were scored statistically higher on the transformation activity than novices.

These experiments showcased a unique method to measure representation use and to utilize multiple representations. In general, experts were able to work with representations more efficiently based on several factors, which included effective transferring across various representations and an understanding of conceptual ideas rather than just surface details.

Summary of Measurements of Representation Use in Chemistry

The use of representations in chemistry have been measured with several different methods. Each method has its own strengths adding to the understanding how individuals use representations. One method of measurement that appears to not be utilized though, is web-based assessments.

Web-Based Assessments

Web-based assessments are tests, which are delivered via the Internet. Commonly, they are discussed as e-assessment, computer-assisted assessment, and online assessment. Web-based assessments have their benefits and obstacles when compared to most traditional methods as paper-and-pencil or in-person interviews.

Benefits

Benefits associated with web-based assessments include automation and data collection. Web-based assessments have the potential to reduce the amount of time and streamline the effort required by instructors for a variety of tasks via automation; thus, allowing them to use more resources for other endeavors for their students. In addition, web-based assessments have the potential to collect more diverse data than standard tests.

The time saved by web-based assessment automation can take several forms, such as automation of scoring and delivery. The automation of scoring eliminates the need to hand score an assessment. This can allow students to receive feedback dramatically faster, even instantaneously in many cases (Jordan & Butcher, 2010). The automation of delivery can reduce the amount of time and effort required to distribute an assessment. Technologies, such as java applets, can allow an assessment to operate on multiple computers without being dependent on the specific application installed (Crisp 2001). In the case of web-based assessments, it would only require a computer to have an internet browser to utilize them.

Another main advantage associated with web-based assessments is their ability collect diverse and rich data (Vendlinksi & Stevens, 2002; Zenisky & Sireci, 2002). This is easily accomplished due to their visual interfaces, which allow for a variety of novel question types, such as drag-and-drop, passage editing, and inserting text (Boyle & Hutchinson, 2009; Zenisky & Sireci, 2002).

Obstacles

Several obstacles need to be overcome before the benefits of web-based assessments can be fully utilized. One of the most important obstacles to overcome when developing a web-based assessment is its usability or ease of use. Since the study of usability is a large field, only a limited sample of usability studies will be discussed.

The visual interfaces of web-based assessments have the possibility of creating problems for users. Ensuring that they are easy to operate for users is crucial because if the assessments are not operated properly, the information they collect may be incorrect. Past studies show that usability tests are an important part in the design of such systems and should be essential for their development (Crowther, Keller, & Waddoups, 2004; Lim, Song, & Lee 2012).

Usability tests can be conducted several ways. When Crowther et al. (2004) wanted to ensure students were able to effectively use an online system for teaching introduction chemistry, they recorded students' feedback while they interacted the system. In addition, they asked several interview questions post-interaction. Crowher et al. (2004) found that the system had functional and navigational issues due to their efforts.

Lim et al. (2012) conducted usability test using a method called iterative design where feedback would be recorded for one version of a system that would then inform the design of the next version then repeated as necessary. Lim et al. (2012) completed two cycles using this method with interviews, questionnaires, and experts' suggestions when developing the KERIS digital notebook. Using these methods they discovered issues such as, ineffective menu layouts that confused users and poor navigation that led to users to use their time searching for information using the system.

The preceding literature regarding usability is only a limited sample of the present research completed in the field. Techniques for conducting usability tests are presented in Chapter 3.

Summary

Representations are an important part of several disciplines. They act as a medium between learners and the abstract concepts that they portray in order to help students learn. Chemistry is a discipline that uses a variety of different representations in order to discuss different phenomena. Ensuring that students are satisfactory and fluent in the use of representations is important. Several studies have been completed using a variety of different methods to better understand what students understand as related to representations. One method that has been not used is the web-based assessment. Web-based assessments allow for measurements that seem impossible with other methods. Extra care is required in order to take full advantage of them though. Usability is one of the areas that requires extra attention.

CHAPTER 3

PHASE ONE: DEVELOPING A WEB-BASED REPRESENTATION ASSESSMENT

This chapter will describe the development of the Variable Representation Assessment (VRA), which is used throughout this research study. Specifically, this chapter will explain what the VRA entails, what the usability test performed using the VRA informed for its development, and the opinions of chemistry faculty members concerning the VRA. After the VRA was fully developed, it was evaluated for its usability using an iterative process where it under went three sets of modifications. Once the modifications were complete, a survey was issued to gather chemistry faculty members' opinions from around the United States about the VRA regarding their interest in it and the general direction of its design.

Variable Representation Assessment

The VRA is a web-based assessment tool that was developed to record how long students view different representations while solving chemistry problems as a means to measure how they use different representations. During development, the VRA could be accessed by any device that had an internet browser. Users with an internet connection and an internet browser could log into the VRA using a username created by the test administer. Once users logged in, a randomized test was generated for the user using a set of pre-defined questions to account for any order item effect (Schroeder, Kristen, & Holme, 2012). Each question had a set of three different molecules as answer choices given in a random order to remove bias in answer choices. Each molecule had a set of five different images with each representing a different chemical representation that appears in a random order to control for bias. The exception was the first representation button, which was always the chemical formula for each molecule. The view of each molecule was restricted to only allow one representation to be viewable at a time. Users could change to any representation while answering questions. The restriction to only allow one representation image per option choice to be viewed at a time was a design choice. Limiting or segmenting the amount of information a student has access to at a given moment can reduce their cognitive load when using an application (Mayer & Moreno, 2003). While users viewed different representations, their choice of representation was recorded along with a timestamp of when they changed the representation using the navigation. Figure 3.1 shows the appearance of the first version of the VRA.

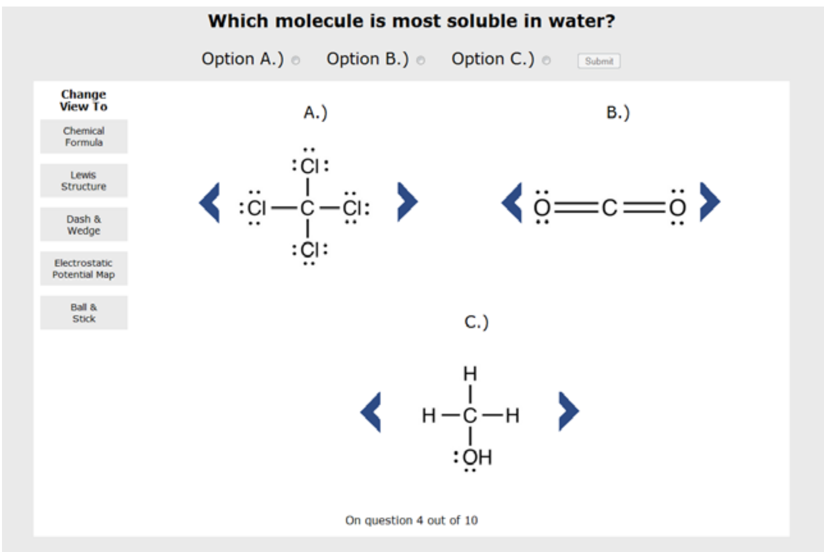


Figure 3.1: First version of the VRA with navigation on the left side and on the sides of each representation for changing views

Usability Test

The effectiveness of the VRA's measurements depended on how well students understood how to change representations using its navigation. Therefore a usability study was conducted.

Before a usability study can start, goals should be defined. Once goals are defined, proper formative and summative forms of evaluation can be selected for the usability test to address those goals. Formative evaluation is used to understand the specific details about an interface while summative evaluation is concerned with the overall interface (Nielsen, 1994; Rohrer 2014). Summative evaluation can also be used to tell how much better or worse an alternative interface is compared to other versions of the same interface (Nielsen, 1994; Rohrer, 2014).

Formative and Summative Evaluation Methods

The primary goal of the usability study was to ensure that users employed the VRA's navigation to change representations. The secondary goal of the usability study was to address any other usability issues users encountered since they may affect users' answer choices in unforeseen ways.

To address the primary goal, a formative evaluation method called Ajax action tracking was selected. Ajax action tracking is a method for observing actions a user preforms while using the VRA. It was inspired by past web-based action tracking techniques (Atterer, Wnuk, & Schmidt, 2006; Atterer & Schmidt, 2007). Using this method, the frequency at which a user changed representations was recorded.

To address the secondary goal, a mix of formative and summative evaluation methods were used. The first formative method was Ajax action tracking. The VRA had functionalities besides being able to change representations such as enlarging representations. Ajax action tracking was used to record how frequently a given functionality was used. In addition to Ajax action tracking, interviews were used as a second formative measure of usability. Interviews are a method to collect users' preferences and issues about an interface (Holzinger, 2005).

Interviews were used for two reasons. First, they were a method of collecting qualitative data regarding actions measured by Ajax action tracking. Second, they were an exploratory method for uncovering usability issues that could not be discovered using the other evaluation methods used in the usability test. The summative method was the System Usability Scale (SUS) (Brooke, 1996). The SUS is a 10-item survey that has been shown to be a reliable and valid quantitative measure of self-reported usability (Bangor, Kortum, & Miller, 2008).

Iterative Design

Development of the VRA used an iterative process. As data was collected from usability testing, the next version of the interface was developed based on the analysis of the data. Jakob Nielsen states that having five users is sufficient to find 85% of usability issues for a given interface (Nielsen, 1994; Nielsen, 2000). If five users can find that many issues from one interface, it may be a more effective use of resources to give new users a new interface to test instead of having them use an interface that was already tested. Therefore, new testers would be given the new interface instead of the already tested interface. Participants were never re-used using this process.

The following steps were used in the iterative design of the VRA:

1. Development of the first version of the VRA
2. Usability testing of the first version (10 participants)
3. Analysis of data from the usability test of the first version of the VRA
4. Use results of the first version of the VRA to inform development of the second version of the VRA

5. Usability testing of the second version of the VRA (8 participants)

6. Analysis of data from the usability test of the second version of the VRA

7. Use results of the second version of the VRA to inform development of the third version of the VRA

8. Usability testing of the third version of the VRA (5 participants)

9. Analysis of data from the usability test of the third version of the VRA

No assessment items were modified during this process.

Components of the VRA

The VRA tool had four distinct components corresponding to the different components of the study. A pre-survey where users can report their experience using different chemical representations, the interface that allows users to navigate through and view different representations, the different assessment items, and a post-survey for measuring the usability of the tool. The following sections describe in further detail each components' design and purpose.

Pre-survey

In order for the usability testing to simulate experimental conditions as accurately as possible, all components of the study were added to the usability test. After logging into the VRA, participants were asked to complete a pre-survey asking for their experience with each representation within the application: chemical formula, Lewis structure, dash-and-wedge model, ball-and-stick model, and electrostatic potential map. For each representation, participants were asked if they had ever solved a chemistry problem using the representation. If yes, when did they first solve the problem and how frequently do they use the representation to solve problems? The order of the pre-survey questions was static. All pre-survey responses were recorded in a MySQL database and analyzed using R, a statistical software package. Data from the pre-survey collected during the usability test is not reported as the focus of the usability test was on the development of the VRA.

Interface

The interface of the VRA was made of two primary components. The representations users could view while solving assessment items and Ajax action tracking, which was responsible for recording areas of functionality users interacted with.

The chemical formula, Lewis structure, and dash-and-wedge model representations in the VRA were created using ChemDraw 13 (Ultra, 2001) software and edited as needed with Microsoft Paint software. Electrostatic potential maps and ball-and-stick models were created using Spartan 2008 (Spartan, 1995) software and default settings.

Ajax action tracking is an adaptable method that can record different pieces of information based on its design. For the VRA, it was designed to record how frequently different functionalities were used to see if they were utilized by users. Similar techniques have been shown to be powerful tools for usability research (Atterer, Wnuk, & Schmidt, 2006; Atterer & Schmidt, 2007; Menezes & Nonnecke, 2014). Below are the different actions for each version of the VRA that were recorded along with a description of the action.

The first version of the VRA recorded three actions:

1. Zoom functionality – Users were able to make one and only one representation larger for a period of time. Users could see the enlarged image by clicking on the smaller image of any representation.

2. Single representation navigation – Users could change the representation that was displayed for an individual molecule. Representations were in a static sequence when users used the single representation navigation. This action was performed when one of the two blue arrows on the left or right side of a representation was clicked using the mouse.

3. Multiple representation navigation - Users could change the representation that was displayed for all three molecules at once regardless of what representation was visible for any molecule. Representations would display according to the label on the button clicked by the user. This action was performed when the buttons on the left side were clicked using the mouse.

The zoom functionality and single representation navigation were removed from the second version of the VRA. The reasons for these modifications are discussed in the data and analysis portion of the usability test section. The following actions were recorded in the second version of the VRA:

1. Multiple representation navigation – Same action as in the first version.

2. Image Answer Selection – Users could select an answer choice by clicking on the representation image, which corresponded to the appropriate answer choice. This action was performed by clicking on the representation using the mouse.

The image answer selection was modified in the third version of the VRA. The reasons for this modification are discussed in the data and analysis part of the usability test section. The third version of the VRA recorded one action:

1. Multiple representation navigation – Same action as in the first and second version.

For this study, Ajax action tracking also recorded the specific time when a user would change which representation was visible. The records allowed for the total amount of time a representation was visible to be calculated. Recording time spent viewing representations is normally done with eye-tracking software allowing only one individual to be recorded at a time (Hinze et al., 2013; Williams et al., 2013). By using Ajax action tracking, similar measurements can be recorded for multiple individuals at one time.

All actions and time measurements recorded with Ajax action tracking were stored into a MySQL database and analyzed using R. Only the frequency of actions are reported in this chapter.

Assessment

The assessment component in the VRA was designed for general chemistry students. In order to develop assessment questions, different general chemistry textbooks were reviewed including Bauer (2009), Brown, LeMay, Bursten, Murphy, & Woodward (2014), Russo & Silver (2001), and Silberberg & Duran (2002). An initial list of 38 questions was created and was later reduced to 11 questions by the consensus of a team of chemistry education researchers. All assessment responses were recorded in a MySQL database and analyzed using R. Results of the assessment are not reported for the usability test but are reported for the pilot study and full experiment in chapter 4.

Post-survey

Participants completed the SUS as a post-survey after completing the assessment component of the VRA. Scoring the SUS gives a score of 0 to 100. 0 being poor usability and

100 being excellent usability (Brooke, 1996). All SUS responses were recorded in a MySQL database and analyzed using R.

Participants

After approval from the Institutional Review Board was obtained, see Appendix A, 23 students from a large Midwestern university during the fall of 2014 participated in the usability test. In-class announcements were used to recruit participants from courses in general chemistry for engineers, advanced general chemistry, and introductory quantum mechanics. All students were informed of the identities of the key personnel who were involved in the study, the need for 20 to 30 minutes of their time for the activity, and the compensation of pizza for volunteering their time. Individuals could express their interest to participant with email or by sign-up during class announcements. Meetings were set at mutual convenience times. All participants stated they were 18 years or older. An example of the in-class recruitment signup can be found in Appendix B.

The logic for asking different levels of chemistry students to participate was to get a more diverse group of users. Students with varying levels of chemistry knowledge were invited to highlight possible usability issues that might depend on familiarity with the subject matter. Differences in users have been discussed as a factor in usability issues (Nielsen, 1989).

Procedure

Participants and the researcher met at a mutually agreed upon location and time. Each participant was then given two copies of the consent form: one for their personal records and the other to be signed confirming their consent. Each participant was given a verbal explanation of the consent form contents and activity. After verbal explanations were given, participants could read the consent and ask questions. Once questions were answered and formal consent was given, each participant was given a username and asked to log into the VRA website. All participants used the internet browser Firefox on the same Windows 8.0 Lenovo laptop for this activity. While participants were using the VRA, the researcher remained in close proximity to participants until completion of the activity to answer any questions from the participant.

Following the activity, interviews were conducted. Each version of the VRA has its own interview guide developed with a common set of questions along with version specific questions. The common set of questions were developed to be unrestrictive with respect to the qualities of the VRA as to allow participants to talk about what they found to be most relevant. The version specific questions were developed to address actions and features that were unique to a given version. Follow-up questions were asked as needed to encourage users to elaborate on their initial statements (Holzinger, 2005). Not all participants were asked all questions. All interviews were recorded using an Olympus WS-600S voice recorder and transcribed verbatim for analysis. A full list of interview questions for the first, second, and third versions of the VRA can be found in Appendix C, D, and E respectively. An example of a transcribed interview for the first, second, and third versions of the VRA can be found in Appendix F, G, and H respectively.

After completing the activity, participants were informed of the day and location where they would receive pizza as compensation for their participation. If the given day was inconvenient for a participant, alternative plans were made for them to receive said compensation. An example of the consent form can found in Appendix I.

Results of the Usability Test and Modifications

First Version: Formative Evaluation

Results from Ajax action tracking showed that the 10 participants used the zoom functionality on three occasions and the single representation navigation on nine occasions. All participants were recorded using the multiple representation navigation. There was a total of 197 instances of participants using the multiple representation navigation (M = 19.7, SD = 6.8).

In the first version interviews, participants who were asked if it was difficult to view the representations while using the VRA, all reported no issue viewing representations. In a follow-up question immediately afterwards, participants were asked if they had known about the existent of the zoom functionality. Responses were a mixture of yes and no with both groups reporting that they did not need it. One participant stated they did not know about the existence of the zoom functionality but if they needed such functionality, they would have searched for it.

When participants reported how they changed representations, all stated they used the multiple representation navigation. Participants described using the multiple representation navigation for two reasons. It allowed them to view a specific type of representation immediately versus using the static sequence of representations via the single representation. In addition, the multiple representation navigation allowed for easier comparison of the representations as it changed all three representations to one specific type.

The final finding of the formative evaluation was related to the triangular layout of the representations. One participant when asked if there was anything they disliked about the VRA, they explained how the triangular layout of the representations did not line up with the radio buttons responsible for selecting answer choices making it difficult at moments to select answers. The triangular layout also created a scrollbar that users did not appreciate.

First Version: Summative Evaluation

The average SUS score of the first version was 90.3 (SD = 6.3). This score was considered high based on past SUS score analysis (Bangor et al., 2008). Generally, scores over 70 are considered acceptable while scores over 85 are given to interfaces of higher quality. While the current SUS score is an important piece of data, it may be more important when compared to the scores of the other versions to see if any version of the VRA scored statistically better or worse than the others.

First Version: Modifications

Four major modifications were complete after a week of analysis and programming. The first response of the usability test on the first version of the VRA was the removal of the zoom functionality. As indicted by Ajax action tracking and interviews, the zoom functionality was rarely used. The representations in the VRA were large enough to read without needing them enlarged. This allowed participants to ignore the zoom functionality.

The second response was the removal of the single representation navigation. Ajax action tracking recorded only nine occasions of the single representation navigation being used. Users stated that they preferred the multiple representation navigation as their method of changing representations. The multiple representation navigation allowed them to change all three representations at one time and to go directly to the representation of their choice. The multiple representation navigation acted as a shortcut for navigating representations. Shortcuts are a type of accelerator that reduce the amount of time required for frequently used functionality (Nielsen, 1994). The multiple representation navigation reduced the amount of time needed to change a representation when compared to the single representation navigation.

The next result was the modification of the triangular layout of the representations. Users stated their dislike of the layout because it required scrolling and did not map intuitively with the answer choice layout. Mapping is the idea of consistency between two sets of two elements that are connected by a specific functionality (Norman, 2002).

The last result was the addition of the image answer selection. It was developed to help improve mapping between the representations and answer choices. If users could select an answer choice by clicking on the representation it corresponded to, then this could reduce errors performed by the user when matching answer choices to representations.

Figure 3.2 shows a view of the second version VRA after modifications were complete.

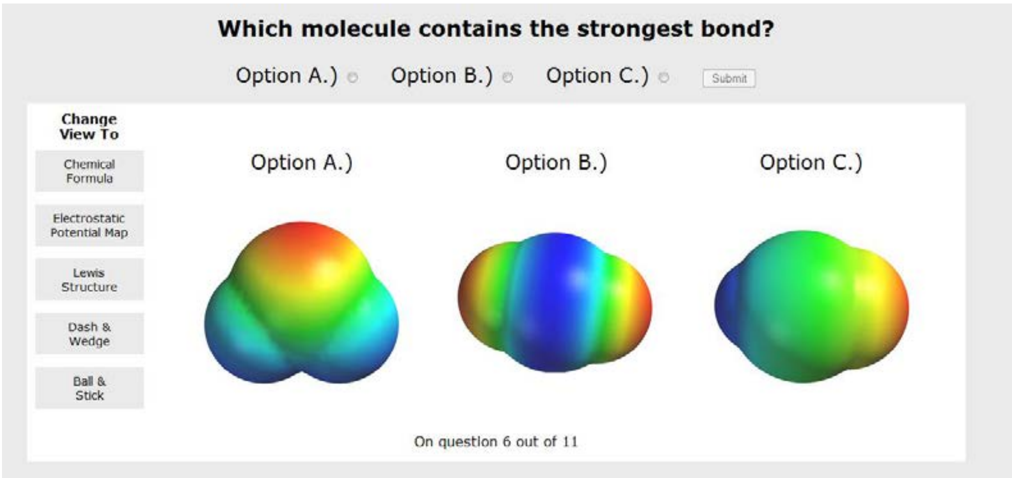


Figure 3.2: Second version of the VRA with a linear layout for representations

Second Version: Formative Evaluation

Results from Ajax action tracking for the second version showed that the 8 participants used the image answer selection on two instances. Participants used the multiple representation navigation a total of 259 occurrences ($M = 32.4$, $SD = 15.1$). Records showed one person had used the multiple representation navigation zero times. It should also be noted that one participant asked about using paper and pencil to draw representations. This participant was informed about the multiple representation navigation capabilities.

The participant who did not use the multiple representation navigation was asked to explain their reasons. They stated that the labels on the multiple representation navigation appeared to be different categories of problems to be completed. When the participant was answering questions, they thought they were given the chemical formula for each molecule because that was the current category of assessment items. The participant also reported that the tutorial did not aid them in learning about the multiple representation navigation and how to use it.

When participants' preference for selecting answer choices was investigated, many reported preferring the traditional answer choice layout on the top of the VRA left of the submit button compared to the image answer selection. This was largely due to users not knowing that the image answer selection existed. After given a description of the image answer selection, many participants thought the idea had merit. One participant stated that it could help individuals select their answer because the images would help discriminate the answer a user would try to select with their mouse.

The final issue was related to the linear layout of the representations. One participant reported that the misalignment of the representations and answer choices created difficulties for them while selecting answer choices.

Second Version: Summative Evaluation

The average SUS score of the second version of the VRA was 91.6 (SD = 6.0). The Wilcoxon ranked-sign test was used to evaluate the difference in the two sets of SUS scores of the VRA (Wilcoxon, 1945). With each sample size being small (n <= 10), a normal distribution was not assumed. No statically difference was found between the two groups based a *p*-value of 0.620. This finding was not surprising as the previous score was high, so the room for improvement was hindered.

Second Version: Modifications

Three major modifications were complete after an additional week of analysis and programming. The first result of the second iteration of the usability test was the modification of the tutorial page. The tutorial was redesigned to require users to complete a practice assessment item before proceeding to the main assessment. This created a low stakes environment where users' scores were not affected by their answer on the tutorial problem before advancing to the high stakes environment. Low stakes environments have been shown to be good tutorials (Atwood, 2014). The new tutorial also acted as a constraint to slow down the user and encourage them to read the tutorial's information. For the purpose of usability studies, a constraint is a design choice that requires users to behave in a specific way (Norman, 2002).

The second response was the modification to the multiple representation navigation. The buttons were updated to have more contrast with the gray background. This was done by coloring the buttons green and by adding a red border around each button corresponding to its state of use. These additions were made to spur users into attempting to use the multiple representation navigation. This changed was inspired when the participant who did not use the multiple representation navigation stated their misunderstanding of the buttons as basic labels.

The last response was the image answer selection modification. While many participants reported they did not know that the image answer selection existed, they stated that the concept was admirable. To address the image answer selection's discoverability, visual cues were added. Discoverability is the extent or degree to which a feature can be identified by a user (Norman, 2002). The first cue was that highlights appeared whenever a user's mouse cursor hovered over an image. The highlight was yellow if the answer was selectable and permanently gray if the answer was already selected. The second cue was the combining of the image answer selection and the buttons from the top section of the interface after their removal. This visual addressed the issue of misalignment between the answer choices and their corresponding representation.

Figure 3.3 shows a view of the third version VRA after modifications were complete.

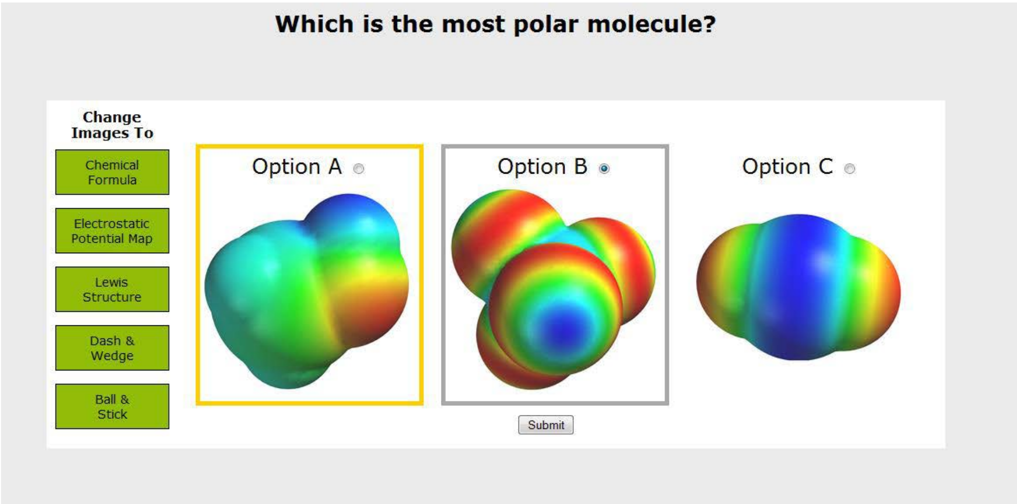


Figure 3.3: Third version of the VRA with the updated multiple representation navigation

Third Version: Formative Evaluation

All participants were recorded using the multiple representation navigation for a total of 118 instances (M = 23.6, SD = 12.6).

Qualitative data collected from interviews did not reveal any possible usability issues.

Third Version: Summative Evaluation

The average SUS score of the third version of the VRA of 90.4 (SD = 6.0). A Kruskal-Wallis rank sum test was performed to find any differences between the three different sets of SUS scores for the VRA (Kruskal & Wallis, 1952). A normal distribution was not assumed as each sample size was at or under 10 participants. No statistical difference was found between the three groups based a *p*-value of 0.856. All three groups of participants generally had the same opinion of the VRA.

Third Version: Modifications

No modifications to the interface of the VRA were made based on the results from the third version usability test. As all participants were recorded as using the multiple representation navigation, no new usability issues were discovered and no statistical difference was found between the SUS scores, goals of the usability test were deemed satisfied.

Discussion of Usability Test

As a result of usability testing done with the VRA, despite scoring high on the SUS on all three version of the VRA, multiple usability issues were discovered. To address the issues discovered during usability testing of the first and second version of the VRA, several modifications were made. These included removing unused functionality, a redesign of the tutorial page, and visual updates to the multiple representation navigation. After these modifications were made, it was assumed there were no more usability issues based on the results of the usability test with the third version of the VRA.

Interest in the VRA

The goal of the VRA was to be a tool that professors could use to gain instructional feedback on their students' ability to use different chemistry representations. To assess if professors would be interested in using the VRA and what that interest was associated with, a survey was designed and given as part of a needs assessment survey study (Reviere, Berkowit., Carter, & Graves 1996) given out yearly by the American Chemical Society – Exams Institute (ACS-EI) (Emenike et al., 2013; Linenberger & Holme 2014). The focus of the survey was to better understand chemistry professors' thoughts on scientific practices, if scientific practices could be assessed, and opinions on novel concepts in chemistry.

Survey

The needs assessment survey began with demographic questions. Participants were asked their institution's location, the type of degrees their institution afforded to students, and the course they taught the most frequently in the past 5 years. Additional questions were present but not reported in this study.

Questions related to the VRA began by asking participants to report how frequently they used chemical formula, Lewis structure, dash-and-wedge model, ball-and-stick model, and electrostatic potential map representations in their exams. Afterwards, a primer briefly explained what the VRA was and looked like. Following the primer, participants were asked to report their interest level in the VRA and if they thought that allowing their students to use a navigation system similar to the VRA would help or hurt them. Lastly, there was an area for participants to give additional comments if they wished. The questions on the survey related to the VRA can be found in Appendix J.

Participants

2,352 professors completed the needs assessment survey in the spring of 2015. Chemistry faculty members from different institutions across the United States of America were emailed inviting them to participate in the online survey. The faculty members' emails were part of a database maintained by the ACS-EI using public information from institutional websites. Faculty members who were emailed, were incentivized to complete the survey in order to be part of random drawing for an Apple IPad. Only those 2,177 who fully completed the VRA questions were included in the analysis reported here.

Participants from the survey covered every state in the United States plus the District of Columbia and Puerto Rico. Table 3.1 shows the type of degrees offered by participants' institutions and their primary chemistry course that they taught.

Table 3.1: Demographic Characteristics of Participants

Characteristics	*n*	%
Degrees by Institution		
Teaching Certification	150	6.9
Associate	285	13.1
Bachelor	1525	70.1
Master	709	32.6
Doctoral	801	36.8
Chemistry Course		
Biochemistry	70	3.2
General	1696	77.9
Inorganic	27	1.2
Instrumental analysis	22	1.0
Organic	264	12.1
Physical	44	2.0
Quantitative analysis	27	1.2
Other	27	1.2

Procedure

Participants were informed the survey took 15 to 30 minutes to complete. Before participants started the online survey, consent was required for the study associated with the survey. Once consent was given, participants could answer any questions they felt comfortable with. A formal time limit of two weeks was given to complete the survey. One week before the deadline, participants were given a reminder stating that they could complete the survey if they desired.

Results of Survey

Tables 3.2 shows the results of how participants self-reported how often they used specific representations on their exams. Each response option was coded as a numerical value in order to calculate an average. “Never” was coded as one, “rarely” was coded as two, and "occasionally" was coded as three, and "frequently" was coded as four. Using this method, chemical formula (M = 3.89) was the most used representation on exams followed by the Lewis structure (M = 3.62) followed by dash-and-wedge models (M = 3.43). Ball-and-stick models (M = 2.53) and electrostatic potential maps (M = 1.90) were the least used.

Table 3.2: Results of How Often Participants Use Representations on Their Exams

Representation	Never	Rarely	Occasionally	Frequently
Chemical formula	9	27	163	1978
Lewis structure	50	118	442	1567
Dash-and-wedge	69	199	641	1268
Ball-and-stick	401	636	725	415
Electrostatic potential map	896	715	438	128

Interest in the VRA was self-reported as following: not interested (n = 199), somewhat interested (n = 720), interested (n = 785), and very interested (n = 473). A similar method of coding as performed with the representation on exam response was complete with the interest level responses. "Not interested" was coded as one, "somewhat interested" as two, "interested" as three, and "very interested" as four. An average of 2.70 was calculated.

Participants' opinions of wether if the VRA's navigation would help or hurt their students was as following: helps (n = 1667), hurts (n = 107), and neither (n = 403). An average for the data was found by coding the results as binary values. The focus of this question was to highlight if participants believed the VRA's navigation was helpful or not. "Help" was coded as one and "hurts" or "neither" were coded as zero. Afterwards an average of 0.76 was calculated.

Since general chemistry students were the intended users of the VRA, a second analysis was completed using only participants who reported their primary class during the last 5 years as general chemistry. The second analysis found that chemical formula (M = 3.91) was the most used representation on exams for general chemistry professors followed by Lewis structure (M = 3.73) then dash-and-wedge models (M = 3.39). Ball-and-stick models (M = 2.57) and electrostatic potential maps (M = 1.91) were again the least used. Table 3.3 shows the results how of general chemistry participants reported how they use representations on their exams.

Table 3.3: Results of How Often General Chemistry Participants Use Representations on Their Exams

Representation	Never	Rarely	Occasionally	Frequently
Chemical formula	3	17	111	1565
Lewis structure	13	49	321	1313
Dash-and-wedge	44	173	552	927
Ball-and-stick	227	507	581	331
Electrostatic potential map	683	579	336	98

Interest in the VRA for general chemistry participants was as follows: not interested (n = 144), somewhat interested (n = 531), interested (n = 625), and really interested (n = 473). The interest average was 2.75.

General chemistry participants' opinions of whether if the VRA's navigation would help or hurt their students was as follows: helps (n = 1292), hurts (n = 84), and neither (n = 320). An average of 0.76 was calculated.

To find if a relationship was present with how participants used representations on their exams and their interest level in the VRA, a Chi-Square test was performed with each representation and interest level (Pearson 1900). If a significance was found, a polychoric correlation was calculated. A polychoric is a correlation between two sets of ordinal variables and results in a value from +1 to -1 indicating a positive or negative relationship (Olsson 1979). Table 3.4 shows the results of each Chi-Square test and polychoric correlation. Each Chi-Square test showed a statistical significance. Chemical formula, Lewis structure, and dash-and-wedge showed weak correlations at or below 0.14. Ball-and-stick and electrostatic potential map showed slightly stronger correlations with values at 0.22 or higher.

Table 3.4: Results of Correlating Participants' Interest with Representations on Exams

Representation	x^2	*p*-value	Polychoric Correlation
Chemical formula	25.71	0.0022	0.14
Lewis structure	41.95	< .0001	0.14
Dash-and-wedge	47.59	< .0001	0.11
Ball-and-stick model	118.39	< .0001	0.22
Electrostatic potential map	115.61	< .0001	0.25

A Chi-Square test was performed to investigate the relationship between participants' opinions on the VRA's navigation and interest in the VRA. If a significance was present, then a rank biserial correlation was calculated. A rank biserial correlation is the strength of a relationship between a set of ordinal of values and a set of binary values (Cureton 1956). The correlation ranges from +1 to -1 indicating a positive or negative relationship. The Chi-Square test produced a significant result of $x^2 = 517.06$; p-value $< .0001$. A rank biserial correlation of 0.63 was calculated.

The statistical analysis performed for all participants was also performed for all participants who reported their primary class during the last 5 years as general chemistry. Tables 3.5 shows the results of each Chi-Square test and polychoric correlation for each representation. Similar to the results of the statistical analysis with all participants, chemical formula, Lewis structure, and dash-and-wedge showed weak correlations while ball-and-stick models and electrostatic potential maps showed slightly stronger correlations.

Table 3.5: Results of Correlating General Chemistry Participants' Interest with Representations on Exams

Representation	x^2	*p*-value	Polychoric Correlation
Chemical formula	22.64	0.0070	0.11
Lewis structure	23.76	0.0046	0.08
Dash-and-wedge	32.43	0.0002	0.11
Ball-and-stick model	112.86	< .0001	0.25
Electrostatic potential map	114.44	< .0001	0.28

The Chi-Square test between general chemistry participants' thoughts regarding the VRA's navigation and interest in the VRA produced a significant result of $x^2 = 415.71$; p-value $< .0001$. A rank biserial correlation of 0.64 was calculated.

Discussion of Survey

Several results were discovered with regards to general interest of the VRA and the different factors of the application. First, the interest level of all chemistry faculty members and those who report themselves as primarily teaching general chemistry in the VRA was always positively correlated with the different representations they use on their exams. The most notable of the representations was the electrostatic potential map and the ball-and-stick model. One reason for this may because creating highly detailed electrostatic potential maps and ball-and-stick models requires expensive software instructors may not have access to.

The second result observed was the large correlation between interest and chemistry faulty members' opinions on the VRA's navigation for their students. The ability to navigate different representations is a key feature of the VRA and to have a correlation of over 0.6 with an individuals' interest in the VRA is a piece of evidence that the development of the VRA is in line with faculty members' opinions.

Summary

This chapter presented the Variable Representation Assessment, a tool developed to properly measure how students use representations in general chemistry. The development of the VRA was completed using a variety of usability methods to ensure users could properly use it, which in turn allowed the VRA to be a proper method of measurement. Several usability issues were found to affect how users interacted with the VRA and modifications were made to address them. After the usability test, a survey was given to assess chemistry faculty members' opinions on the VRA. Data analysis indicated there was a positive relationship between interest in the VRA and several factors present within the VRA with the most important possibly being the navigation system.

CHAPTER 4

PHASE TWO: MEASURING HOW STUDENTS USE REPRESENTATIONS AS A METHOD OF IMPROVING A WEB-BASED ASSESSMENT

There are several factors that should be considered when developing an academic website. In the case of the Variable Representation Assessment (VRA), the first factor to be examined was its usability. The next factor, which this chapter will discuss, is the state of its assessment materials. This was done by completing a pilot study and a full study where students from general chemistry courses from two Midwestern universities used the VRA. The difficulty and discrimination along with time measurements of how they used chemistry representations were analyzed as means to improve the quality of the VRA's assessment. Usability was also measured as a continued effort to ensure users were able to effectively use it.

Pilot Study

The first study completed with the VRA was a full-scale pilot study that was conducted to inform the direction of future assessment content for the VRA. This experiment was exploratory in nature as there is no literature related to how general chemistry students would use representations to solve chemistry problems when given several representations to view. The following sections will highlight the modifications made to the VRA during the usability test described in Chapter 3 and the pilot study, the results of the pilot study, and steps going forward based on those results.

Components of the VRA

Pre-Survey

There were no modifications made to the pre-survey between the usability test and pilot study.

Interface

The first modification made to the VRA's interface before the start of the pilot study was to the image layout. Users started on a set of blank images instead of a set of chemical formula images at the start of each question. This was done to prevent confounding results due to differences in the amount of time a representation image was used. Activities such as reading the stem of a question would increase the amount of time recorded for whatever representation was used at the beginning of each item. During analysis, the time for such activities could be assumed to be the time spent on the blank representations increasing the accuracy of the results and analysis of time measurements. Corresponding multiple representation navigation buttons for blank images were added also to the left hand side of the interface as part of the change.

The second modification to the interface was a visual addition to the left-hand side multiple representation navigation buttons. A red outline was present for a button whenever the corresponding representation was present. This was done to allow the program to be more consistent between the buttons and the representation visible at any given time.

The last modification to the interface was to the electrostatic potential map images. The settings used to create them using Spartan 2008 software were changed from the usability. Settings of a 0.004 isoValue and -200 to 180 kJ/Mol scale were used to create a more consistent visual representation where red corresponded to negative charge and blue corresponded to positive charge. Similar settings have been used previously for electrostatic potential map representations (Williamson et al., 2013). Figure 4.1 shows the interface of the VRA after the modifications were complete.

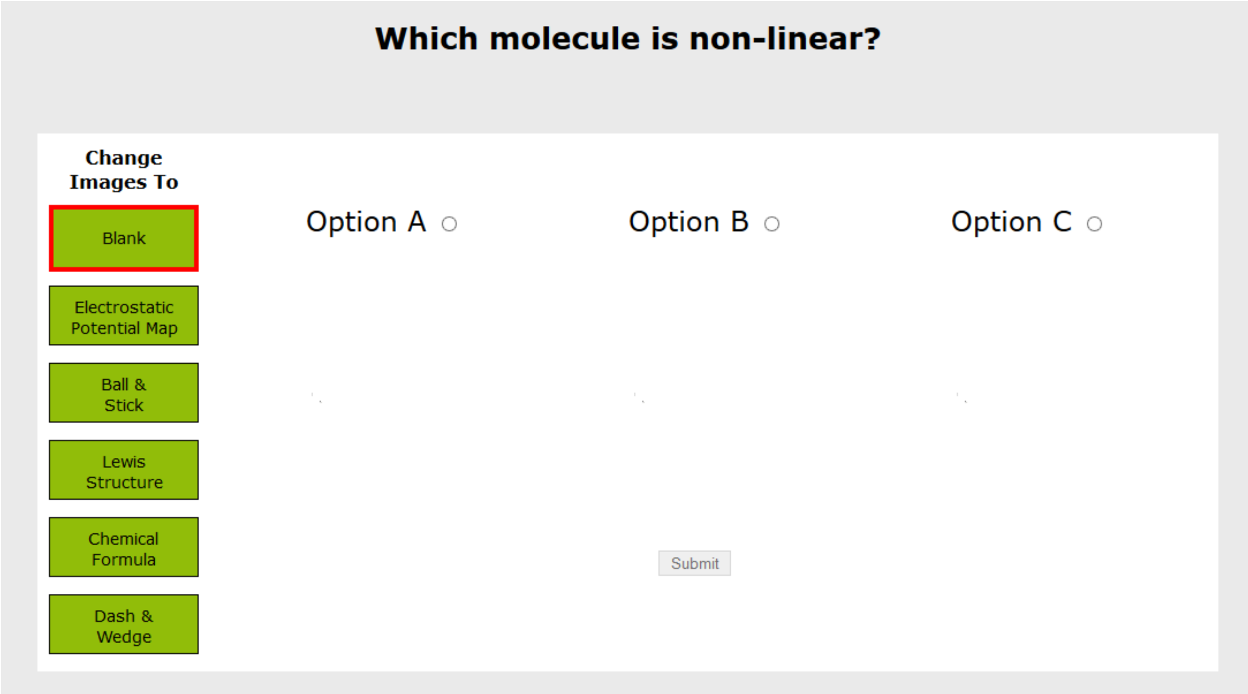


Figure 4.1: The VRA after modifications were complete before the pilot study

Assessment

Several questions were modified to be more grammatically correct. All questions, corresponding answer choices, and corresponding representation images for the pilot study can found in Appendix K.

Post-Survey

In addition to the SUS (Brooke, 1996), users could write comments regarding their thoughts in a textbox after the assessment.

Participants

A total of 327 students from a second semester general chemistry at a large Midwestern university during the spring of 2015 used the VRA as part of an in-class activity. Before the activity, all students were informed who the key personnel involved in the study were, the need for them to bring a laptop or tablet to class to participate, and the compensation of extra credit for participating. A paper alternative was available for the purposes of earning extra credit for those who did not bring a laptop or tablet. The paper alternative was not analyzed as part of the study. The paper alternative can found in Appendix L.

All 327 participants stated they were 18 years of age or older and agreed to the study consent from. 14 participants were removed from analysis using list wise deletion (Allison & Data, 2001) as they did not fully complete the activity leaving a total of 313.

Participants completed a pre-survey where they reported whether they had ever used a specific representation in order to see what representations they used most frequently. If participants reported themselves as using a specific representation, they were also asked when they first used the representation and how frequently. Table 4.1 shows how participants reported themselves as using representations that appeared on the VRA. Table 4.2 shows when participants reported themselves as first using each representation. Table 4.3 shows what participants reported as frequency for using each representation.

Table 4.1: Number of Students Based on Whether They Had Previously Used Each Representation

Representation	Yes	No
Chemical formula	311	2
Lewis structure	310	3
Dash-and-wedge model	202	111
Ball-and-stick model	239	74
Electrostatic potential map	95	218

Table 4.2: Number of Students Based on When They had First Used Each Representation

Representation	Middle School	High School	College	Other
Chemical formula	97	203	8	3
Lewis structure	41	228	38	3
Dash-and-wedge model	4	103	95	0
Ball-and-stick model	36	162	40	1
Electrostatic potential map	1	46	47	1

Table 4.3: Frequency of Use of Each Representation Among Study Participants

Representation	Never	Rarely	Occasionally	Frequently
Chemical formula	3	21	109	178
Lewis structure	1	41	203	65
Dash-and-wedge model	10	79	99	14
Ball-and-stick model	18	119	87	15
Electrostatic potential map	4	73	17	1

Procedure

Two days before the in-class activity during lecture, possible participants were informed to bring a laptop or tablet if they wanted to participate in the study. During the day of the activity, a test administrator visited each section, gave a verbal explanation of the consent form, and answered students' questions as needed. Once questions were answered, participants were given a username and URL with which they could access the VRA if they brought a laptop or tablet. The paper alternative was given out as needed. Participants were given approximately 10 minutes in class to read the consent form and to use the VRA. The test administrator for each classroom section was present in order to answer questions as needed for each section. After participants completed the activity, their email addresses were given to the professor of the course for the process of assigning extra credit. An example of the consent form can found in Appendix M.

Results of Pilot Study

Usability Results

Ajax action tracking recorded that all users successfully used the multiple representation navigation. The majority of participants used the multiple representation navigation on their first question. The remaining participants used it first on the second, third, or seventh question. Table 4.4 shows the question on that participants first used the multiple representation navigation and how many did use the navigation on those questions.

Table 4.4: Question on Which Participants First Used the Multiple Representation Navigation

Navigation Used On Which Question	n	%
First	307	98.1
Second	4	1.3
Third	1	0.3
Seventh	1	0.3

The average SUS score for the pilot study was 78.2 (SD = 15.3). There was a drop of 12.8 in the average SUS score from the usability test.

The majority of comments from the post-survey discussed how participants enjoyed the VRA and having the ability to change to different representations as needed. Several comments questioned the purpose of the blank option. One participant stated that test was initially confusing to use. Another stated that the wording for some questions were too ambiguous without stating what questions were problematic. The remaining comments gave suggestions for the VRA such as the rotation of the 3D representation images.

Assessment Results

The difficulty and discrimination indices of each VRA question were calculated. Difficulty is a measure of how many participants correctly answered a specific item (Adams & Wieman, 2011; Ding & Beichner, 2009). Discrimination is a measure of how well a question distinguishes between higher and low performing students with regards to their assessment score (Adams & Wieman, 2011; Ding & Beichner, 2009). This type of evaluation was performed in order to see what questions caused participants to struggle. In addition, this evaluation helped identify problem items that could be altered for future versions of the assessment. Figure 4.2 shows the difficulty and discrimination indices for each question of the VRA.

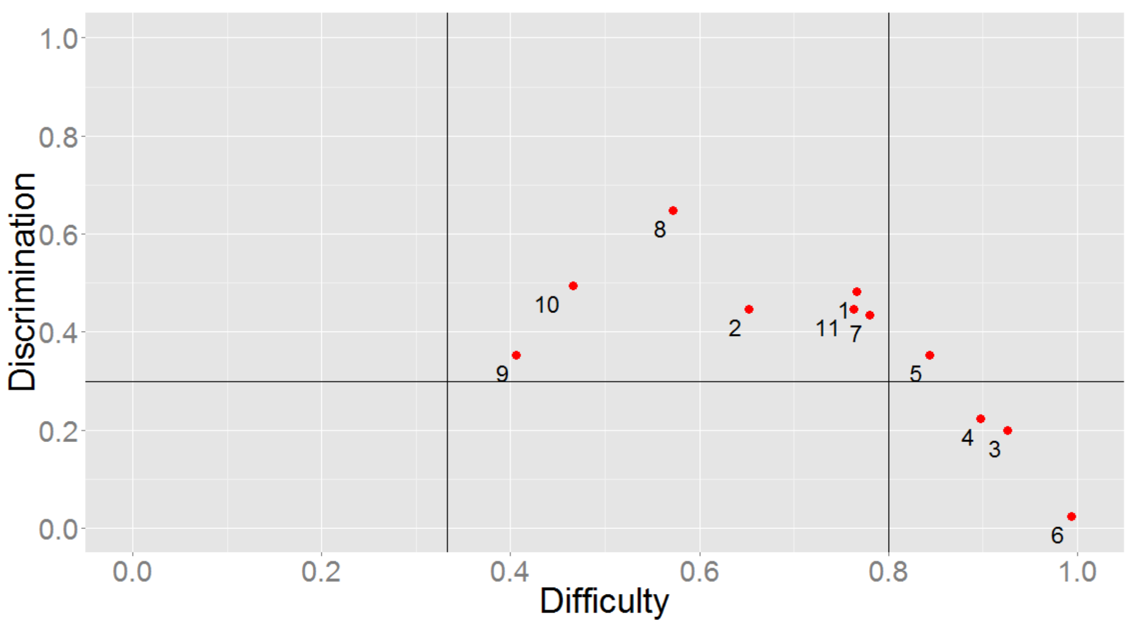


Figure 4.2: Difficulty and discrimination indices for each of the 11 items for the pilot study

The difficulty of a question can range from 0 to 1. Questions with a lower value are considered harder items while higher valued questions are considered easier. Figure 4.2 shows that the difficulty of VRA questions ranged from near about 0.4 to near 1.0. Ideally questions would be in the range of 0.33 to 0.8 range for difficulty. 0.33 is the level of guessing for a multiple choice question with three answers. Based on the difficultly indices, most difficult question was 9. This question asked, “Which molecule has the unknown halogen with the largest partially negative charge?” This question could only be logically answered by using electrostatic potential map representation images. Question 9, its answers, and its representation images can be found in Appendix K.

Item discrimination is the calculated difference between the top and bottom 27% individuals determined by total score who answered the question correctly as a percent (Adams & Wieman, 2011; Ding & Beichner, 2009). The discrimination of a question ranges from -1 to +1. Positive values indicate that higher performing students have answered that specific question correctly more often than lower performing students. Values above 0.30 are considered to have a high ability of distinguishing between higher and low performing students (Doran, 1980). Most VRA questions were above 0.30 with the exception of three questions: 3, 4, and 6. This set of questions can be seen in the bottom right hand corner of Figure 4.2. These questions were too easy for participants resulting in poor discrimination.

Inspection of participants’ total scores revealed all participants earned a score of 3 or higher on the assessment. This may be because several of the questions had a difficulty of 0.89 and higher. The average score was 8.1 (SD = 1.7). In general, the assessment was easy as illustrated by the histogram of total scores (Figure 4.3) that is negatively skewed. The purpose for discussing participants' total scores during the pilot study is for discussion later in the chapter.

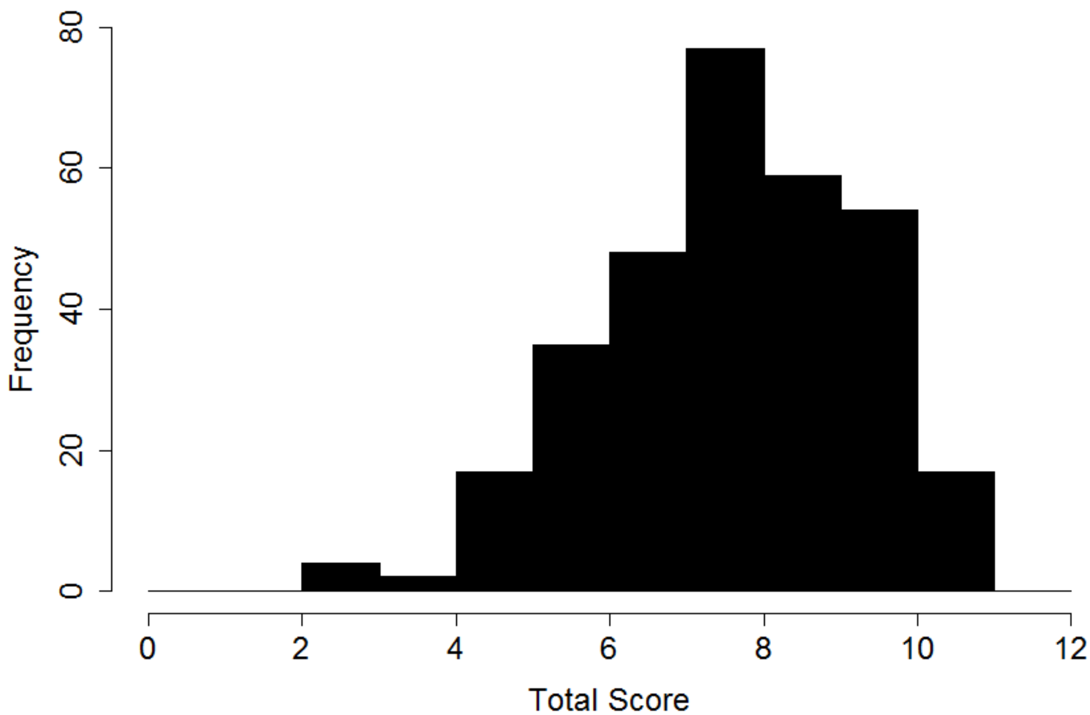


Figure 4.3: Distribution of participants' total scores for the pilot study

Time Measurement Results

The focus of the VRA's time measurements was to determine which representation was the most used for each question. The first statistics calculated were the mean and standard deviation for the amount of time each representation was used for all questions. These were calculated using Ajax action tracking records that are described in Chapter 3. Non-normal distribution was assumed for the measurements as most representation / question combination had a standard deviation larger than their mean. The Friedman Test was used to statistically analyze the measurements (Friedman, 1937). The Friedman Test is the non-parametric alternative to the one-way repeated measures ANOVA and determined whether a statistically significant difference was present among any of the representation's total used time amounts. If a difference was detected, pairwise comparisons using the Wilcoxon signed rank tests were performed to determine where these differences lie (Wilcoxon, 1945). Bonferroni corrections were used in the multiple pairwise comparisons to account for the higher possibility of Type I errors (Bonferroni, 1936).

Table 4.5 shows the mean, standard deviation, and statistical analysis for the amount of time each representation was used for all questions. While the Friedman Test was calculated for each question, the primary focus of the analysis was to determine which representation was the most used for each question. The pairwise comparisons using Wilcoxon signed rank test and Bonferroni corrections *p*-value between statistically significant difference between the most used on average representation and second most used on average representation is reported for that position. The only questions where Lewis structure was not the most used representation on average were question 5, "Which molecule is organic by definition," and question 9, "Which molecule has the unknown halogen with the largest partially negative charge?" For both instances, Lewis structure was the second used most used representation but no statistically significant difference found between with the most used representation for each question.

Table 4.5: Mean, Standard Deviation, and Statistical Analysis of Time Used in Seconds for all Representations for Each Question - All Pilot Study Participants

Question	CF [a]	LS [b]	D&W [c]	B&S [d]	EPM [e]	Total Time	Friedman Test *p*-value	Pairwise Wilcoxon Test *p*-value [f]
1	1.4 (3.1)	**4.1 (5.1)**	**2.6 (4.3)**	1.2 (2.4)	1.2 (2.3)	14.0 (8.5)	< .0001 [g]	< .0001 [g]
2	2.5 (4.6)	**5.7 (6.8)**	**2.8 (5.6)**	1.1 (2.4)	0.9 (2.4)	18.0 (12.5)	< .0001 [g]	< .0001 [g]
3	0.7 (1.8)	**2.3 (4.2)**	1.6 (2.9)	**1.6 (2.6)**	0.5 (1.2)	10.0 (6.2)	< .0001 [g]	0.0580
4	1.1 (4.3)	**3.4 (5.2)**	2.0 (3.5)	**2.4 (3.3)**	0.8 (2.2)	13.1 (7.7)	< .0001 [g]	< .0001 [g]
5	**3.1 (4.4)**	**2.7 (3.9)**	1.6 (3.9)	0.9 (1.8)	0.7 (1.5)	12.3 (7.4)	< .0001 [g]	> .9999
6	0.7 (1.7)	**1.8 (1.8)**	**1.0 (1.8)**	0.6 (1.8)	0.3 (0.9)	7.8 (5.7)	< .0001 [g]	0.0014 [g]
7	1.5 (3.5)	**3.9 (5.9)**	**3.1 (4.6)**	1.5 (2.8)	2.6 (4.5)	16.2 (9.5)	< .0001 [g]	> .9999
8	**4.9 (7.0)**	**5.1 (6.1)**	2.9 (5.6)	1.8 (3.0)	2.0 (4.7)	20.3 (13.5)	< .0001 [g]	> .9999
9	3.7 (5.1)	**6.7 (10.0)**	5.5 (6.5)	3.6 (5.6)	**7.0 (10.0)**	33.3 (16.8)	< .0001 [g]	0.0920
10	4.1 (7.3)	**5.1 (8.2)**	4.1 (7.4)	1.8 (4.7)	3.7 (4.4)	23.7 (15.8)	< .0001 [g]	0.3100
11	**3.0 (5.8)**	**4.4 (6.5)**	1.7 (3.6)	1.8 (3.1)	2.0 (2.5)	17.1 (9.3)	< .0001 [g]	0.0533

[a] CF: chemical formula, [b] LS: Lewis structure, [c] D&W: dash-and-wedge, [d] B&S: ball-and-stick, [e] EPM: electrostatic potential map, [f] With Bonferroni correction between most and second most used viewed representation, [g] Significant *p*-value

Most Used Representation, **Second Most Used Representation**

Analysis of the overall data revealed Lewis structure as the primary representation used by students during the pilot study. Further analysis was completed to investigate whether this trend was true for additional contexts. To do this, four subsets of the overall time measurement data were created for analysis. They were as follows: (1) a subset of time measurements from participants who answered the question correctly, (2) a subset of time measurements from participants who answered the question incorrectly, (3) a subset of time measurements from participants who answered the question fast, and (4) a subset of time measurements from participants who answered the question slow.

Identifying the subsets of fast and slow were inspired by the Dual Processing Theory which states humans have two distinct modes of thought (Evans, 2008). One cognitive mode which is fast, automatic, and without much effort while the other which is slow, reactive, and requires conscious ability. To create the fast and slow subset for each questions, a histogram was created using the R statistical software package with 30 breaks of no set bin size. For the purpose of this analysis, an individuals' time interacting with a particular representation that fell between zero seconds and the highest frequency of time used to complete a question in seconds was considered fast. Any other time was considered slow. There is nothing particularly significant about the time of maximum frequency for student attention. Nonetheless, it is likely that, given the empirical evidence acquired, people beyond this frequency are unlikely to be using the fast processing, heuristic cognitive pathway. Figure 4.4 shows an example of this separation for question 9.

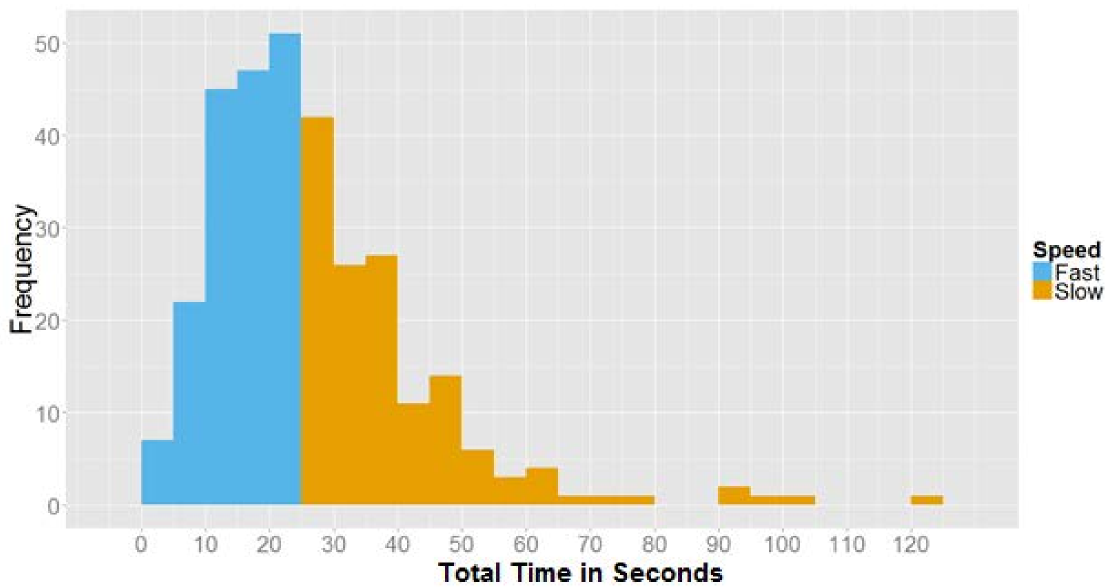


Figure 4.4 Separation of fast and slow participants based on total time for question 9

Similar to the analysis to complete for the overall data, the mean, standard deviation, and statistical analysis for the amount of time each representation was used for all questions for the correct, incorrect, slow, and fast subset. The results can be found in Appendix N, O, P, and Q respectively. The measurements for question 6 in Appendix O are absent as only two participants answered the question incorrectly making it impossible to accurately perform analysis. All four subsets saw Lewis structure as the most used representation or the second most used representation with no statistically significant difference between the most used representation with each question.

Discussion of Pilot Study

Usability of the VRA during the pilot study was satisfactory. The average SUS score for the VRA was 78.2 which is acceptable as based previous research (Bangor, Kortum, & Miller, 2008). In addition, nearly all participants were recorded using the multiple representation navigation which was the primary goal from the usability test. Two questions arose based on the data though. What was the cause of the dramatic drop in SUS score from the final usability test prior to the pilot study and why did only 307 of the 313 participates use the multiple representation navigation on their first question? Participants' comments do not answer these questions.

Results of the time measurement analysis suggest that Lewis structure was participants' primary representation during the pilot study. Lewis structure was calculated as the most used representation for many questions using the full set and subset of time measurements. When the Lewis structure was not the most used representation, it always was the second most used representation with no statistically significant difference found with the most used representation.

Based on these observations, it is unsurprising that when the Lewis structure was not the most appropriate representation for a question, it appeared participants struggled. The most difficult item based on the evaluation of the assessment was question 9 which could only be answered using a non-Lewis structure representation, the electrostatic potential map.

Full Experiment

After the large scale pilot study was complete, modifications were made to the VRA based on the results. Once they were complete, the VRA was administered to two groups of first semester general chemistry students during the spring of 2015 at two large Midwestern universities to see how students from different universities preformed. The universities will hereafter be referred to as University A and University B. The modifications made to the VRA for the full experiment along with its result will be discussed in the following sections.

Components of the VRA

The pre-survey and post-survey were not modified between the pilot study and full experiment. This section will cover the modification to the interface and assessment components of the VRA before the full experiment was administered.

Interface

All arrows representing bond polarities were removed from dash-and-wedge representations. This modification made dash-and-wedge representations more consistent with what is present in classroom textbooks (Brown, LeMay, Bursten, Murphy, & Woodward 2014).

Assessment

Analysis of the pilot study results suggested that general chemistry students primarily relied on Lewis structure for a variety of questions including items that the Lewis structure is not the most appropriate representation to use. The pilot study had one question that was impossible to solve by solely using the Lewis structure. In order to further investigate this phenomena of Lewis structure usage, additional questions were added where it was impossible to solve by solely using the Lewis structure.

In addition to adding new items, several questions were removed as they were too easy based on difficulty indices from the pilot study or they contained content closely related in other items.

The final assessment contained 10 items overall. Six items from the pilot study and four new items. Five assessment items were questions where it was impossible to solve them by solely using the Lewis structure. The other five questions were able to be solved logically using only the Lewis structure.

All questions, corresponding answer choices, and corresponding representation images for the full experiment can found in Appendix R.

Participants

232 students from University A completed the VRA. Of those, 40 participants were removed from analysis for not completing the full activity or showing lack of effort. Lack of effort was operationalized as any participant who started using the multiple representation navigation then at some point stopped using it. This left a total of 192 participants for analysis. The results of University A's pre-survey results can be found in Appendix S.

142 participants from University B completed the VRA. Using the same criteria as University A, 21 participants were removed leaving a total of 121 participants for analysis. The results of University B's pre-survey results can be found in Appendix T.

All participants from both universities stated being 18 years or older and agreeing to the study.

Procedure

A procedure similar to that used in the pilot study was used for the full experiment. Announcements were used to inform all participants at their corresponding university of the key personnel involved in the study, the activity they were being asked to complete, the formal limit of a week to complete the activity, and the compensation of extra credit for participating. After announcements were made, emails were sent to all possible participates giving them a username and URL to access the VRA. After the time limit was reached, students' email addresses were sent to their corresponding instructors for extra credit. The same consent form the pilot study was used for this experiment and can found in Appendix M. An example of the email sent to participants can be found in Appendix U.

Results of Full Experiment

Usability Results

Not all University A and B students were successful in using the multiple representation navigation. Table 4.6 and 4.7 shows what questions University A and B participants first used the multiple representation navigation respectively.

Table 4.6: Number and Percentage of Participants from University A Who Used the Multiple Representation for the First Time on Each Question

Navigation Used On Which Question	n	%
First	188	97.4
Second	4	2.1
Never	1	0.5

Table 4.7: Number and Percentage of Participants from University B Who Used the Multiple Representation for the First Time on Each Question

Navigation Used On Which Question	n	%
First	111	91.7
Second	4	3.3
Sixth	1	0.8
Never	5	4.1

Participants who were recorded as not using the multiple representation navigation were removed from further analysis. Individuals who never changed what representation they used could not supply accurate information as related to the results of the SUS, assessment, or time measurement. Therefore, the number of participants from University A was reduced down to 191 participants and the number of participants from University B was reduced to 116.

The average SUS score for University A was 74.7 (SD = 16.1) and the average SUS score for University B was 66.3 (SD = 17.4).

Many of the comments made during the full experiment were similar to the comments made in the pilot study: enjoyed the system and having the ability to use multiple representations, having confusion about the blank option, and wanting rotation for the 3D objects. Two participants, one from each university, stated that the directions were unclear. One also stated they didn't know when the test started versus the tutorial.

Assessment Results

The majority of items for University A and University B had difficulty levels between 0.33 and 0.8 with the exception of one item above 0.8 difficulty. This was a shift in difficulty from the pilot study where four of the items were over 0.80 difficulty. The assessment had become harder than previously. In terms of results between University A and University B, University B had fewer items above a 0.60 difficulty than University A indicting the assessment items may have been easier for them.

Discrimination indices for all items for University A and University B were above 0.30 indicting all items were able to distinguish between high-performing and low-performing students. This discrimination seems to be a result of the items not being extremely easy or hard in terms of difficulty. These observations about difficulty and discrimination appear to be a result of the modifications done to the questions between the pilot study and the full experiment.

Difficulty versus discrimination graphs for University A and B can be found in Figures 4.4 and 4.5 respectively. For reporting purposes, assessment items used in both the pilot and full study were assigned the same code and items used only for the full study have a new code. Specifically the following questions were used for same for each study: 1, 2, 5, 8, 9, and 10. The following items were used only on the full study: 13, 14, 15, and 16.

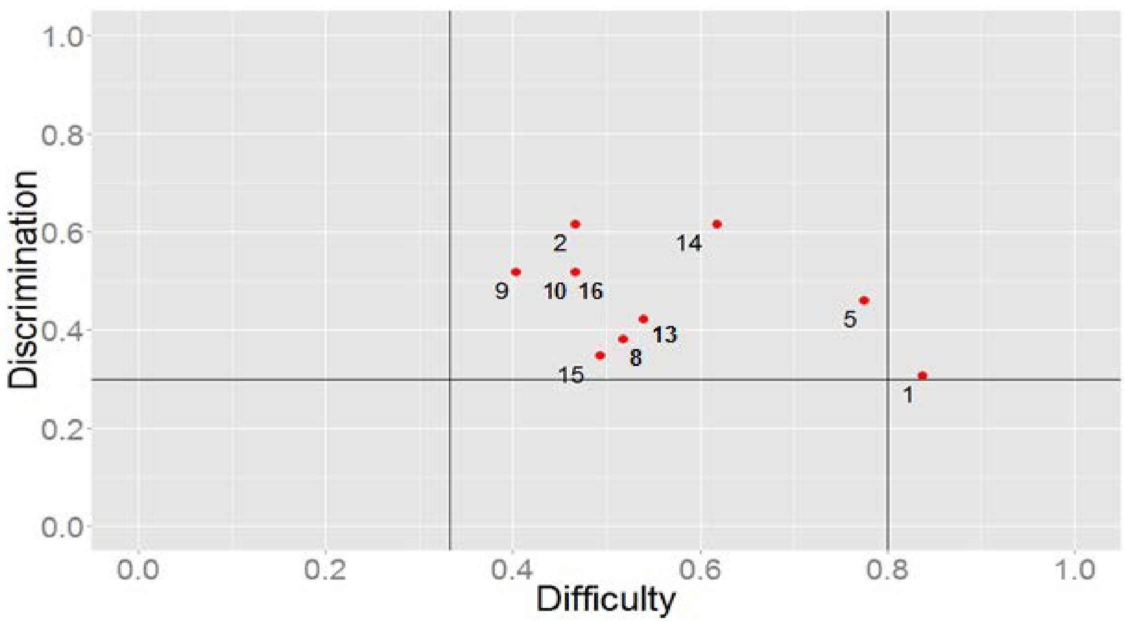


Figure 4.5: Difficulty and discrimination indices for each of the 10 items for the full experiment for University A students

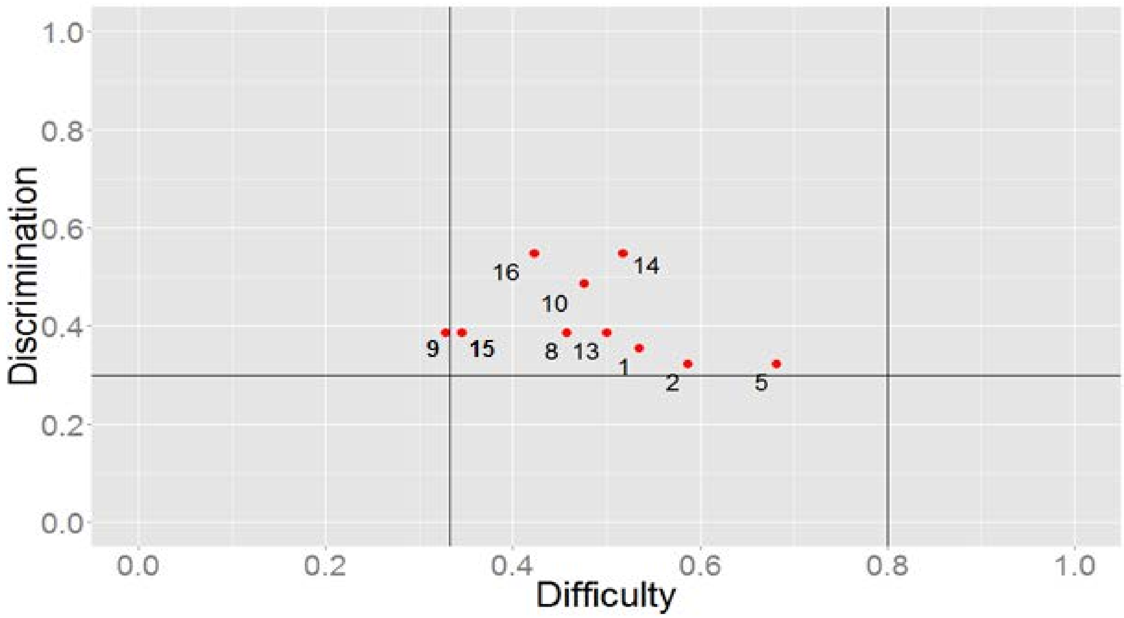


Figure 4.6: Difficulty and discrimination indices for each of the 10 items for the full experiment for University B students

Early in analysis, there was concern regarding whether both universities should be pooled together. One way to address this concern was to compare performance on the VRA, so analysis was performed using the total scores of students from both universities. The average for University A was 5.6 (SD = 1.9) while the University B was 4.9 (SD = 1.7). To statistically compare the two group of scores Welch's T-test was used. The Welch's T-test is a method for comparing two samples uneven in size (Welch 1947). Using Welch's T-test, a statistically significant difference was found between total scores with a *p*-value of 0.0006. Further analysis showed a statistically significant difference between University A between University B when the subsets of questions are compared. For questions where participants could use Lewis structures freely in their efforts to the questions, University A scored 3.1 (SD = 1.2) while

University B scored 2.7 (SD = 1.1). Using Welch's T-test, statistical significance was found based on a p-value of 0.0140. For questions that could not be answered solely with the Lewis structure, University A had an average score of 2.5 (SD = 1.2) and University B scored on average 2.1 (SD = 1.2). Welch's T-test revealed a statistically significant difference based on a *p*-value of 0.0043. Based on this analysis, the groups were not combined. This analysis also shows that students, on average, did better on Lewis questions compared to non-Lewis questions.

The distribution of participants' total scores from University A and B can found in Figure 4.7 and 4.8 respectively.

Since participants' total scores indicted the assessment had become more difficult, it was hypothesized there would may be a relationship between total scores and SUS scores. Participants may have had a hard time distinguishing whether they just did not like the assessment about chemistry and the application itself. A Pearson correlation was calculated to better understand this relationship (Pearson, 1895). University A participants had 0.13 correlation indicting a weak relationship between the two variables. University B participants had a slighter strong correlation at 0.25.

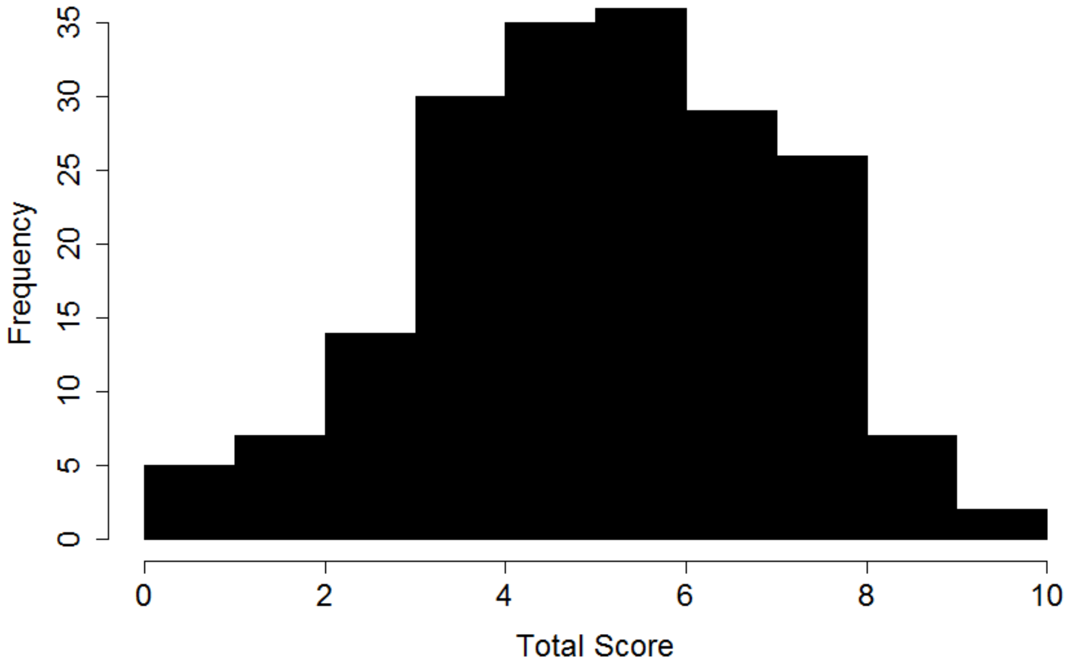


Figure 4.7: Distribution of University A participants' total scores for full experiment

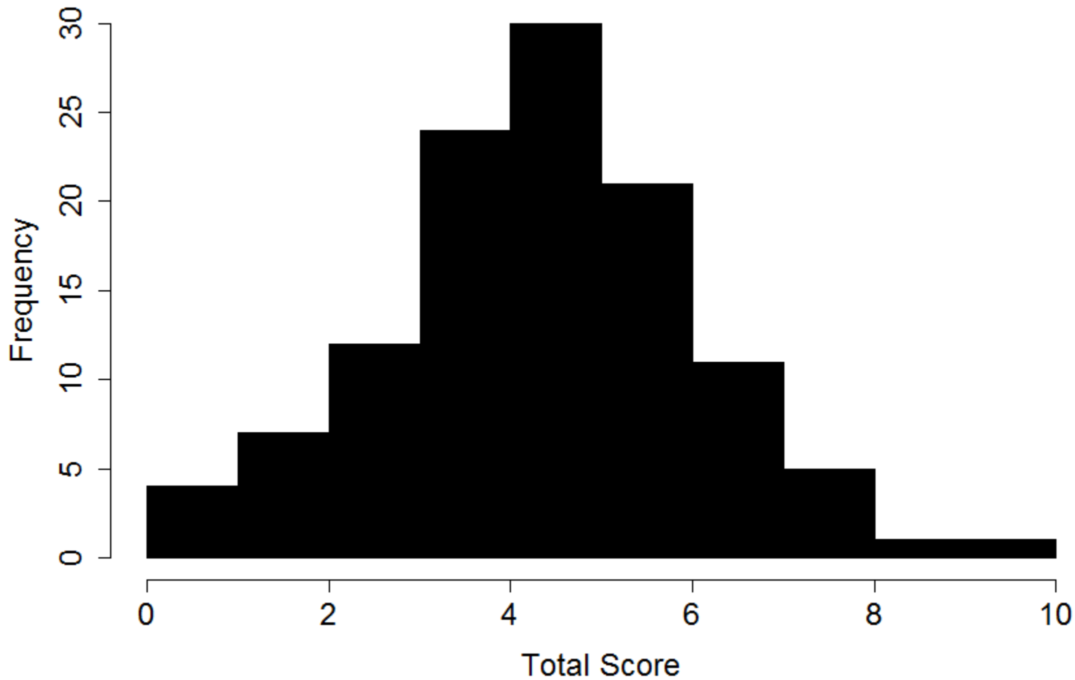


Figure 4.8: Distribution of University B participants' total scores for full experiment

Table 4.8: Mean, Standard Deviation, and Statistical Analysis of Used Time in Seconds for all Representations for Each Question - All University A Participants

Question	CF [a]	LS [b]	D&W [c]	B&S [d]	EPM [e]	Total Time	Friedman Test *p*-value	Pairwise Wilcoxon Test *p*-value [f]
1	1.0 (2.3)	**5.0 (15.1)**	**2.8 (6.8)**	2.4 (16.0)	1.3 (4.6)	16.6 (26.2)	< .0001 [g]	0.0348 [g]
2	4.8 (10.3)	**8.0 (13.9)**	**4.6 (7.5)**	2.3 (8.2)	1.4 (2.9)	26.1 (20.3)	< .0001 [g]	0.0062 [g]
5	**5.7 (10.6)**	**3.9 (14.4)**	2.0 (4.1)	1.1 (3.0)	0.7 (1.7)	17.7 (19.2)	< .0001 [g]	0.0045 [g]
8 [h]	**8.8 (15.6)**	**4.6 (8.1)**	2.8 (4.7)	2.3 (7.8)	1.6 (3.4)	26.1 (27.6)	< .0001 [c]	0.0019 [g]
9	4.3 (7.0)	**7.6 (12.0)**	4.2 (7.5)	3.7 (7.1)	**6.4 (11.3)**	35.0 (26.1)	0.0002 [g]	> .9999
10	**5.7 (14.8)**	**6.6 (14.1)**	2.6 (9.8)	1.4 (3.8)	3.7 (7.9)	25.2 (24.3)	< .0001 [g]	> .9999
13	3.9 (11.5)	**12.9 (21.4)**	**11.9 (15.7)**	10.4 (19.4)	2.4 (6.3)	47.9 (39.7)	< .0001 [g]	> .9999
14	2.0 (3.0)	**10.1 (9.6)**	**3.7 (4.4)**	3.2 (3.9)	2.5 (5.5)	26.2 (15.3)	< .0001 [g]	< .0001 [g]
15 [h]	5.8 (10.9)	**18.0 (32.3)**	**7.9 (15.0)**	5.8 (14.7)	3.5 (7.1)	46.3 (54.6)	< .0001 [g]	< .0001 [g]
16	**9.9 (17.1)**	**15.7 (23.2)**	7.5 (13.6)	4.1 (6.1)	3.0 (6.2)	44.6 (37.5)	< .0001 [g]	0.0070 [g]

[a] CF: chemical formula, [b] LS: Lewis structure, [c] D&W: dash-and-wedge, [d] B&S: ball-and-stick, [e] EPM: electrostatic potential map,
[f] With Bonferroni correction between most and second most used viewed representation, [g] Significant *p*-value, [h] Removed outliner(s)
Most Used Representation, **Second Most Used Representation**

Table 4.9: Mean, Standard Deviation, and Statistical Analysis of Used Time in Seconds for all Representations for Each Question - All University B Participants

Question	CF [a]	LS [b]	D&W [c]	B&S [d]	EPM [e]	Total Time	Friedman Test *p*-value	Pairwise Wilcoxon Test *p*-value [f]
1	3.0 (7.6)	**5.0 (12.2)**	**2.9 (6.2)**	1.6 (2.8)	1.7 (3.4)	18.6 (20.1)	0.0015 [g]	0.5152
2	**5.8 (14.6)**	**6.2 (9.4)**	4.0 (12.7)	2.1 (4.8)	1.0 (2.2)	24.5 (23.4)	< .0001 [g]	0.8089
5	**3.5 (6.4)**	**3.7 (10.2)**	1.7 (3.9)	1.0 (2.3)	0.7 (1.0)	15.1 (16.5)	< .0001 [g]	> .9999
8 [h]	**6.2 (10.1)**	**6.2 (11.4)**	3.0 (5.9)	2.4 (7.2)	1.7 (3.5)	23.9 (18.8)	< .0001 [g]	> .9999
9	3.8 (4.7)	**6.5 (7.9)**	3.7 (7.1)	3.3 (5.3)	**6.2 (10.9)**	31.0 (23.4)	< .0001 [g]	> .9999
10 [h]	5.2 (10.7)	**8.9 (19.1)**	1.7 (3.2)	1.4 (2.7)	3.5 (5.3)	25.7 (24.9)	< .0001 [g]	> .9999
13	3.0 (9.2)	6.5 (12.3)	**6.6 (12.4)**	**6.7 (7.5)**	2.3 (6.7)	35.0 (48.2)	< .0001 [g]	0.4120
14	2.3 (2.9)	**6.5 (8.0)**	3.0 (4.3)	**4.1 (4.7)**	2.6 (3.9)	25.2 (27.6)	< .0001 [g]	0.1399
15	5.1 (6.8)	**18.0 (27.4)**	**6.5 (13.5)**	3.7 (7.0)	3.1 (6.1)	41.2 (36.6)	< .0001 [g]	< .0001 [g]
16	**8.3 (8.8)**	**21.1 (33.7)**	4.9 (7.4)	3.8 (4.9)	2.8 (4.0)	44.9 (37.9)	< .0001 [g]	< .0001 [g]

[a] CF: chemical formula, [b] LS: Lewis structure, [c] D&W: dash-and-wedge, [d] B&S: ball-and-stick, [e] EPM: electrostatic potential map, [f] With Bonferroni correction between most and second most used viewed representation, [g] Significant *p*-value, [h] Removed outliner
Most Used Representation, **Second Most Used Representation**

Time Measurement Results

During the analysis of the time measurements, several outliers were removed. Any participant who required over 500 seconds for a question was removed from the analysis of that question. After outliers were removed, analysis was completed for three sets of time measurements for each university: overall, correct, and incorrect.

Table 4.8 and 4.9 detail the analysis of University A and University B students' time spent on each question. As was seen in the pilot test, students tend to view Lewis structures as the primary representation to answer questions in the VRA. The key numbers to notice in these tables are those that indicate the number of items where the most used is the Lewis structure. Unlike the pilot test, in this experiment there were seven instances where Lewis structure was not the most used representation, nor the second most used representation, or had a statistically significant difference between it and the most used representation when it was the second most used representation. This happened for questions 1, 5, 8, 10, and 13. Question 10 and 13 were questions much could not be answered using only the Lewis structure. Of those questions 1, 5, and 8 could be answered using only the Lewis structure. These can be seen in Appendix V, W, X, and Y where the analysis for correct and incorrect for University A students and correct and incorrect for University B students' sets of time measurements can be found respectively.

Discussion of Full Experiment

Modifying the assessment items for the main experiment had little apparent effect as to what representation participants primarily used. This observation was made despite the fact that the newly added items were specifically designed to reward students for accessing representations other than the Lewis structures. Students from both universities generally used

Lewis structure for every question with the exception of question 13. Question 13 asked students to correctly identify what molecule had the inconsistent representations requiring them to look all representations for the molecule. Students may have been more confident in their ability to identify inconsistences in Lewis structures than other representations. As a result, students spent more time viewing other representations that were not the Lewis structure. Question 13, its answers, and its representation images can be found in Appendix R.

The score on this assessment was, on average, lower than the pilot study. This seems to be a result of two actions: removal questions what were too easy based on pilot study difficulty indices and the addition of questions what required students to not solely rely on the Lewis structure. This overall score result provides additional evidence that general chemistry students struggle when they cannot use Lewis structures.

As for usability, both universities had participants who did not use the multiple representation navigation. Ajax tracking action does not give enough data to figure out why this occurred. One idea is that it is related to the method of administration. During the pilot study, test administrators were in the general area while participants were using the VRA. In the full experiment, participants could use the VRA at any time they chose, possibly creating an environment where participants tried less. In order to find out, qualitative data would need to be collected or the method of administration would need to be methodologically controlled.

SUS scores, since the usability test, declined in each study. Is it the administration of the VRA in one-on-one environment vs. an environment where the test administer in the general area vs. an environment where participants could use the VRA without structure a possible reason? Is the increasing difficulty of the assessment a possible cause? Comments given by participants at the end of the experiment do not give enough detail to prompt what might have caused the decline. Qualitative data, controlling for the method of administration, or better understanding the relationship between total scores and SUS scores may help to answer this.

Summary

This chapter described a pilot study and full experiment, which used Ajax tracking and other analysis to improve a web-based assessment. The results from pilot study suggested that students in general chemistry courses primarily rely on Lewis structure for chemistry questions when given the choice between multiple representations. When the Lewis structure was not the most appropriate for a question, students often answered that question incorrect. This observation suggests that general chemistry students may not be fluent in using representation beyond the Lewis structure. These trends related to Lewis structures were further investigated in the full experiment by modifying the assessment to have more items that could not be answered solely through the use of Lewis structure. The changes resulted in an assessment that was harder than the pilot version and further suggested the general chemistry students do not use a diverse amount of representations beyond the Lewis structure.

CHAPTER 5

CONCLUSIONS

This chapter will summarize the results of this study in several ways. First, the research questions presented in Chapter 1 are answered. Then, the implications for teaching general chemistry are presented along with the limitations of the study. Finally, potential directions for future research based on the results are provided.

Research Questions Answered

1. *How is the usability of the VRA perceived by users?*

Overall, users' perception of the VRA's usability was inconsistent over the two phases of the study. During the initial usability test, despite several changes being made because the results of interviews and Ajax action-tracking, users perceived the VRA as having high usability based on the results of the SUS (Bangor et al., 2008; Brooke, 1996). The SUS scores for each iteration was 90.3, 91.6, and 90.4, respectively. Many of the changes made tended to relate to users' preferences completing their goal with one method over other. An example of this would be users' preference of the multiple representation navigation rather than the single representation navigation since the multiple representation navigation allowed the user to change three representations at once compared to one at a time.

During the second phase of the study, the VRA's usability was not perceived as highly as in the first phase. This is inferred because of a consistent drop in SUS scores dropped with each administration of the VRA during the second phase. A SUS score of 78.2 was calculated for the pilot study while SUS scores of 74.7 and 66.3 were calculated respectively for the universities that participated in the full experiment. In addition, Ajax action-tracking recorded a number of participants not using the multiple representation navigation. The lower SUS scores and individuals failing to use the multiple representation navigation suggested that a change between the first phase and second phase maybe due to a drop in users' perception of the VRA's usability. Alternatively, increasingly difficult assessment content may have influenced users' perception if they did not have clear distinction between the challenge of the overall assessment and its usability.

2. *What does the data produced by the VRA tell us about how general chemistry students use representations?*

The data produced by the VRA revealed that general chemistry students strongly favor using the Lewis structure to answer chemistry questions. This observation can be seen even in instances when the Lewis structure is not the most appropriate representation to use. The average time measurements collected using Ajax action-tracking revealed that the Lewis structure was always the most used representation or the second most used representation while not having a significant difference with the most used representation except in few instances. The small number of exceptions to this observation arise from seven times when the average time measurements were analyzed for participants who answered items quickly or slowly in the full experiment. In these cases, the Lewis structure was third, fourth, or fifth most used representation, or it had a statistically significant difference with the most used representation when it was the second most used representation. The Lewis structure seems to be the only representation many general chemistry students are comfortable with using.

When the difficulty for each question was calculated, items that required the use of representations beyond solely the Lewis structure were never found to have a difficulty index above 0.62. By contrast, questions that could be solely answered using only the Lewis structure were capable of being 0.99 in difficulty. Thus, the evidence suggests that students tend to correctly answer questions that could be solved using only the Lewis structure compared to questions that required using representations beyond solely the Lewis structure.

Implications

There are three broad implications for instructors based on the results on this study. First, the VRA's interface could be a tool with which to deliver new and novel assessments. Currently, the VRA is designed for general chemistry, but could be modified for other content areas, such as higher level chemistry courses for which multiple visualization methods seem to be more important.

Specifically related to general chemistry, it would be beneficial for instructors to know if their students had possible difficulties using chemistry representations beyond the Lewis structure. If instructors could use the VRA to know a difficulty was present in their classroom, they could adjust what they teach. For example, and keeping in mind that the amount of time classes have is limited, instructors could adjust their courses to diversify the number of representations they teach in their classroom.

Lastly, this study seems to suggest that students, such as novices in chemistry, become comfortable using one specific skill and then keeping using that skill for every similar situation they encounter.

Limitations

While this study used a novel technique to investigate general chemistry students' use of chemistry representations, its results should be interpreted with certain limitations in mind.

While the usability test was successful in refining the VRA's interface, several participants (6 from the pilot study and 15 from the full experiment) did not choose to use the multiple representation navigation during their respective experiments. In addition, average SUS scores ranged from 66.3 to 91.3 during the various phases of the study. This inconsistency could be the result of several factors such as the different methods of administering the assessment (e.g. one researcher with one student during the usability test versus one researcher with several students during the pilot test versus no formal structure during the full experiment). For future use with the VRA, this inconsistency should be addressed before proceeding.

Another limitation with the study is that this study only included students from two Midwestern universities. The findings and results of the study are limited by those who were involved. Diversifying the sample with general chemical students from other universities would address with this limitation.

Future Research

One direction for future research could be to administer the VRA in a high stakes environment instead of a low stakes environment. The data gathered in this fashion would be more realistic since individuals' effort would directly affect their respective course grades.

While this study did identify trends in how students use representations in general chemistry, the reasons for why these trends occurs cannot be explained with the data collected from the study. Qualitative measures, such as think alouds where participants talk about their actions (Virzi, Sorce, & Herbert, 1993), would help to explain why these trends appeared. With a better understanding of why these trends occur, proper steps could be taken to address or change what is causing them to happen if needed.

Another possible direction for research would be to administer the VRA to organic or higher level chemistry students to see what patterns appear. Identifying patterns that are and are not shared between general chemistry and higher level students, when using the VRA, could be valuable in understanding how students transform their thinking the further they are in a chemistry curriculum.

REFERENCES

Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. International Journal of Science Education, 33(9), 1289-1312.

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

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183-198.

Allison, P. D., & Data, M. (2001). Sage. *Thousand Oaks, CA: University Papers Series on Quantitative Applications in the Social Sciences*.

Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *Educational studies in mathematics*, *52*(3), 215-241.

Atterer, R., Wnuk, M., & Schmidt, A. (2006, May). Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In *Proceedings of the 15th international conference on World Wide Web* (pp. 203-212). ACM.

Atterer, R., & Schmidt, A. (2007, April). Tracking the interaction of users with AJAX applications for usability testing. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1347-1350). ACM.

Atwood, Jeff. (October 9, 2014). Level One: The Intro Stage. Retrieved from http://blog.codinghorror.com/level-one-the-intro-stage/.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, *24*(6), 574-594.

Bauer, B. (2009). Marks, Introduction to Chemistry.

Bergqvist, A., Drechsler, M., De Jong, O., & Rundgren, S. N. C. (2013). Representations of chemical bonding models in school textbooks–help or hindrance for understanding? *Chemistry Education Research and Practice, 14*(4), 589-606.

Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilita. Libreria internazionale Seeber.

Boyle, A., & Hutchison, D. (2009). Sophisticated tasks in e-assessment: what are they and what are their benefits?. *Assessment & Evaluation in Higher Education*, *34*(3), 305-319.

Brandriet, A., & Holme T. A. (March 22, 2015). Creating the Exams Data Analysis Spreadsheet (EDAS) as a tool to help instructors conduct customizable analyses of student ACS exam data and compare the results to national normative statistics. Paper presented at the 249th American Chemical Society National Meeting in Denver, CO.

Brenner, M. E., Mayer, R. E., Moseley, B., Brar, T., Duran, R., Reed, B. S., & Webb, D. (1997). Learning by understanding: The role of multiple representations in learning algebra. *American Educational Research Journal, 34*(4), 663-689.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry, 189*(194), 4-7.

Brown, T. E., LeMay, H. E. H., Bursten, B. E., Murphy, C., & Woodward, P. (2014). *Chemistry The Central Science 13th Edition*. Prentice Hall.

Cooper, M. M., Underwood, S. M., & Hilley, C. Z. (2012). Development and validation of the implicit information from Lewis structures instrument (IILSI): do students connect structures with properties? *Chemistry Education Research and Practice*, *13*(3), 195-200.

Crisp, G. (2002). Using java applets to help make online assessment interactive. In *ASCILITE* (pp. 799-802).

Crowther, M. S., Keller, C. C., & Waddoups, G. L. (2004). Improving the quality and effectiveness of computer-mediated instruction through usability evaluations. *British Journal of Educational Technology*, *35*(3), 289-303.

Cureton, E. E. (1956). Rank-biserial correlation. *Psychometrika*, *21*(3), 287-290.

Dermo, J. (2009). E-Assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology*, 40(2), 203-214.

Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. Physical Review Special Topics-Physics Education Research, 5(2), 020103.

Doran, R. L. (1980). Basic Measurement and Evaluation of Science Instruction. National Science Teachers Association, 1742 Connecticut Ave., NW, Washington, DC 20009 (Stock No. 471-14764; no price quoted).

Emenike, M. E., Schroeder, J., Murphy, K., & Holme, T. (2013). Results from a national needs a assessment survey: A view of assessment efforts within chemistry departments. *Journal of Chemical Education*, *90*(5), 561-567.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. Annu. Rev. Psychol., 59, 255-278.

Fredlund, T., Airey, J., & Linder, C. (2012). Exploring the role of physics representations: an illustrative example from students sharing knowledge about refraction. *European journal of physics*, *33*(3), 657.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association, 32(200), 675-701.

Hinze, S. R., Rapp, D. N., Williamson, V. M., Shultz, M. J., Deslongchamps, G., & Williamson, K. C. (2013). Beyond ball-and-stick: Students' processing of novel STEM visualizations. *Learning and instruction*, *26*, 12-21.

Hoffmann, R., & Laszlo, P. (1991). Representation in chemistry. *Angewandte Chemie International Edition in English*, *30*(1), 1-16.

Holzinger, A. (2005). Usability engineering methods for software developers. *Communications of the ACM*, *48*(1), 71-74.

Johnstone, A. H. (1982). Macro-and micro-chemistry. *School Science Review*, *64*(227), 377-379.

Johnstone, A. H. (1993). The development of chemistry teaching: A changing response to changing demand. *Journal of Chemical Education*, *70*(9), 701.

Johnstone, A.H. (2000). Chemical education research: where from here? *University Chemistry Educatio*n, 4 (1), pp.32-36.

Johnstone, A. H. (2010). You can't get there from here. *Journal of Chemical Education*, *87*(1), 22-29.

Jong, T. D., Ainsworth, S., Dobson, M., Hulst, V. D. A., Levonen, J., Reimann, P, Sime, J. A., Someren, M. van, Spada, H., & Swakk J. (1998). Acquiring knowledge in science and mathematics: The use of multiple representations in technology based learning environments.

Jordan, S., & Butcher, P. (2010). Using e-assessment to support distance learners of science. In *Physics Community and Cooperation: Selected Contributions from the GIREP-EPEC and PHEC 2009 International Conference* (pp. 202-216).

Kozma, R. B., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of research in science teaching*, *34*(9), 949-968.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, *47*(260), 583-621.

Kubiszyn, T., & Borich, G. (1990). Educational testing and measurement.

Linenberger, K. J., & Holme, T. A. (2014). Biochemistry Instructors' Views toward Developing and Assessing Visual Literacy in Their Courses. *Journal of Chemical Education*.

Luxford, C. J., & Bretz, S. L. (2013). Moving beyond definitions: what student-generated models reveal about their understanding of covalent bonding and ionic bonding. *Chemistry Education Research and Practice*, *14*(2), 214-222.

Luxford, C. J., & Bretz, S. L. (2014). Development of the Bonding Representations Inventory to identify student misconceptions about covalent and ionic bonding representations. *Journal of Chemical Education*, *91*(3), 312-320.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, *38*(1), 43-52.

Menezes, C., & Nonnecke, B. (2014). UX-Log: Understanding Website Usability through Recreating Users' Experiences in Logfiles. *Journal ISSN*, *2368*, 6103.

Murphy, K., Holme, T., Zenisky, A., Caruthers, H., & Knaus, K. (2012). Building the ACS Exams anchoring concept content map for undergraduate chemistry. *Journal of Chemical Education*, *89*(6), 715-720.

Nielsen, J. (1989, November). The matters that really matter for hypertext usability. In *Proceedings of the second annual ACM conference on Hypertext* (pp. 239-248). ACM.

Nielsen, J. (1994). *Usability engineering*. Elsevier.

Nielsen, J. (March 19, 2000) Why you only need to test with 5 users. Retrieved from http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

Norman, D. A. (2002). *The design of everyday things*. Basic books.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*(4), 443-460.

Pearson, K. (1895). Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 240-242.

Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *50*(302), 157-175.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). Knowing what students know. *The Science and Design of Educational Assessment. Washington*.

Reviere, R., Berkowit, S., Carter, C. C., & Graves Ferguson, C. Needs Assessment: A Creative and Practical Guide for Social Scientists; Taylor Francis: Washington, DC, 1996. *There is no corresponding record for this reference*.

Rohrer, C. (October 12, 2014). When to use which user experience research methods. Retrieved from http://www.nngroup.com/articles/which-ux-research-methods/

Russo, S., & Silver, M. (2001). *Essentials of Introductory Chemistry*. Prentice Hall.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, *18*(2), 119-144.

Schroeder, J., Murphy, K. L., & Holme, T. A. (2012). Investigating factors that influence item performance on ACS exams. *Journal of Chemical Education*, *89*(3), 346-350.

Silberberg, M., & Duran, R. (2002). *The molecular nature of matter and change*. McGraw-Hill.

Spartan, V. (1995). 4.1, Wavefunction. *Inc., Irvine, CA.*

Stieff, M., Hegarty, M., & Deslongchamps, G. (2011). Identifying representational competence with multi-representational displays. *Cognition and Instruction*, *29*(1), 123-145.

Tang, K. S., Tan, S. C., & Yeo, J. (2011). Students' multimodal construction of the work–energy concept. *International Journal of Science Education*, *33*(13), 1775-1804.

Treagust, D. F., & Chittleborough, G. (2001). Chemistry: A matter of understanding representations. *Advances in research on teaching*, *8*, 239-268.

Tsui, C. Y., & Treagust, D. F. (2003). Genetics reasoning with multiple external representations. *Research in Science Education*, *33*(1), 111-135.

Ultra, C. (2001). *Cambridge Soft Corporation, Cambridge, USA.*

Vendlinksi, T., & Stevens, R. (2002). Assessing student problem-solving skills with complex computer-based tasks. *The Journal of Technology, Learning and Assessment*, *1*(3).

Virzi, R. A., Sorce, J. F., & Herbert, L. B. (1993, October). A comparison of three usability evaluation methods: Heuristic, think-aloud, and performance testing. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 37, No. 4, pp. 309-313). SAGE Publications.

Welch, B. L. (1947). The generalization of student's' problem when several different population variances are involved. Biometrika, 28-35.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, 80-83.

Williamson, V. M., Hegarty, M., Deslongchamps, G., Williamson III, K. C., & Shultz, M. J. (2013). Identifying student use of ball-and-stick images versus electrostatic potential map images via eye tracking. *Journal of Chemical Education*, *90*(2), 159-164.

Zenisky, A. L., & Sireci, S. G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education*, *15*(4), 337-362.

APPENDIX A

INSTITUTIONAL REVIEW BOARD APPROVAL

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
1138 Pearson Hall
Ames, Iowa 50011-2207
515 294-4566
FAX 515 294-4267

Date: 8/11/2014

To: Dr. Thomas Holme
0213 Gilman Hall

From: Office for Responsible Research

Title: Electronic Delivery of Scaffolded Visualization Tutorials and Assessments in Chemistry

IRB ID: 13-342

Study Review Date: 8/8/2014

The project referenced above has been declared exempt from the requirements of the human subject protections regulations as described in 45 CFR 46.101(b) because it meets the following federal requirements for exemption:

- (1) Research conducted in established or commonly accepted education settings involving normal education practices, such as:
    - Research on regular and special education instructional strategies; or
    - Research on the effectiveness of, or the comparison among, instructional techniques, curricula, or classroom management methods.

- (2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey or interview procedures with adults or observation of public behavior where
    - Information obtained is recorded in such a manner that human subjects cannot be identified directly or through identifiers linked to the subjects; or
    - Any disclosure of the human subjects' responses outside the research could not reasonably place the subject at risk of criminal or civil liability or be damaging to their financial standing, employability, or reputation.

The determination of exemption means that:

- **You do not need to submit an application for annual continuing review.**

- **You must carry out the research as described in the IRB application.** Review by IRB staff is required prior to implementing modifications that may change the exempt status of the research. In general, review is required for any modifications to the research procedures (e.g., method of data collection, nature or scope of information to be collected, changes in confidentiality measures, etc.), modifications that result in the inclusion of participants from vulnerable populations, and/or any change that may increase the risk or discomfort to participants. Changes to key personnel must also be approved. The purpose of review is to determine if the project still meets the federal criteria for exemption.

  Non-exempt research is subject to many regulatory requirements that must be addressed prior to implementation of the study. Conducting non-exempt research without IRB review and approval may constitute non-compliance with federal regulations and/or academic misconduct according to ISU policy.

  Detailed information about requirements for submission of modifications can be found on the Exempt Study Modification Form. A Personnel Change Form may be submitted when the only modification involves changes in study staff. If it is determined that exemption is no longer warranted, then an Application for Approval of Research Involving Humans Form will need to be submitted and approved before proceeding with data collection.

Please note that you must submit all research involving human participants for review. **Only the IRB or designees may make the determination of exemption**, even if you conduct a study in the future that is exactly like this study.

Please be aware that **approval from other entities may also be needed.** For example, access to data from private records (e.g. student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **An IRB determination of exemption in no way implies or guarantees that permission from these other entities will be granted.**

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.

Assigned IRB ID: **13-342**

RECEIVED JUL 29 2014 By IRB

INSTITUTIONAL REVIEW BOARD (IRB)
Modification Form for Non-Exempt Research

Title of Project: Electronic Delivery of Scaffolded Visualization Tutorials and Assessments in Chemistry

Principal Investigator (PI): **Thomas Holme**			Degrees: PhD
University ID: **544911377**	Phone: 294-9025	Email Address: taholme@iastate.edu	
Department: Chemistry			

FOR STUDENT PROJECTS *(Required when the principal investigator is a student)*		
Name of Major Professor/Supervising Faculty:		
University ID:	Phone:	Email Address: @iastate.edu

Alternate Contact Person:	Email Address:
Correspondence Address:	Phone:

Please notify the IRB Office if your contact information has changed since the last review.

ASSURANCE

- I certify that the information provided in this application is complete and accurate and consistent with any proposal(s) submitted to external funding agencies. Misrepresentation of the research described in this or any other IRB application may constitute non-compliance with federal regulations and/or academic misconduct.
- I agree to provide proper surveillance of this project to ensure that the rights and welfare of the human subjects are protected. I will report any problems to the IRB. See Reporting Adverse Events and Unanticipated Problems for details.
- I agree that modifications to the approved project will not take place without prior review and approval by the IRB.
- I agree that the research will not take place without the receipt of permission from any cooperating institutions when applicable.
- I agree to obtain approval from other appropriate committees as needed for this project, such as the IACUC (if the research includes animals), the IBC (if the research involves biohazards), the Radiation Safety Committee (if the research involves x-rays or other radiation producing devices or procedures), etc., and to obtain background checks for staff when necessary.
- I understand that IRB approval of this project does not grant access to any facilities, materials, or data on which this research may depend. Such access must be granted by the unit with the relevant custodial authority.
- I agree that all activities will be performed in accordance with all applicable federal, state, local, and Iowa State University policies.

[redacted] 28Jul14

Signature of Principal Investigator — Date

Signature of Major Professor/Supervising Faculty — Date
(Required when the principal investigator is a student)

For IRB Use Only	Full Committee Review: ☐	Review Date: August 8, 2014
	Approval Not Required: ☐	Approval/Determination Date: August 8, 2014
EXEMPT per 45 CFR 46.101(b): 1, 2	*Not Research:* ☐	Approval Expiration Date: NA - exempt
EXPEDITED per 45 CFR 46.110(b): Category Letter	*No Human Subjects:* ☐	
	Not Approved: ☐	Risk: Minimal ☑ More than Minimal ☐
IRB Reviewer's Signature	[redacted]	August 8, 2014

APPENDIX B

IN-CLASS RECRUITMENT SIGNUP

Hi, my name is Jack Polifka, and I am a second semester Masters student in Human Computer Interaction, and I work in the Department of Chemistry at Iowa State University. My advisor is Dr. Thomas Holme. My research study involves creating an online assessment and identifying how students interact with the representations on the assessment. I am interested in having people participate in a **short 30 minute interview** where you will: 1) take a short survey about your chemistry experiences, 2) use the online tool, and 3) take a final survey intended to provide feedback about the application. Your participation or non-participation has no influence on your grade in this course. Willing participants will be offered a pizza party after the study, if they wish. By volunteering, you will be helping us improve education in chemistry. Once I have completed the interviews, I will destroy the name and email information below. For those students who are interviewed, your identity will be known only to the interviewer. By signing below, you indicate that you are at least 18 years old.

If you do not wish to participate in this research project, leave this form blank.

If you are willing to participate in this research project, please complete the information below.

Name: ________________________________

E-mail address: _________________________

PLEASE PLACE THIS FORM IN THE DESIGNATED ENVELOPE

APPENDIX C

FIRST VERSION INTERVIEW QUESTIONS

1. What did you like about the application and what would you keep?
2. What did you dislike about the application and what would you remove?
3. Is there anything you would improve about the application?
4. Were there any representations you felt you used more than others?
5. Were there any representations you felt you used the least?
6. Was it hard to change images?
7. Was the size of the image big enough?
8. If you were given a chemistry exam in this format would you like it? Would you think it would be fair?

APPENDIX D

SECOND VERSION INTERVIEW QUESTIONS

1. What did you like about the application and what would you keep?
2. What did you dislike about the application and what would you remove?
3. Is there anything you would improve about the application?
4. Does the layout of the application help you?
5. Did you know you could select answers two different ways?
6. Was the tutorial page of any help to you?
7. If you were given a chemistry exam in this format would you like it? Would you think it would be fair?

APPENDIX E

THIRD VERSION INTERVIEW QUESTIONS

1. What did you like about the application and what would you keep?

2. What did you dislike about the application and what would you remove?

3. Is there anything you would improve about the application?

4. Was selecting your answers easy or hard? What made it easy or hard?

5. On the right hand side, there are 5 green boxes with labels, what were those things?

6. Was the tutorial page of any help to you?

7. If you were given a chemistry exam in this format would you like it? Would you think it would be fair?

APPENDIX F

EXAMPLE OF RECORDED INTERVIEW FROM FIRST VERSION

Jack: "Ok. So my first question is what did you like about the application and what would you keep?"

Participant: "I just like being to switch between the different ways to look at different molecules just because depending the problem you want to see different parts of the atom. So I liked being to switch between them all."

Jack: "Could you give me an example?"

Participant: "Umm there was one where it wanted to look at the shape of the molecule, so looking at the chemical formula was going to be helpful but the ball-and-stick method was extremely helpful."

Jack: "Was there anything you liked and you would keep?"

Participant: "Well that was... there were times where different representations of the molecule weren't necessary, so maybe if there was a way to get rid of some of them so people are confusing themselves with each ones they need. I think that's probably complex for a program but having all five options every time might too much for some people, but part of it is knowing which representation to use."

Jack: "Did you think ever having, do you think having all of the representations would confuse someone in how it is setup? Cause it's set up so you only have three molecules at a time and three pictures."

Participant: "It shouldn't be too much of an issue as long as the questions are multiple choice like that. If they weren't, I see how it would really apply and as long as people know what representations are. If someone goes in not knowing what ball and stick represents, it will be confusing to them. They'll be like what am I looking? Especially that electrostatic one, if people have never seen those before then that could be confusing. Otherwise it was fairly simple to use as long as you know what the representations were."

Jack: "Ok. Umm was there anything you didn't like about the application and that you would remove?"

Participant: "Umm just the layout of the different atoms. They were in like a triangle. A, B, C. But then the options were linear. A, B, C. So there were a couple times we was started with clicking B, but I meant C and so on. So if there's a way to stack them linearly or move the.... options you click to represent that better that would help."

Jack: "You think having Umm a consistent between the options and the images would help?"

Participant: "That would be more helpful then it was. Obviously you just need to read it to click the right button you want. But if you're not thinking about you might click the wrong one if you're just to click through buttons. So that would be the only thing I would change about it that I saw while using it."

Jack: "Umm that answers my third question I was going to ask. What would you improve about the application? You would improve the layout and how the answer line up with the representations or the pictures. Would there by anything you would improve?"

Participant: "I thought it was really easy to use. It was nice to just click on whatever representation you wanted. I thought it was a really good program."

Jack: "Cool. Umm what representations did you think you used the most?"

Participant: "I definitely used the ball and stick when it was talking about shapes. I used the electrostatic potential one when it was talking about electronegativity or when has the negative potential."

Jack: "Why?"

Participant: "Because... well going into it, I assumed it was going to show me where the negative charges were going to be and positive charges being blue. Umm never having used those before, I probably could have gotten those wrong but it looked like... I kind of used context clues to figure that one out."

Jack: "Wait, wait. Can you explain the colors? You said blue was positive charge?"

Participant: "Yes."

Jack: "Ok. And what were the other colors?"

Participant: "Red would be the more negative charge. Umm because of the way I looked at it the first time."

Jack: "So blue is more positive and the red areas are more negative?"

Participant: "Yes."

Jack: "Ok. Ok. Cool. Umm were there any, what representations did you use the least?"

Participant: "Umm probably just the chemical formulas themselves. They're nice to know what atom was talking, using however it necessarily show me too much about the atom. So I spend a lot of time on the ball-and-sticks and then occasionally the wedges... Umm then I liked the way the electrostatic looked even I didn't necessarily know how I was using them so."

Jack: "Ok. Umm ever the images big enough in the application?"

Participant: "Yea."

Jack: "Did you know there was a zoom option?"

Participant: "I saw the zoom option. I didn't use it. I didn't think it, I was able to see them just fine."

Jack: "Can you think of any instance when you would use or when you would need it?"

Participant: "The only time I think I would want to use a zoom option is when I could change the orientation of the picture to look at it from the back or something."

Jack: "Ok."

Participant: "If you're just zooming in, I think that just draws away from entire molecule."

Jack: "Ok. Umm was changing representations easy?"

Participant: "Changing was easy. You just had to click different buttons."

Jack: "So you always change them all at once?"

Participant: "As in ... what are you trying?"

Jack: "So, so there's an option to change them all at once and there's an option to change one molecule at a time."

Participant: "O. I didn't know that. I just changed them all at once."

Jack: "Ok. Would you... why is that? Why did you change them all at once? Because you couldn't find, you didn't know about the single?"

Participant: "I didn't know you change them single. But I would probably still change them all at the same time so I see what they all look like. For example if I hit the electrostatic ones for two of them and ball and stick for the 3rd, I wouldn't have known about the shape for the electrostatic ones."

Jack: "So would it be correct to say you did it for comparison reasons?"

Participant: "Yea."

Jack: "Ok. So there would there were be an instance you would want to change just one molecule at a time."

Participant: "Umm if I had narrowed down between two of the options and I just wanted to see differences between the two going from like a ball and stick to the chemical formula maybe. Seeing which atoms specially hover, it's still going to be a comparison between two of them, so I would probably not use that option as much as I would change all three of them."

Jack: "Ok. Cool. Umm then my last question is if you were given a test or exam in chemistry in this format, would you like it? Would you think it's fair?"

Participant: "Umm yea. I thought... the questions were easier to answer being able to look at the all of the different representations for the molecule. If he would ask questions like that without showing the molecules on test. They would only give us one and except us to know how to do that rest. Taking a test with the program could only be helpful cause, it would show you different representations of each molecule. I don't see how that would hurt your chances on being well on it."

Jack: "Ok. Cool. What other questions become harder and they required you to look at multiple representations? Like you couldn't just get the answer from one, you had to like get it from two."

Participant: "Well then, definitely I think using the program would be much better because then you can switch between two and look at both and if it was just a paper test and they just printed off the one picture. Even if they print two, it's the same kind of idea. That thing that is hard is that you have to know which to use for which question."

Jack: "Ok. Cool. That's all I have. Do you have any addition comments or feedback?"

Participant: "Not that I can think of."

Jack: "Awesome. Thank you."

APPENDIX G

EXAMPLE OF RECORDED INTERVIEW FROM SECOND VERSION

Jack: “Cool, so my first question is… what did you like about the application and what would you keep?”

Participant: “Umm… it’s very easy to use and you could switch between each of the different categories from like Lewis structure to the electrical field. That really helped out with solving the other problems.”

Jack: “Was there anything you would keep or that you liked and would keep?”

Participant: “Umm honestly I think the whole system. It ran smoothly. I didn’t have any problems with how, how the system was really easy to understand. I picked it up right away, so I thought it would be very simple for anyone to use.”

Jack: “Was there anything you dislike about the application and you would remove?”

Participant: “Umm… not necessarily. No.”

Jack: “Is there anything you would improve upon in the application?”

Participant: “I don’t know if it was just the, how the test was setup or if it was randomly generated but a lot of my answers right at the beginning were probably, I see first ones were all C but otherwise.”

Jack: “Ok. Ok. Umm how was the layout of the application? How was the layout of the molecules in the application, side by side in that straight line?”

Participant: “Side by side, it was Umm it wasn’t that bad to pick them a part. Umm… with the answers though I guess like the A, B, and C up top, so times I had to go back and check but to figure out, check which offset.”

Jack: “They were offset, so like… I have an idea what you mean. They weren’t exactly lined up so like the top answers where the random buttons were A, B, and C and these were kind of offset by the answer. Umm would you line those up if you could?”

Participant: “Yea. I mean I picked it up and it wasn’t the hard for me but it wasn’t that hard to distinguish but I don’t know if in the future a different student if you know what they might get confused.”

Jack: “Ok. Do you think any other shape for the layout would umm help?

Participant: “Umm if you have more… let’s say between just A, B, and C and let’s say you have A through E maybe if you did horizontal that might save. I don’t know but necessarily save space but make it easier for them to keep track.”

Jack: “Ok. Umm did you know you could select your answers two different ways?”

Participant: "No, I do not."

Jack: "So you can click the radio buttons which I'm assuming which you did."

Participant: "Yea."

Jack: "And then you can click on the image and that will select the radio button for you."

Participant: "I didn't know that. That would work it…"

Jack: "Would that have made it easy for you if you would have known that?"

Participant: "I don't know if it would have made it easier but it would have made it… it be giving us less of a chance of making a mistake on choosing the hard answer I guess. If that makes sense."

Jack: "Can you explain? What would make you make you less likely to choose a hard with that way?"

Participant: "Less likely by clicking the picture."

Jack: "Yea."

Participant: "Clicking on the picture would make it less likely because like, if you on a hand written test and you write down B and you meant A. It's kind of like that, so instead of hitting, clicking in the buttons up top instead of hitting the B, you want A. You could click on the picture which you know is what you want and that make sure what you want."

Jack: "O. So… there's a stronger distinction …"

Participant: "Stronger distinction. Yup."

Jack: "Ok. Ok. That's my word choice. Umm… how would you put it?"

Participant: "Umm… it make it less, less of a change of making a mistake by giving the wrong answer but knowing… You can have the confidence that you're choosing the once you want to put in and you want to submit your own answer."

Jack: "Umm was the tutorial page helpful to you?"

Participant: "Yea. Yea."

Jack: "Was it?"

Participant: "What do you mean the tutorial page?"

Jack: "For the tutorial page, it's just a static image that says that this layout displays the functionality or something to those words of the test. If you want to continue, click this hyperlink."

Participant: "Yea. That's how I figured out how to switch thorough all of the different layouts."

Jack: "But didn't, so then the ways for selecting answers was on there too."

Participant: “I was of just skipped over that then.”

Jack: “Ok. Ok. Umm is there anything that would help that tutorial page to get that idea of being able to select answers another way a crossed?”

Participant: “If you say that, if explains how you can choose two different ways and that’s on there, then no. I would say it’s good. I would say that I should probably read through the instructions before.”

Jack: “No, don’t assume that it was your fault, because like tutorial pages are meant to be read but people still don’t read them most of the time. At least I think. Umm... ok. If you were given a test or an exam in this format for chemistry, would you think it’s fair to you umm and would you like it?”

Participant: “I love this. If this is how we took our last couple of tests, it would, well I hope, I know it would improve our grade because I would be able to switch through with, especially with the Lewis structure. I haven’t really gotten to the circle…. I don’t how to put descriptions for each of the molecules. It would help me answer the questions more knowing more of the background of the molecule, so I think it would definitely make testing a lot easier.”

Jack: “Ok. Cool. Do you have any more comments? Ok.”

Participant: “I just wish I could may use this again in the future.”

Jack: “Alright. Cool.”

APPENDIX H

EXAMPLE OF RECORDED INTERVIEW FROM THIRD VERSION

Jack: "Cool. So my first question is what did you like about the application and what would you keep?"

Participant: "I definitely liked the different options of viewing the different problems for sure. To look at each problem in a different way is helpful. Specially looking here at all five different representations of the problem was, it allowed you to answer the problem differently if you saw it in a different way, so I like that."

Jack: "Do you think people see problems in different ways?"

Participant: "Absolutely, I think people do. I think some people would be more like visual with Lewis-dot diagrams or just the chemical formula will allow people to know the answer right away or looking at the electro...."

Jack: "Static."

Participant: "Static Potential Map will definitely get a way a lot of answers, so definitely different people see different ways."

Jack: "Ok. Umm was there anything else you liked about the application and that you would you keep?"

Participant: "It was easy. It was very easy. You don't need any help figuring it out. Umm I wish myself I would have spent a little bit more time on a question or two and I would of gotten more right but Umm yea definitely very easy. I like it."

Jack: "Ok. What made it easy?"

Participant: "Umm the flow. It was just easy to click all five Umm options were on the side, left-hand side. You just click and maneuver each through and different pictures would pop up and show you the different ways the problem can be solved by looking at it. Umm it was very simplistic."

Jack: "Was there anything else that made it easy to use besides pressing the buttons?"

Participant: "Umm the visuals were great. Umm they were clear, clear cut. Umm besides that no. It was easy flow and the visuals were very clear."

Jack: "Umm was there anything about the application you didn't like and that you would remove?"

Participant: "Umm looking back at it. I don't think of anything I could add or take away. No, I think it was, like I said, very well thought out."

Jack: "You just said you wouldn't add or remove anything, Umm but would there anything you would improve on or change in the application?"

Participant: "Umm let me think. (Pause) I don't think I... I wouldn't add anything or change anything. I think that it was, I can't change it myself at least. No, I didn't see anything I could change."

Jack: "Well, pretend you can do anything. Because we aren't talking within your abilities."

Participant: "Ok. Umm. I don't know. I mean color schemes maybe to give more people. I wouldn't... looking back at it, the diagrams were perfect. I wouldn't change the diagrams at all. They represented that they needed to. The questions were just basic chemistry questions that people should know. Umm I don't know. I don't think I could change anything honestly."

Jack: "Ok."

Participant: "Well having the.... five different options on the side to click was in the perfect position and very close to the problem. You didn't to go looking for it. Umm maybe the one thing I didn't have a problem with. At the beginning was realizing that I could change to different views."

Jack: "Ok."

Participant: "Maybe having a statement saying, you know, there are options on the side to change. At first, I read the question and I answered it based on just using that very first picture it give me. I didn't initially realize that there were other options. I know that's what we would be doing but it didn't click right away. I had to go through the first problem and then "O" realize here are the options. I guess I would say let the user know by a statement, a small statement, on the top that those are the different the different options he or she click through. That is one thing I would change."

Jack: "Would you do that? Would you just put something plainly that says that all the time or would you just give a quick remainder?"

Participant: "Maybe something as simple as a quick remainder on each slide. Or maybe on the initial quiz, or after the quiz. Have a slide between the quiz and the post quiz and the test. Something between saying this is what you'll be doing. There will be options on the side able to change the different views of the problems or something like that. So kind of initial statement or be something simple on each slide to remind them."

Jack: "What if I told you that was already there?"

Participant: "Umm I mean. Myself. I forgot initial that would be there on the side. I mean um, the user would probably forgot, or you can't assume every user would know or would remember. You know some people are definitely use visuals vs hearing it but they need to see each to be reminded. Umm for me, I'm definitely a visual. Someone can tell me something. A lots going on. Things go outside my head. I should just completely forgot something that someone told me and I would need to be reminded by some visual."

Jack: "Ok. Between thee, the pre-survey and the quiz itself, there's a tutorial page where thee..."

Participant: "It shows you what."

Jack: "Yea, it... there's like a little blurb on the top and fades in the question for the thing and then asks you to change the representation and then select an answer."

Participant: "I guess I jumped and I didn't see it guess. Yea."

Jack: "Ok. Umm would you do...."

Participant: "Then I guess that initial... I didn't actually, when I got to the test, I realized I could change, but when I first saw the quiz I didn't read that up on the top. I just answered the question based on that first diagram. But as soon, that was, like an example the test would look like. I didn't realize that there. But once, I got to the test, I did. So I guess I didn't realize initially which is you're supposed to figure out then."

Jack: "How would you help Umm how could I have helped you known that was a test question tutorial thing?"

Participant: "It said it. So I did read it and it said it was a preview of what the test would look like. Maybe... make that page Umm look a little bit different from the test so it's more distinguishable. Because I thought we jumped straight into the test. But then I did read that was preview of what the test question would look like and the test question looked exactly like the same format. So maybe some sort of different format for the pretest would allow someone to that's completely different, even though it's a same type of question. They would be asked the same. Just a little bit different. Maybe that would have helped me. That's all I got"

Jack: "Ok. Umm was selecting your answers easy or hard?"

Participant: "Absolutely easy."

Jack: "Why?"

Participant: "It was very clear cut. I mean I had three different options for each problem. All I had to do was click the little dot for multiple choice. It was very easy. Then I all I had to do was submit."

Jack: "Ok. Umm you already said this bunch, but I just want to ask for more information. Umm there are five boxes on the right hand side. They are green. They have labels. Don't tell me what the labels are. What are those boxes?"

Participant: "What were those boxes?"

Jack: "Yea, those five things on the right, left-hand side. Excuse me that were green."

Participant: "The different options for representing the problem."

Jack: "And what were those used for?"

Participant: "They were used for giving you different visual look at what the problem was asking for. Or the problem would ask you a question and umm typically it would present a formula or different representation of what you need to know to answer that problem. I mean each question included a formula, chemical formula, and a different way of representing that formula. Different visuals... I guess. Does that answer?"

Jack: "Yea. Umm how did you know they would give you a different representation?"

Participant: "By the name. Umm I've seen the names in the past. I've two different chemistry courses here before in high school. Umm so I had worked. I've worked with every map that they give the option with but the commonly didn't work with the... I worked with Lewis-dot structures a lot, just the plain chemical formula, and the dot and ... I didn't actually call it that, but yes those three were the most. I didn't really work with these two. Four and five. Very much. I have seen them definitely a lot but we did more work with one, two, and three."

Jack: "Ok, so because you know of those representations you know you selected those boxes..."

Participant: "I knew what it would look like."

Jack: "You knew what it would like. Did you know that when you hit them or something that they would change those?"

Participant: "Not initially no. I supposed though during the pre-quiz, clicking the button with the name would change the form."

Jack: "What made you suppose that?"

Participant: "Umm when you first bring up the question, you see a representation of Umm the question at hand and I saw the different names and I saw the name that it was representing currently. The one that first popped up and I assumed that I if clicked another button it would change to that form so the current form was of the one of the ones listed in the five. So that's how I knew."

Jack: "Ok. Umm and I had really asked. Was the tutorial page helpful? But you said you would change it to look like something else."

Participant: "Yea, I mean Umm I think would just because it distinguishable the actual pre-quiz from the actual quiz. I mean the overall idea is to make it look similar because that's what it's going, you're going to be representing is the real quiz but umm maybe somehow make it a little bit distinguishable from the actual quiz so people know that the actual tutorial or way to understand what you will giving them."

Jack: "Umm would you, if you were given a test, exam, quiz, whatever in chemistry in this format would you think it would be fair? Would you like it?"

Participant: "Personally, I would like it. Umm it would help me answer the problem easier. Umm"

Jack: "From your point of view, not the teachers. I'm jumping the gun here."

Participant: "For the teachers, Umm."

Jack: "Don't worry about the teachers."

Participant: "Ok. I would. Definitely, from my point of view, definitely would help me. It would, I could. For some of the questions I looked at, if it give me #4 form, I haven't worked with them very much. I don’t know how to correctly answer with that form. I had to switch it to number one form, the chemical formula, to answer it. Or umm looking at number two, the form. So definitely having every option helped me answered the question. That's in my point of view. In teacher's point of view,"

Jack: "Don’t worry about the teachers' point of view. Umm that's all I have. Did you have any Umm other comments that you think that would help me?"

Participant: "Umm besides what I said. No. I think overall was really great. I actually liked looking at it differently like that and it helped answering the chemistry questions."

APPENDIX I

CONSENT FORM GIVEN TO PARTICIPANTS FOR USABILITY TEST

Informed Consent Document

Title of Study**: Electronic Delivery of Scaffolded Visualization Tutorials and Assessments in Chemistry**
Participants: Principle Investigator: Thomas Holme, PhD

INTRODUCTION

The main purpose of this study is to develop tools for on-line learning of molecular level details about chemical systems. The types of chemistry content included in this study usually span multiple course levels in the undergraduate chemistry curriculum. To accomplish this task, the on-line tools, referred to as "digital learning objects" (DLOs) will attempt to scaffold the manner in which you interact with the information present. The current phase of the project is designed to investigate how students interact with the DLO and how the design of the DLO affects those interactions. The findings of this stage of the project will inform the refinement of these learning tools for visualizing chemical representations.

DESCRIPTION OF PROCEDURES

This study will last for approximately 30 minutes**.**

During the study you may expect the following study procedures to be followed.

1) The researchers will contact prospective participants to schedule a study time and will send the Informed Consent Document to allow prior review.

2) On the selected date of the test study, you will be given a copy of the Informed Consent Document for review and to sign prior to the start of the session. If you agree, and sign the Informed Consent Document the session will begin.

3) Information regarding the project will be read and explained to you before session begins.

4) The study will begin with a set of still images from a range of possible chemical and biochemical structure depictions. Accompanying each image there will be a question that asks you to identify a trait or a property associated with the molecule depicted. You will also be video recorded to match your spoken words to the images on the screen.

5) The next step will be to work with the live DLO. You will have an open browser on the screen to use and below it will be an additional window with a question for you to consider. Voice recording will take place during this phase of the study. Your actions to use the DLO, such as the links you click on, will also be recorded.

7) You will fill out a brief survey.

RISKS

There are no foreseeable risks in this study. Participants may leave the study at any time without penalty.

BENEFITS

Participants will receive limited direct benefits from this study. Because the topic of these new DLOs is related to course work they may be currently taking or taking in the future, participation in the study may result in a better understanding of the science related to multiple types of chemical representations. However, the data gained from this usability study can be expected to significantly improve the design of the DLOs and thereby result in benefits for many students who use the DLO later, after it has been revised.

COSTS AND COMPENSATION

You will not have any costs from participating in this study. Participants who wish, may have light food or soft drink refreshments provided by the researchers after the interview.

PARTICIPANT RIGHTS

Your participation in this study is completely voluntary and you may refuse to participate or leave the study at any time. If you decide to not participate in the study or leave the study early, it will not result in any penalty or loss of benefits to which you are otherwise entitled. During the testing, if you feel uncomfortable at anytime you can quit.

CONFIDENTIALITY

Records identifying participants will be kept confidential to the extent permitted by applicable laws and regulations and will not be made publicly available. However, federal government regulatory agencies and the Institutional Review Board (a committee that reviews and approves human subject research studies) and representatives from the National Science Foundation (who funded the study) may inspect and/or copy your records for quality assurance and data analysis. These records may contain private information.

To ensure confidentiality to the extent permitted by law, the following measures will be taken. Only the researchers will have access to the data. The data will be entered and kept in password-protected computer files located on the researchers' computers. Any questionnaires used will be shredded after all the information is entered into the computer. The participants will be identified by a sequential number that will not be traceable to them.

QUESTIONS OR PROBLEMS

You are encouraged to ask questions at any time during this study. For further information about the study contact Jack Polifka (jpolifka@iastate.edu) or Thomas Holme (taholme@iastate.edu).

If you have any questions about the rights of research subjects or research-related injury, please contact IRB Administrator, (515) 294-4566, IRB@iastate.edu, or Director, Office for Responsible Research, (515) 294-3115, 1138 Pearson Hall, Ames, IA 50011.

****

SUBJECT SIGNATURE

Your signature indicates that you voluntarily agree to participate in this study, that the study has been explained to you, that you have been given the time to read the document and that your questions have been satisfactorily answered.

Subject's Name (printed) ____

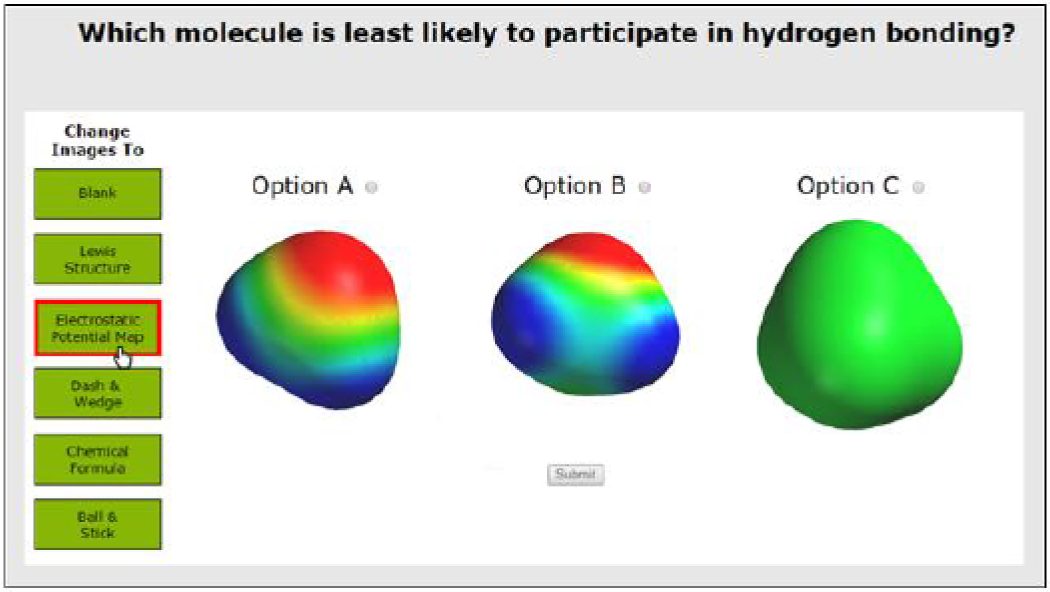
_______________________________________ __________________

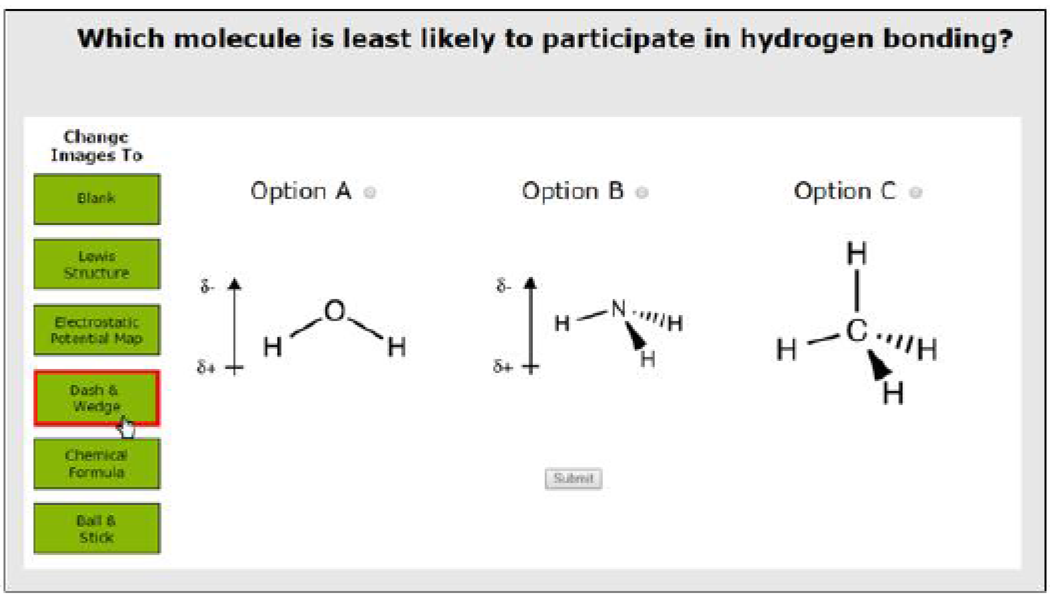
(Subject's Signature) (Date)

APPENDIX J

NEEDS ASSESSMENT SURVEY VRA QUESTIONS

At this time, the ACS Exams Institute is developing a classroom assessment that can assess students' understandings of various representations using an online platform. Students can toggle through different representations in order to answer the question An example in which a student is clicking between two representations is shown belo

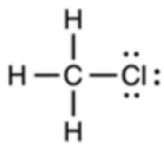




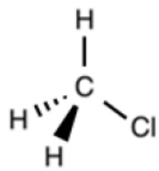
The following include a few different types of representations that are sometimes used in chemistry textbooks to depict compounds and molecules.

CH_3Cl

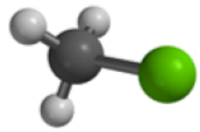
(1)



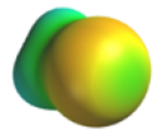
(2)



(3)



(4)



(5)

19. How often do you use these representations on your course exams?

	Never	Rarely	Occasionally	Frequently
(1)	◯	◯	◯	◯
(2)	◯	◯	◯	◯
(3)	◯	◯	◯	◯
(4)	◯	◯	◯	◯
(5)	◯	◯	◯	◯

20. Please indicate your level of interest.

	Not Interested	Somewhat Interested	Interested	Very Interested
To what extent would you be interested in using an assessment like the one shown?	◯	◯	◯	◯

21. Do you believe that allowing students to navigate among different representations of molecules in the manner shown above helps or hurts the assessment of their chemistry knowledge?

◯ Helps

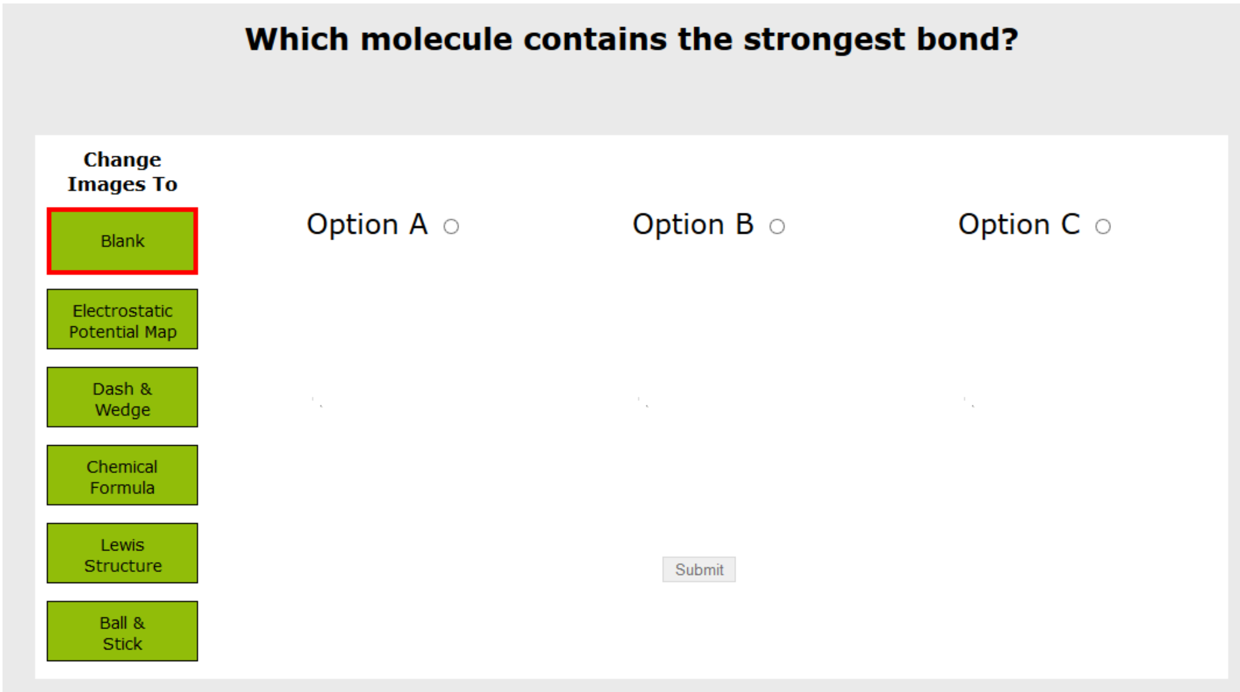
◯ Hurts

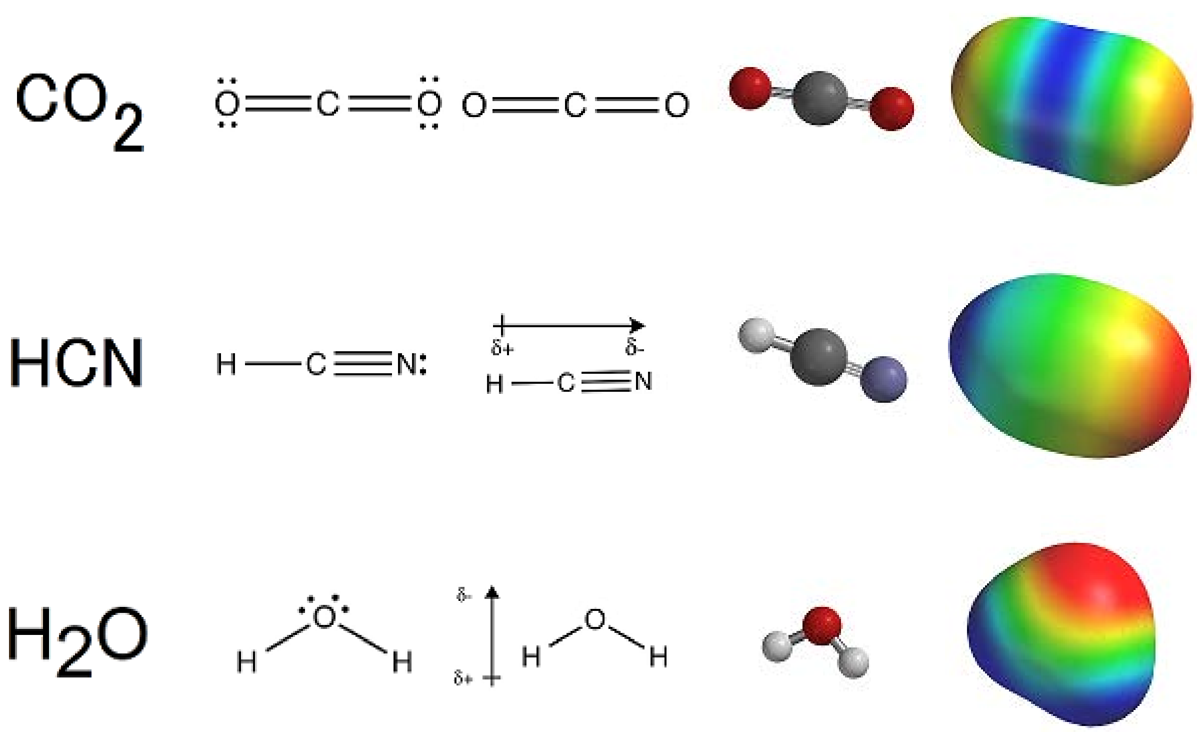
◯ Neither

APPENDIX K

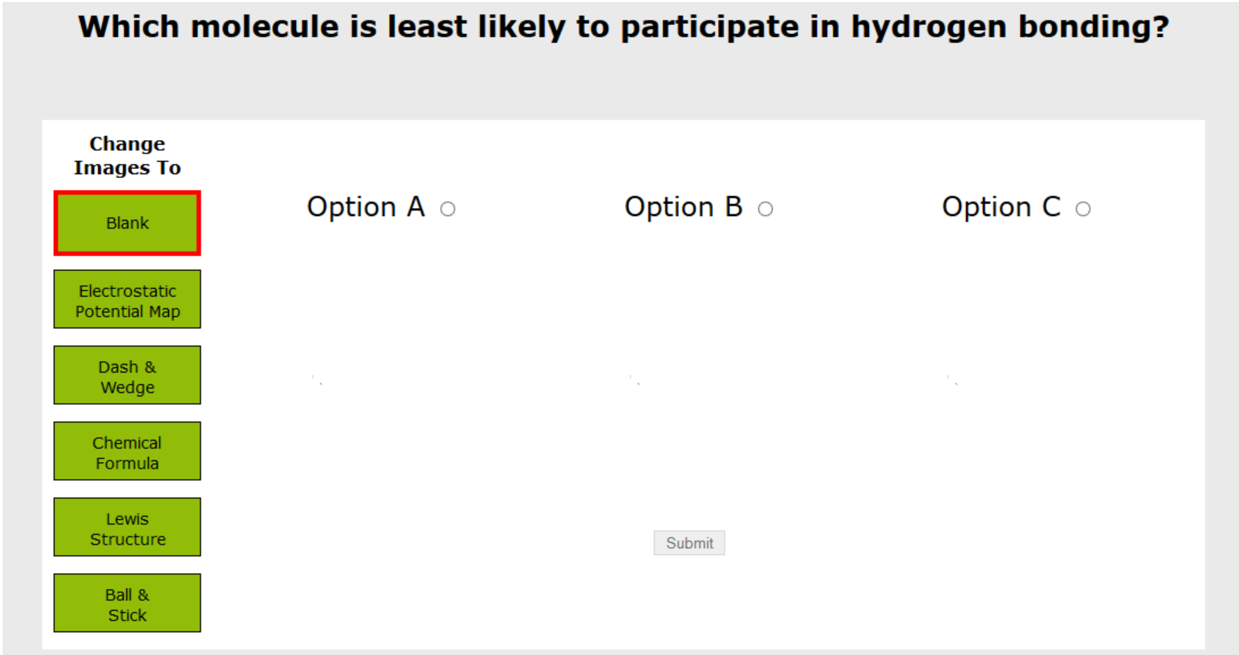
QUESTIONS, ANSWER CHOICES AND REPRESENTATION IMAGES FOR PILOT STUDY

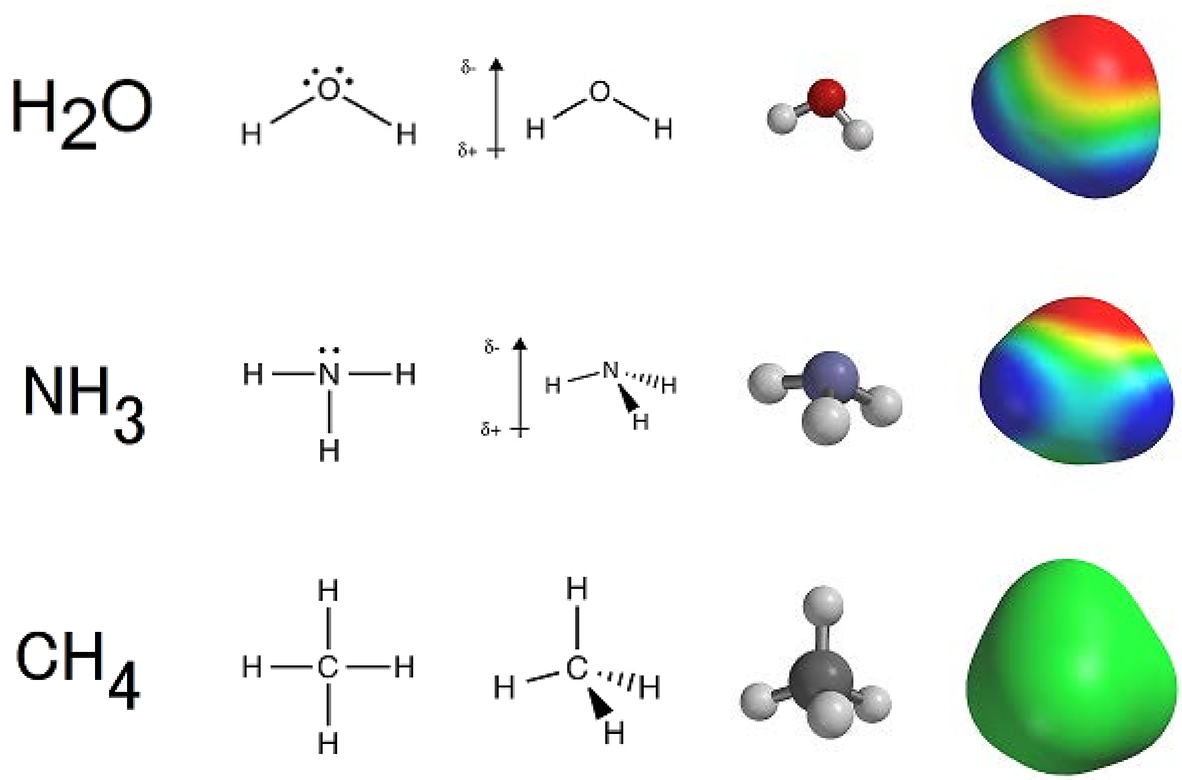
Question 1.)



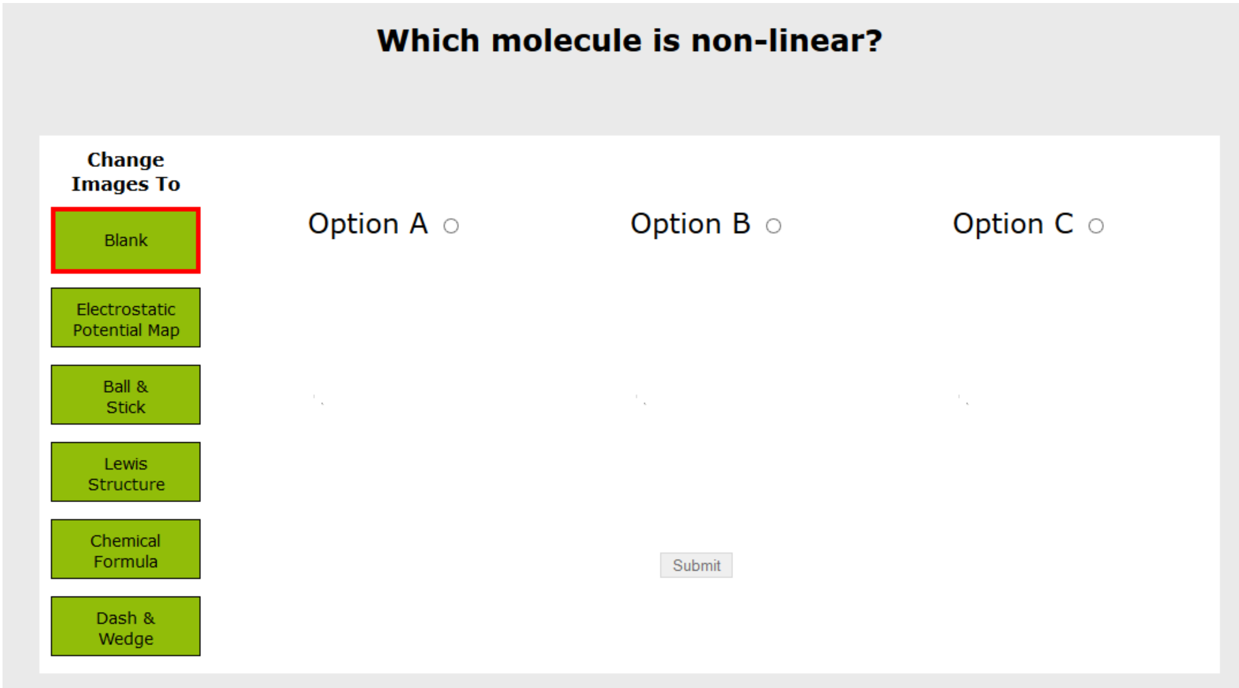


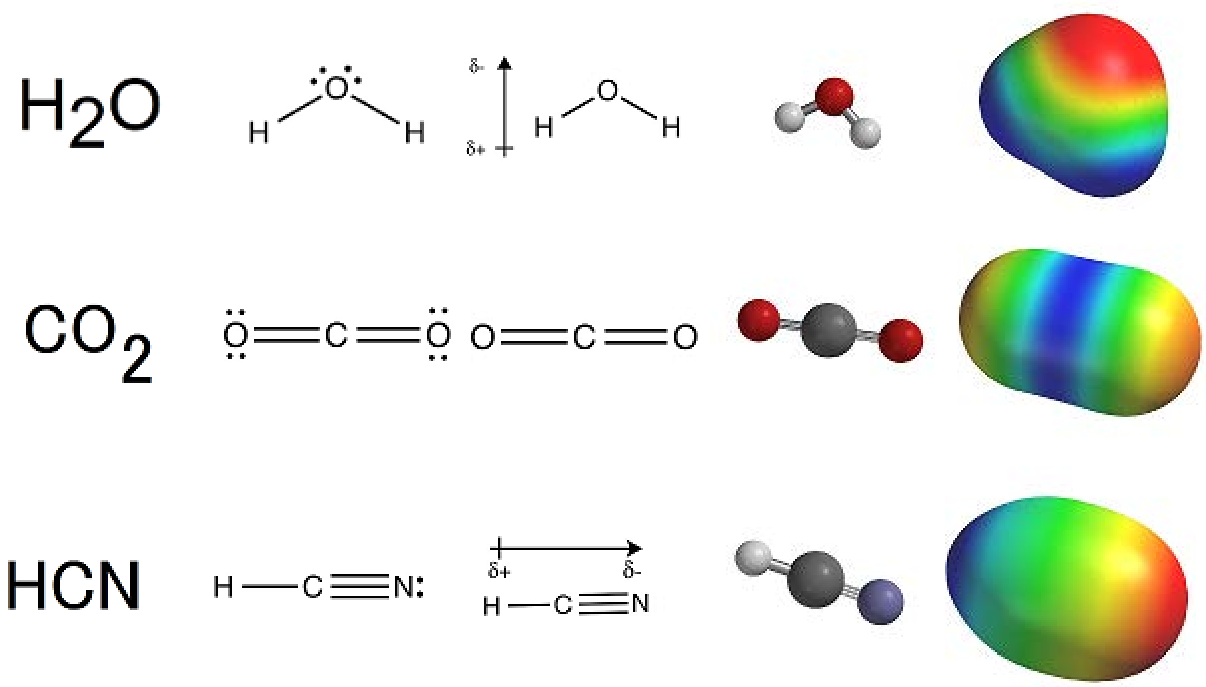
Question 2.)


Which molecule is least likely to participate in hydrogen bonding?
Change Images To
Blank
Electrostatic Potential Map
Dash & Wedge
Chemical Formula
Lewis Structure
Ball & Stick
Option A
Option B
Option C
Submit

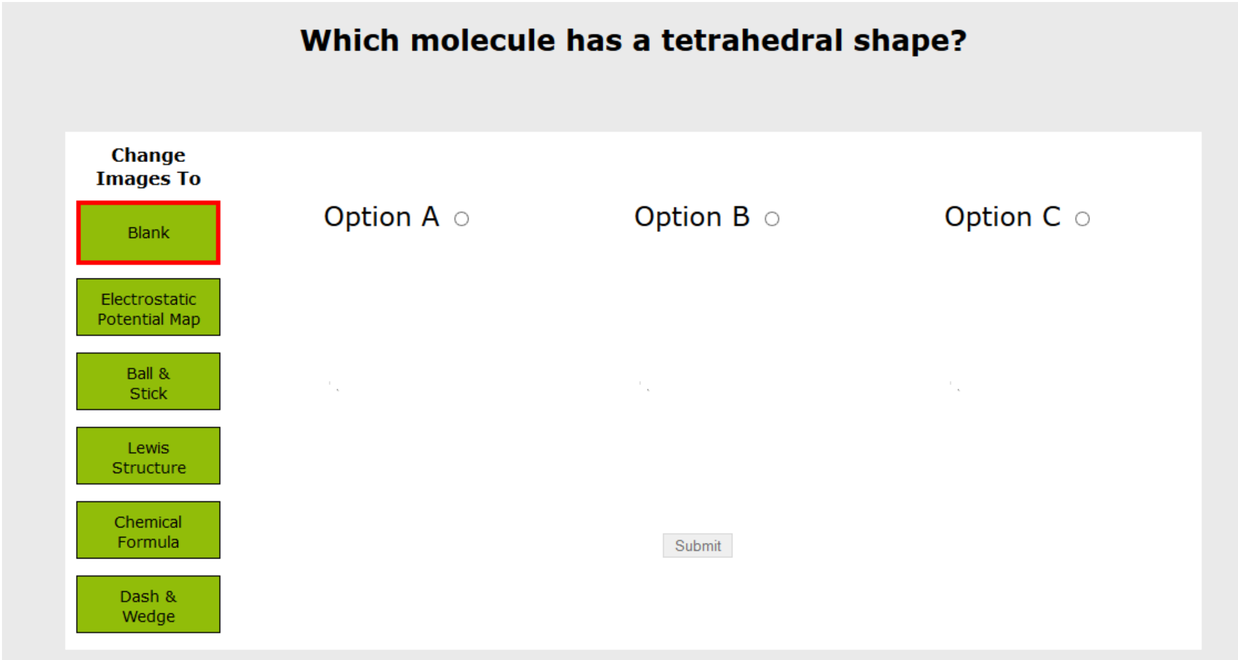

H2O
NH3
CH4
δ-
δ+

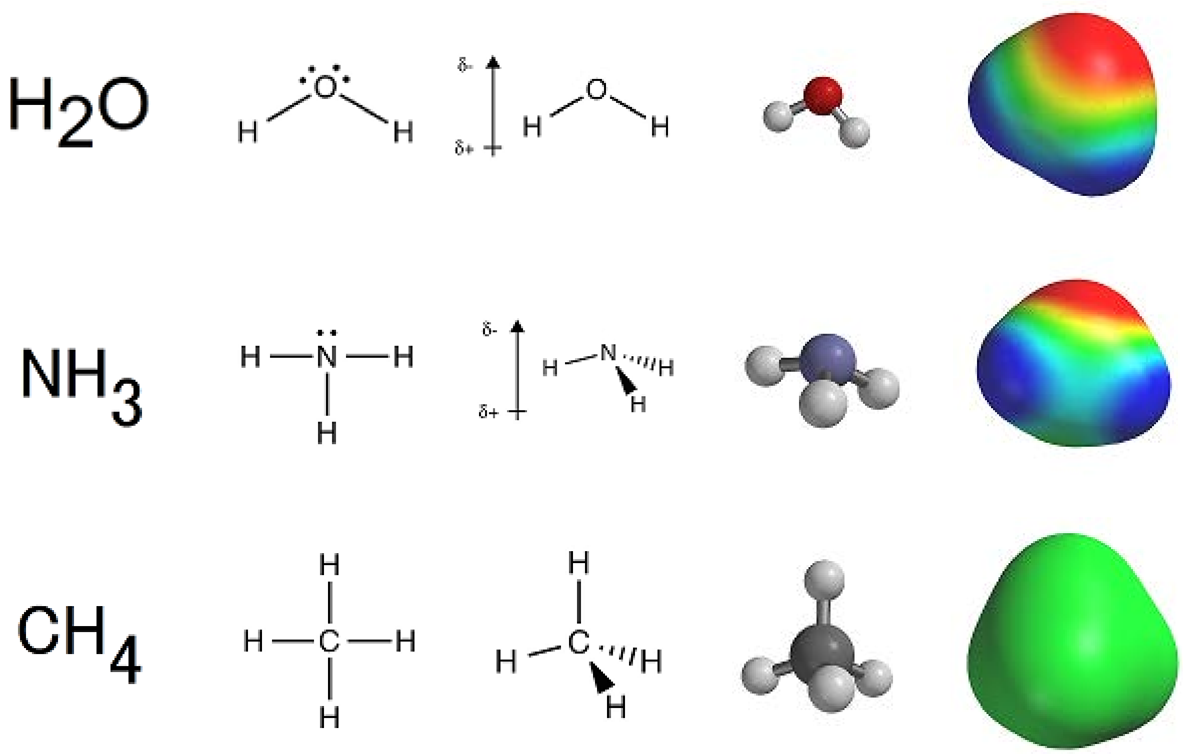
Question 3.)


Which molecule is non-linear?
Change Images To
Blank
Electrostatic Potential Map
Ball & Stick
Lewis Structure
Chemical Formula
Dash & Wedge
Option A
Option B
Option C
Submit

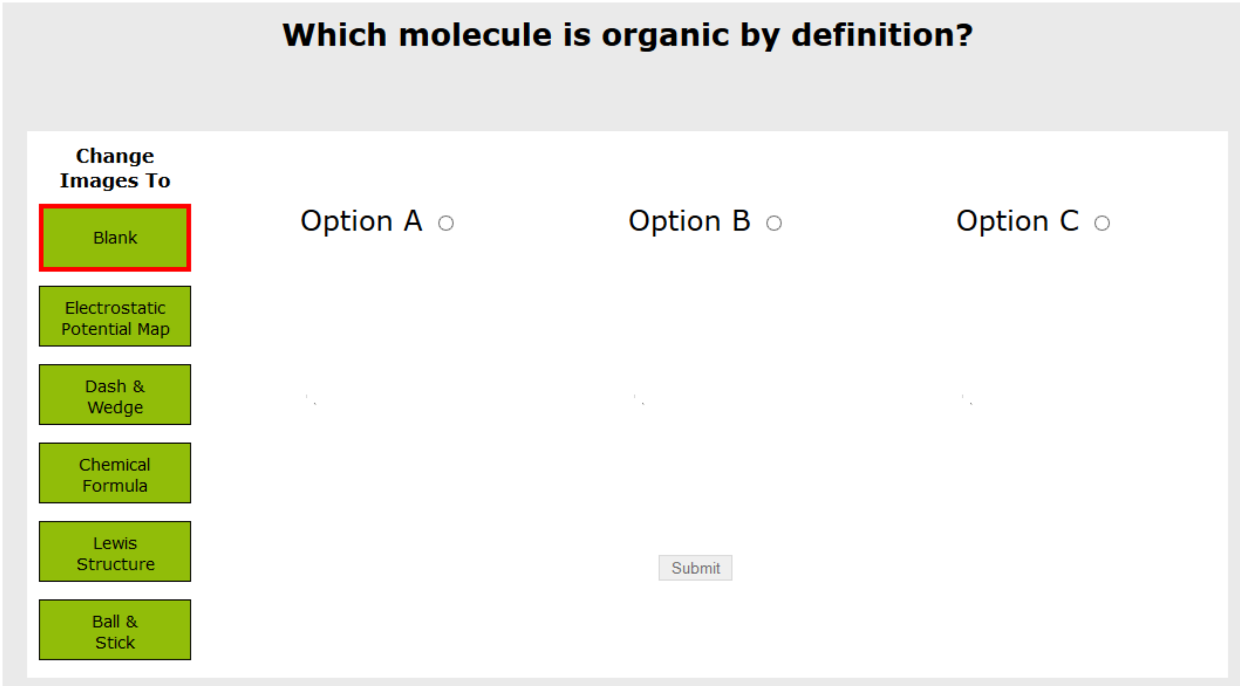

H2O
CO2
HCN
δ-
δ+
δ+
δ-

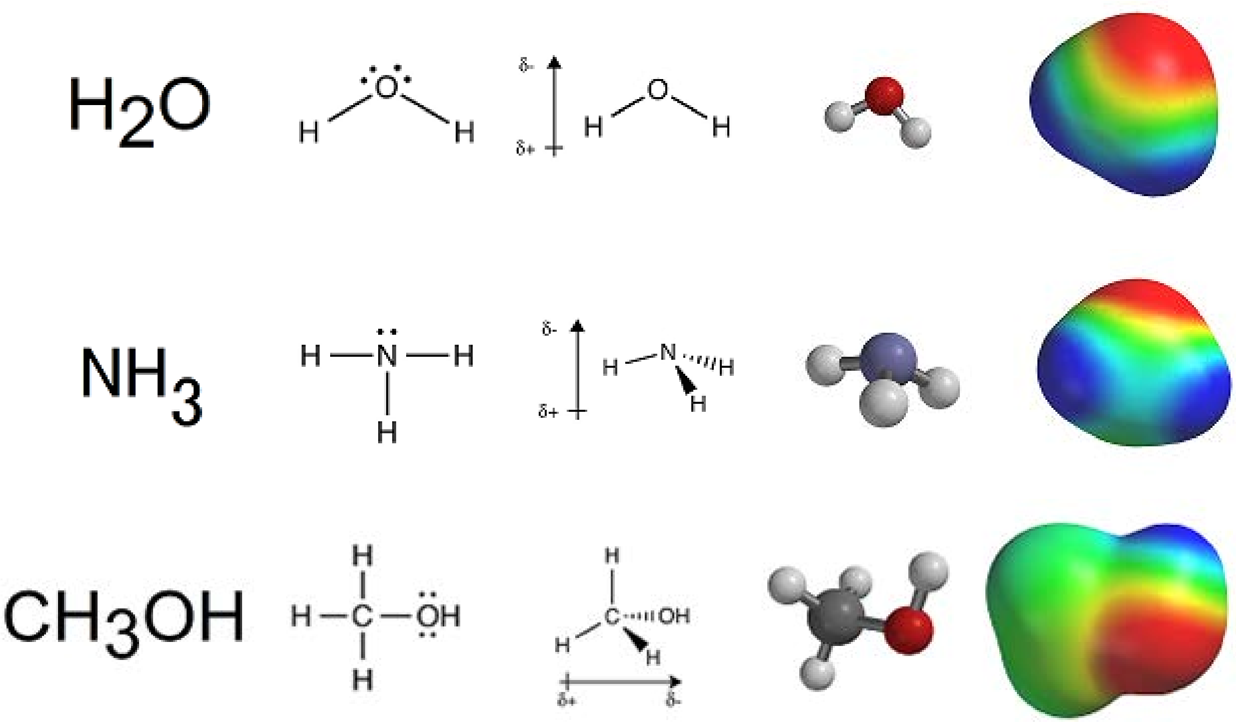
Question 4.)


Which molecule has a tetrahedral shape?
Change Images To
Blank
Electrostatic Potential Map
Ball & Stick
Lewis Structure
Chemical Formula
Dash & Wedge
Option A
Option B
Option C
Submit

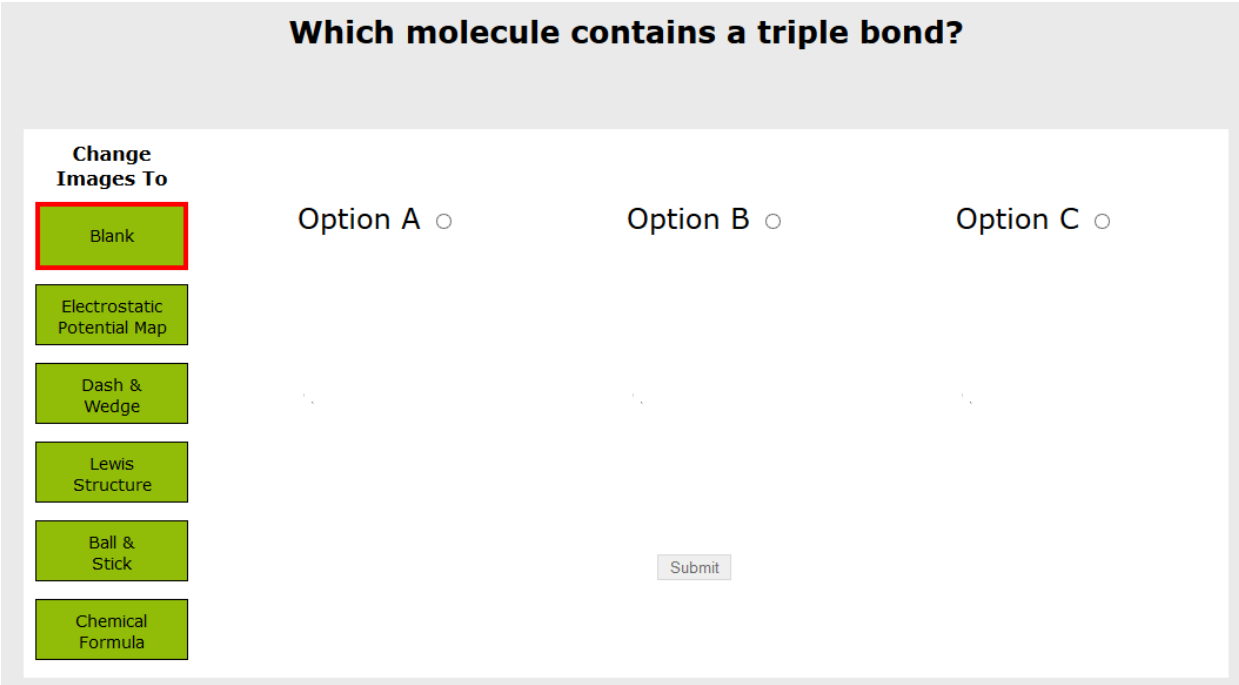

H2O
NH3
CH4
δ-
δ+

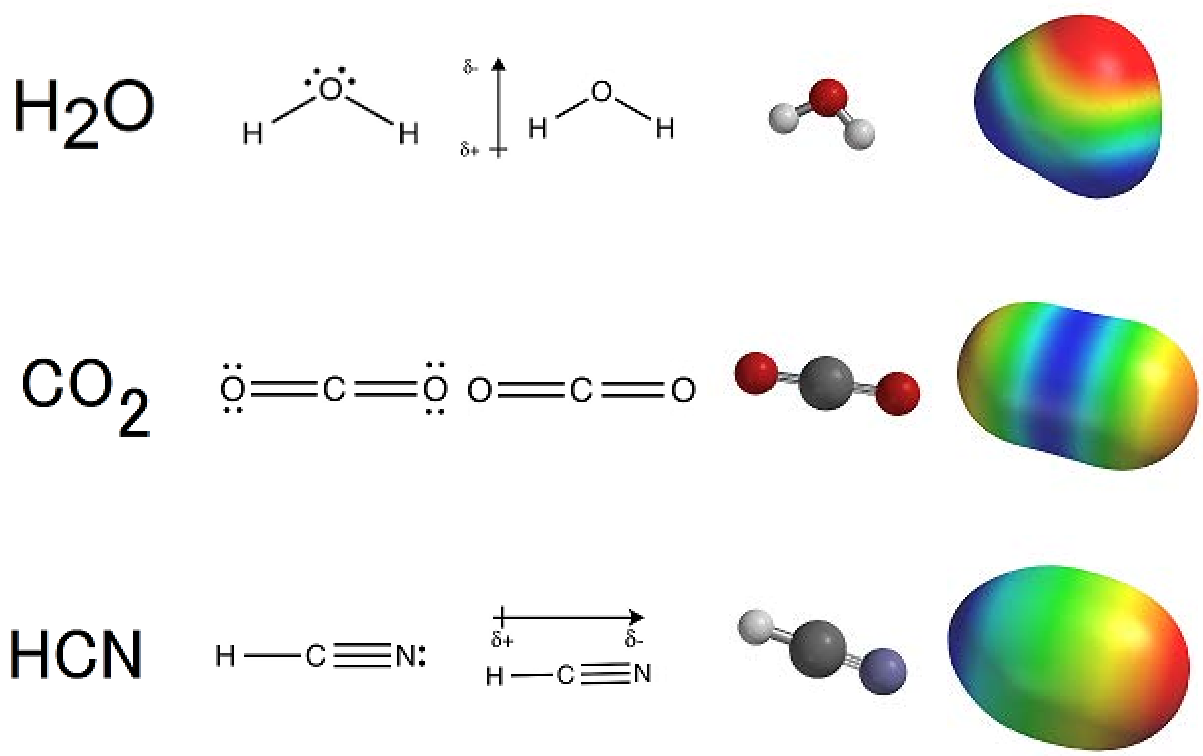
Question 5.)


Which molecule is organic by definition?
Change Images To
Blank
Electrostatic Potential Map
Dash & Wedge
Chemical Formula
Lewis Structure
Ball & Stick
Option A
Option B
Option C
Submit

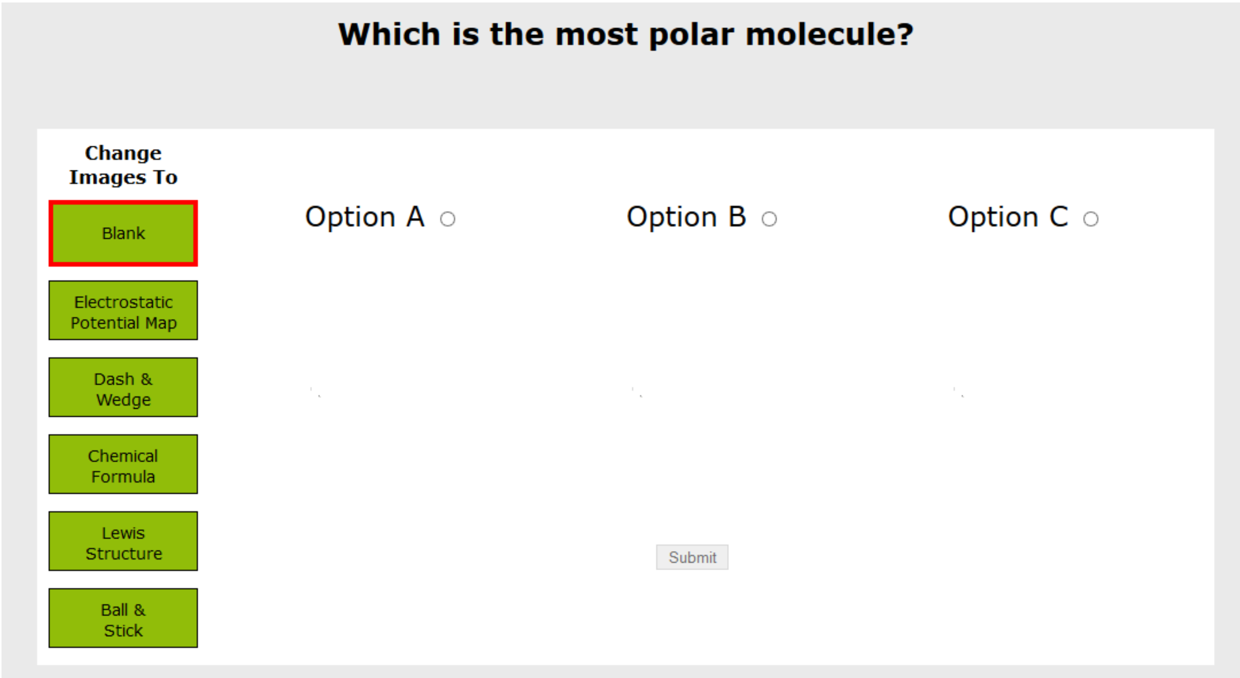

H2O
NH3
CH3OH
δ-
δ+

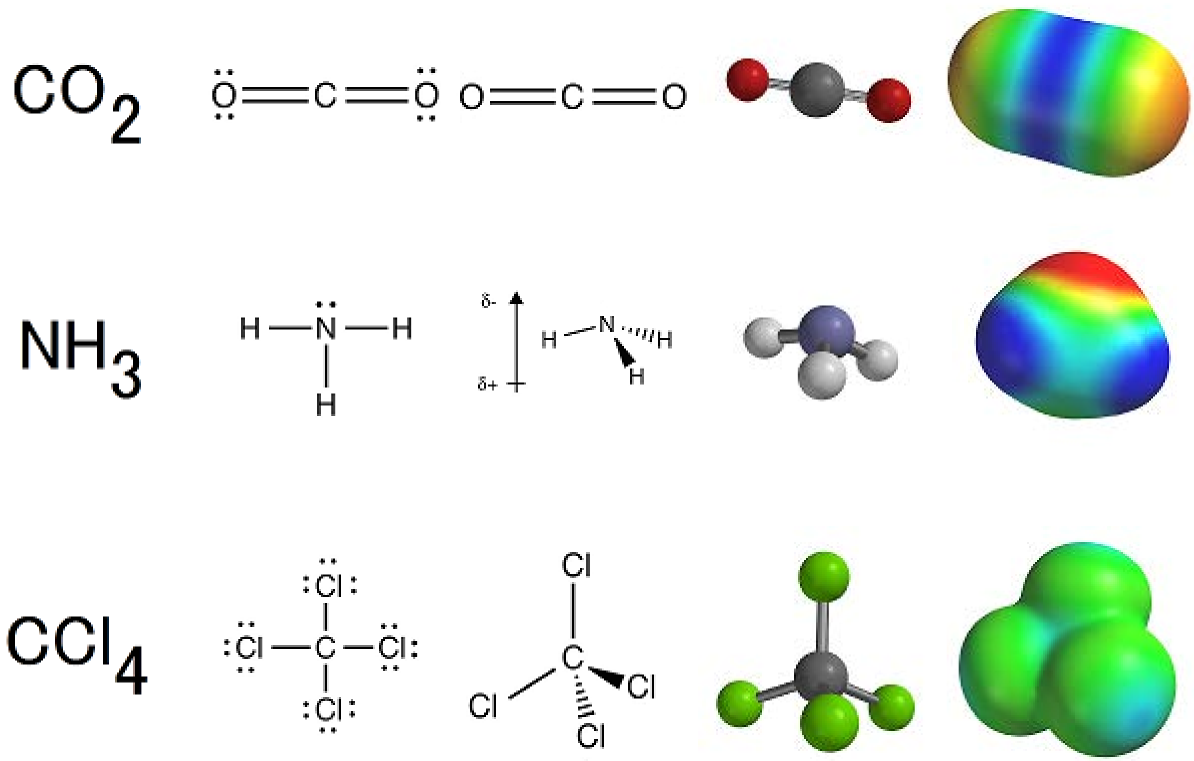
Question 6.)


Which molecule contains a triple bond?
Change Images To
Blank
Electrostatic Potential Map
Dash & Wedge
Lewis Structure
Ball & Stick
Chemical Formula
Option A
Option B
Option C
Submit

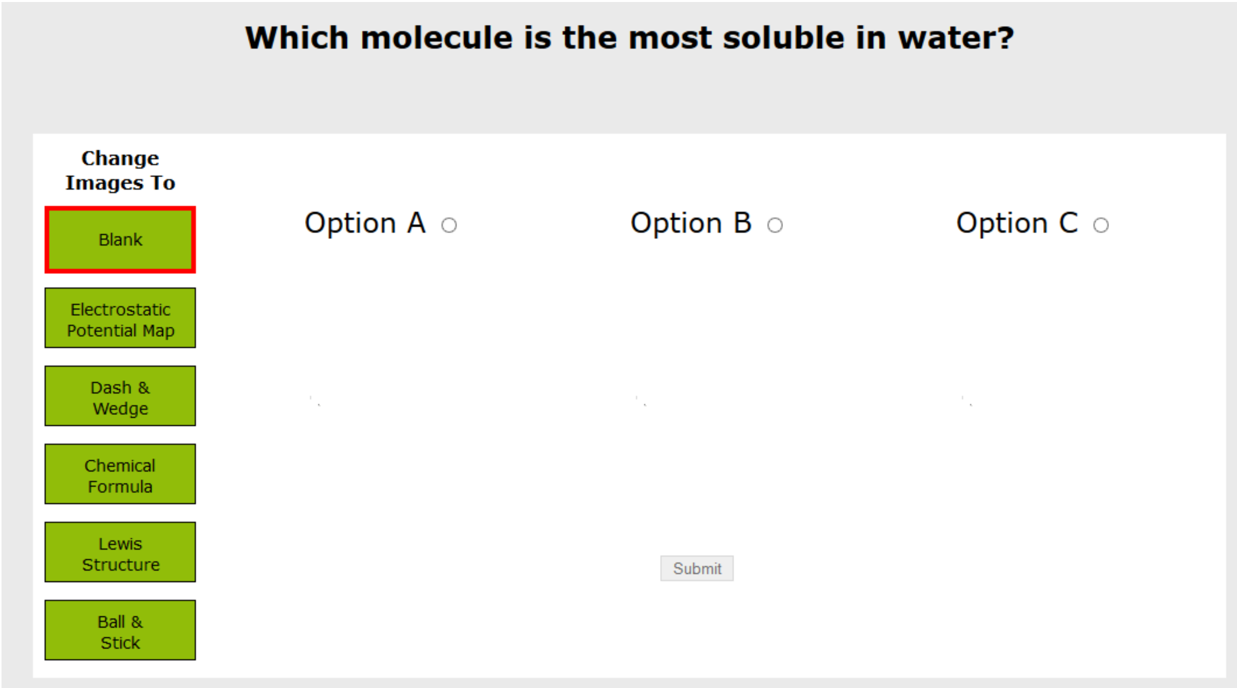

H2O
H
O
H
δ-
δ+
CO2
O=C=O
HCN
H—C≡N:
δ+
δ-
H—C≡N

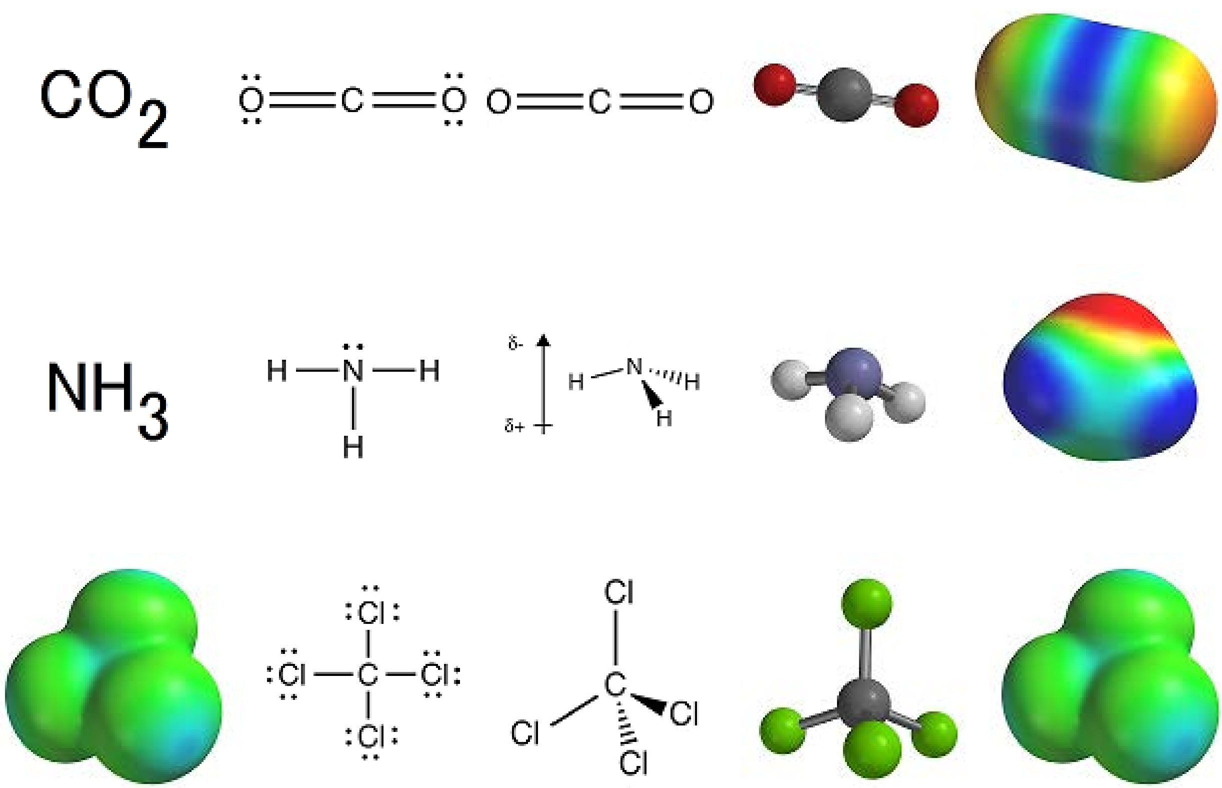
Question 7.)


Which is the most polar molecule?
Change Images To
Blank
Electrostatic Potential Map
Dash & Wedge
Chemical Formula
Lewis Structure
Ball & Stick
Option A
Option B
Option C
Submit

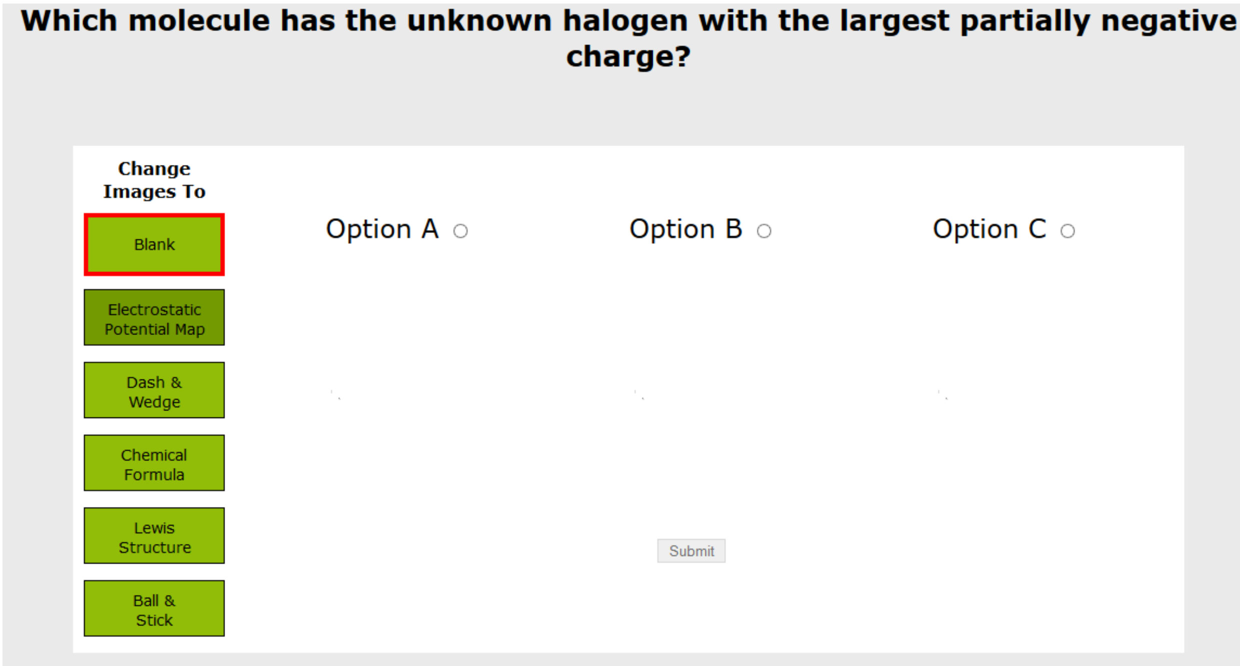

CO2
NH3
CCl4

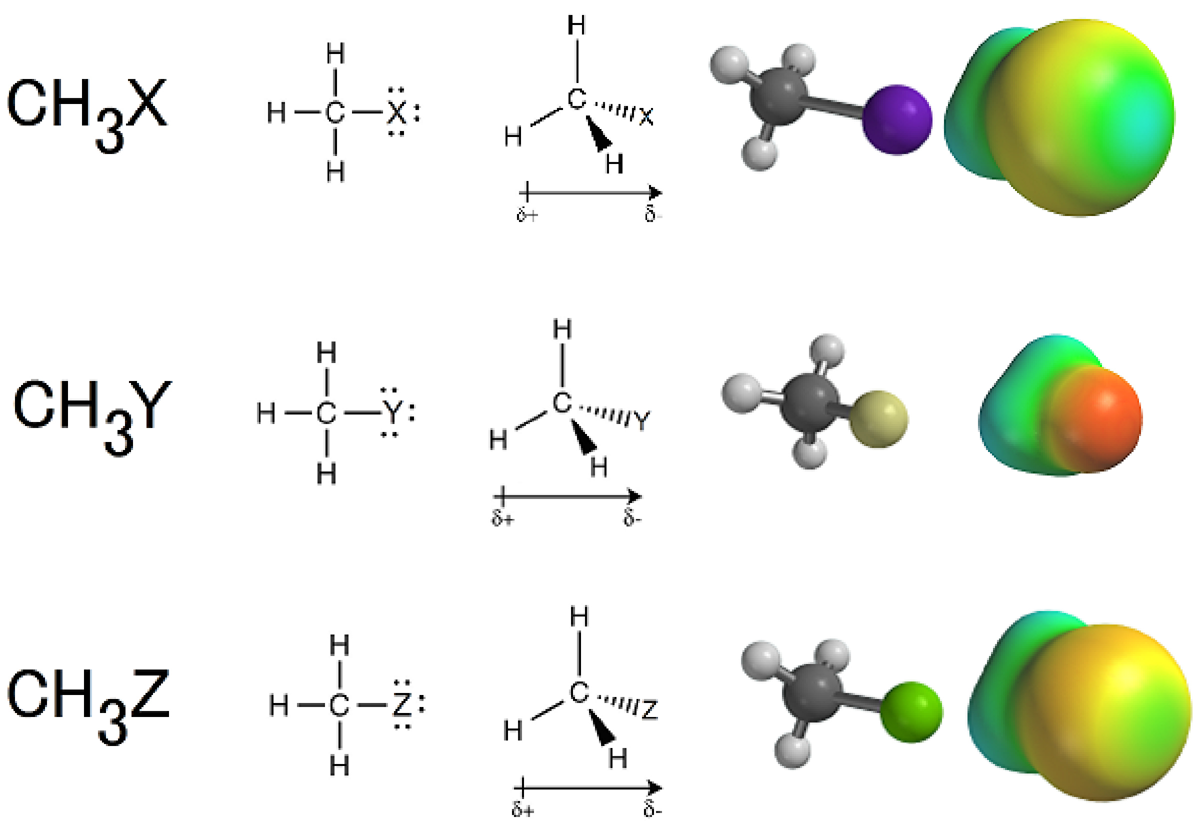
Question 8.)


Which molecule is the most soluble in water?
Change Images To
Blank
Electrostatic Potential Map
Dash & Wedge
Chemical Formula
Lewis Structure
Ball & Stick
Option A
Option B
Option C
Submit

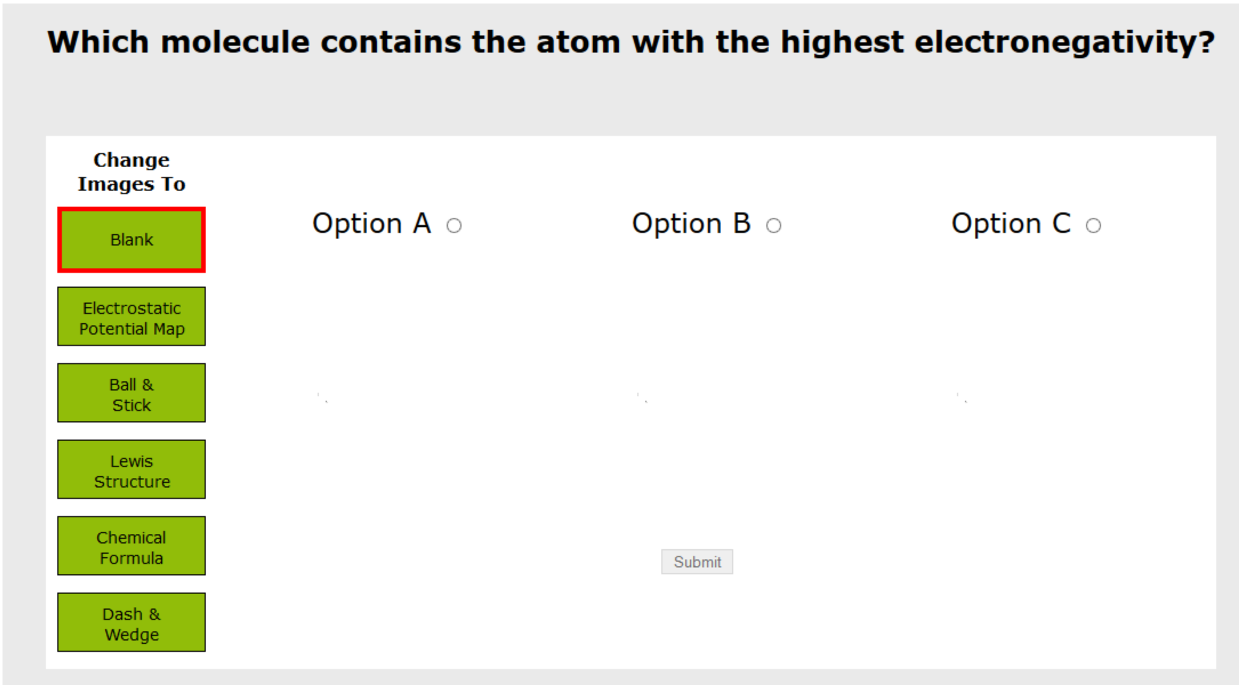

CO2
O=C=O
NH3
H—N—H
H
δ-
δ+
Cl
C

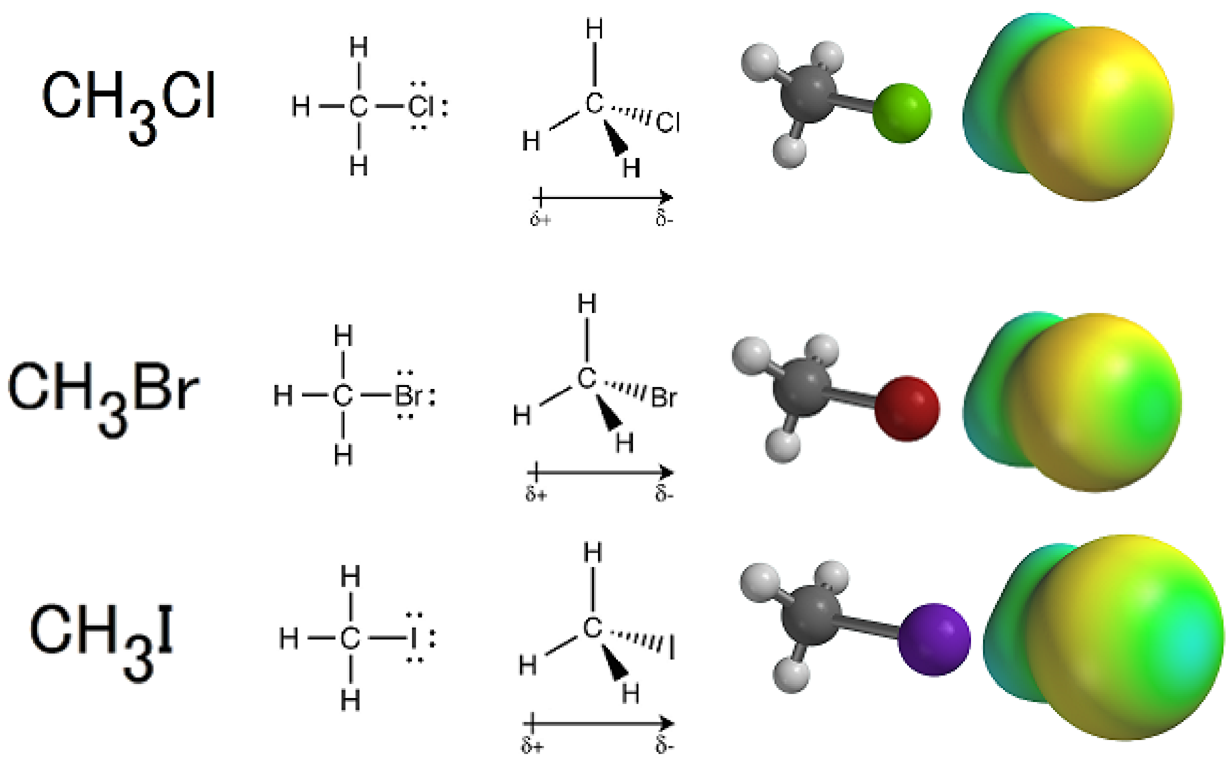
Question 9.)


Which molecule has the unknown halogen with the largest partially negative charge?
Change Images To
Blank
Electrostatic Potential Map
Dash & Wedge
Chemical Formula
Lewis Structure
Ball & Stick
Option A
Option B
Option C
Submit

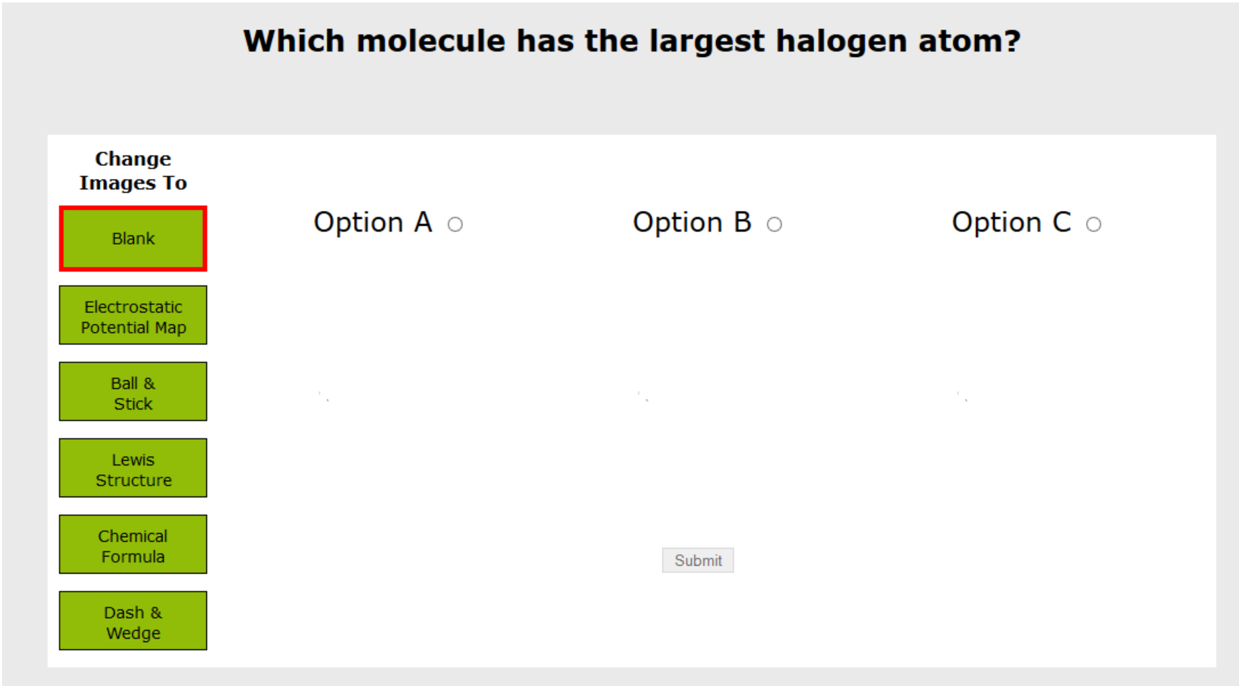

CH3X
CH3Y
CH3Z
H
C
X
Y
Z
δ+
δ-

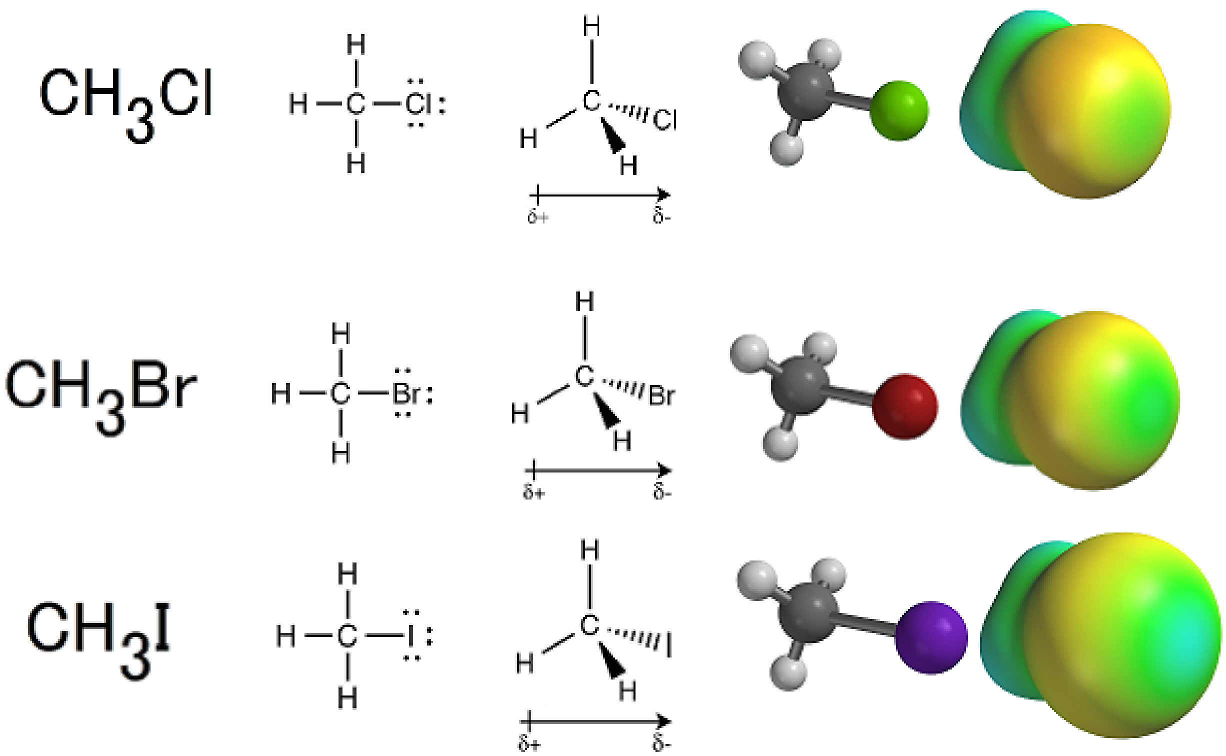
Question 10.)


Which molecule contains the atom with the highest electronegativity?
Change Images To
Blank
Electrostatic Potential Map
Ball & Stick
Lewis Structure
Chemical Formula
Dash & Wedge
Option A
Option B
Option C
Submit


CH3Cl
CH3Br
CH3I
δ+
δ-

Question 11.)


Which molecule has the largest halogen atom?
Change Images To
Blank
Electrostatic Potential Map
Ball & Stick
Lewis Structure
Chemical Formula
Dash & Wedge
Option A
Option B
Option C
Submit


CH3Cl
CH3Br
CH3I
δ+
δ-

APPENDIX L

PAPER ALTERNATIVE FOR PILOT STUDY

Diagnostic Test Jan, 23, 2015

Name:____________________

Circle the answer which best answers the question. Images for each molecule are given on another sheet of paper.

1.) Which molecule contains the strongest bond?

H20 CO2 HCN

2.) Which molecule is least likely to participate in hydrogen bonding?

NH3 CH4 H20

3.) Which molecule is non-linear?

H20 CO2 HCN

4.) Which molecule has a tetrahedral shape?

NH3 H2O CH4

5.) Which molecule is organic by definition?

NH3 H2O CH3OH

6.) Which molecule contains a triple bond?

H2O CO2 HCN

7.) Which is the most polar molecule?

CO2 CH3OH CCl4

8.) Which molecule is the most soluble in water?

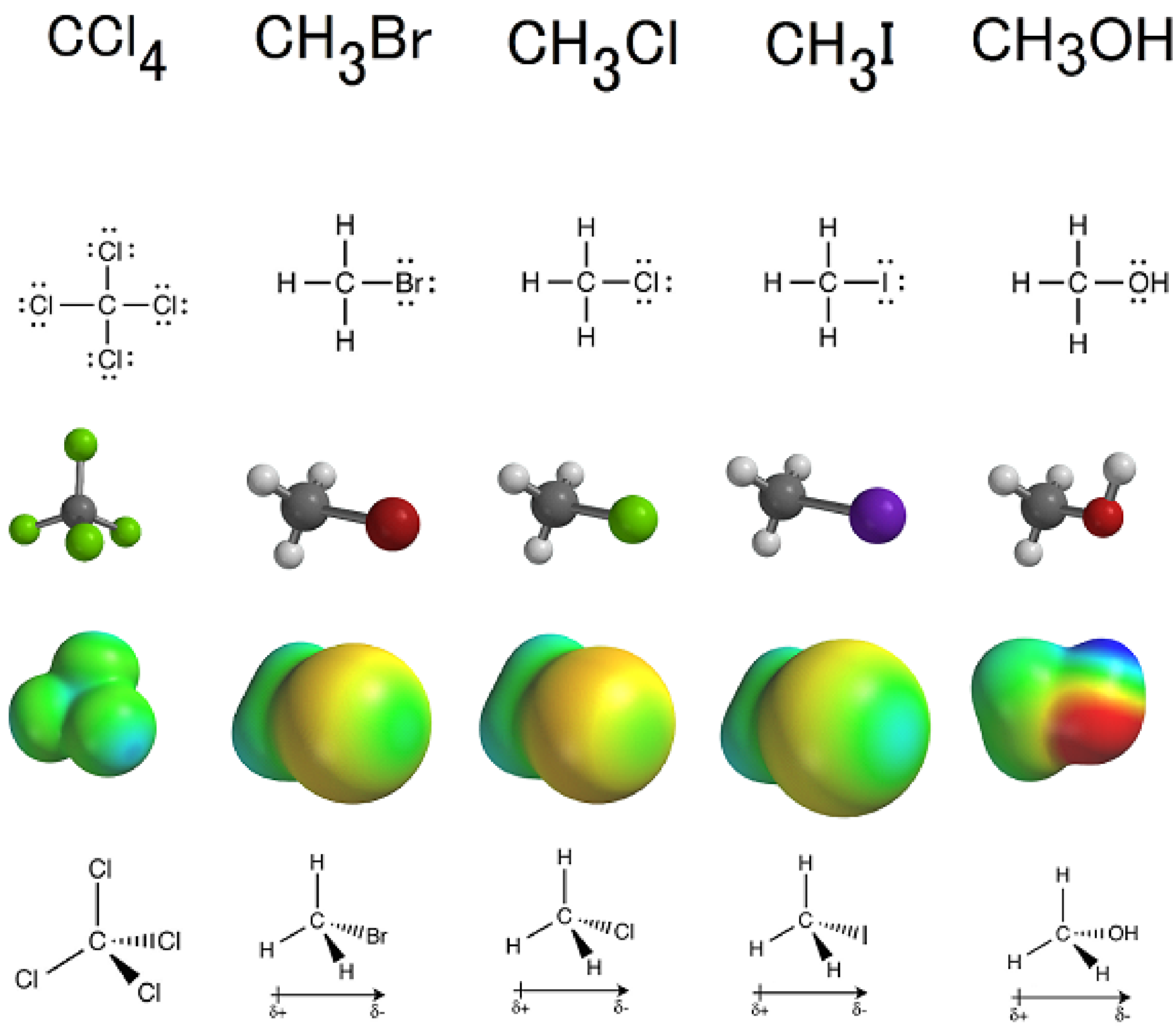
CO_2 CH_3OH CCl_4

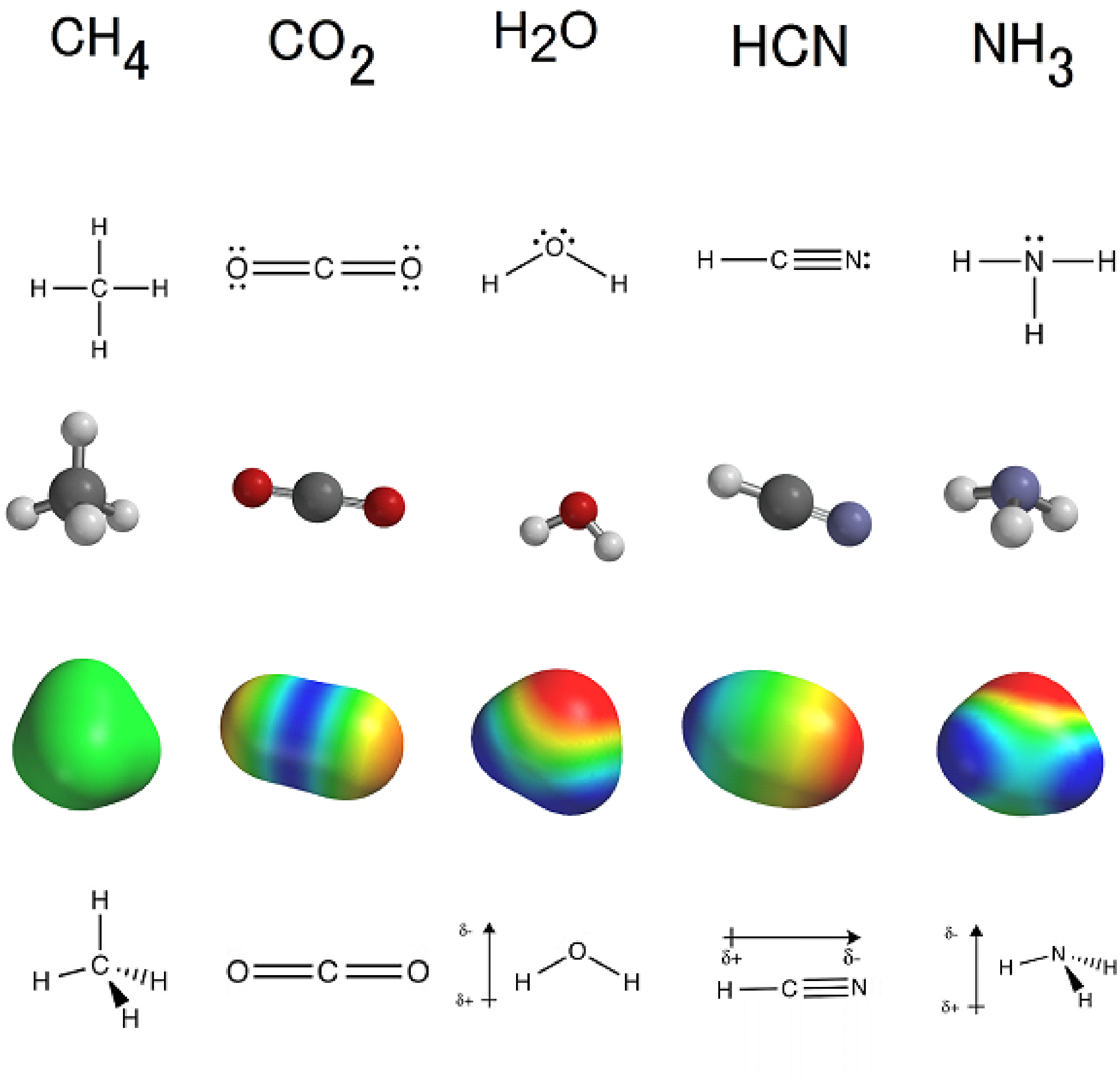
9.) Which molecule has the largest (by size) partially negative charged atom?

CH_3Cl CH_3Br CH_3I

10.) Which molecule contains the atom with the highest electronegativity?

CH_3Cl CH_3Br CH_3I


CCl4
CH3Br
CH3Cl
CH3I
CH3OH


CH4
CO2
H2O
HCN
NH3
δ-
δ+
δ-
δ+
δ-
δ+

APPENDIX M

PAPER VERSION OF THE CONSENT FORM FOR PILOT STUDY

Consent Form

Thank you for taking the time to work with the beta-test of a new tutorial and assessment system about visualizing chemistry representations and participate in this research study. This initial page will be used to outline the rationale for this study. Please take your time in deciding if you would like to participate, and feel free to contact the researchers with any questions at any time.

Introduction

The purpose of this study is to conduct a thorough investigation into how on-line tools can be constructed to help students visualize different representations in chemistry. These specific tools are designed to give you the option to look at different aspects of molecules so that you can choose visual representations that you already understand well, perhaps from your experience in previous courses. Concepts in chemistry often influence understanding in other areas of science, and often these concepts involve several layers of information about how molecules are visualized. This on-line system uses Digital Learning Objects (DLOs) that are designed to build educational materials that explicitly help students.

Description of procedures

If you agree to participate in this beta-testing component of this study, you will be asked to answer questions within a Digital Learning Object (DLO). The system will include the ability to access visual information about molecules using different models of the molecules involved so that you can explore the way the molecules are represented until you find the ways that most help you answer the questions about the chemistry in the DLO. During the study you may expect the following study procedures to be followed

National Beta Test of the System

1. You will first be provided with a brief description of the basic navigational elements of the system.
2. From this point forward, each page will have a frame at the bottom that will remind the you that you may stop at any time along with a link to allow you to exit the system.
3. Once you feel comfortable with the navigational instructions, you will be able to click on a link to start interacting with the DLOs on the site themselves.
4. You will be allowed to interact with the elements present in the site at your own pace. The software will capture any navigational actions (links clicked, etc.) as well as responses to questions that are part of the DLO.

5. Feedback on the chemical accuracy of your responses will be provided during the use of the DLO. Before the screen that provides the feedback (to tell you if you answered a question right or wrong, for example), a text-box that will allow you to provide, open-response, feedback about the DLO.
6. This process, with questions related to different components of the chemistry content will continue for up to eleven questions.
7. When you complete the final DLO question, including receiving the feedback and having the change to provide us with feedback about the system in a text-box, the system will let you know that you reached the end of the beta-test.
8. You will have a chance to see a summary of their responses and you can check to make sure the accurately reflect your impressions before you exit the system.
9. You will have a chance to fill out some basic demographic information that may help us design better tools to help student, or improve the design of the current tools.

Risks

There are no known physical, psychological, social, or medical risks associated with this research. Any information you provide will be kept confidential, and if this work is to be published, your name will not be used.

Benefits

If you decide to participate in this study there will be no direct benefit to you. Although, it is hoped that the information you provide will help the American Chemical Society Exams Institute create an assessment tool that is both valid and reliable and develop general chemistry exams that reflect the types of representations used in a general chemistry course of the 21st century.

Costs and Compensation

You will not incur any costs from participating in this study. You may be receiving some form of homework or extra credit from an instructor for participating in this project. If that is the case, you will be requested to enter your email and your email address will only be used to report to your instructor that you participated in this study. It will NOT be used as any kind of identifier, and it will be discarded once your participation has been verified with your instructor.

Participant Rights

Your participation in this study is completely voluntary and you may refuse to participate or leave the study at any time. If you decide to not participate in the study or leave the study early, it will not result in any penalty or loss of benefits to which you are otherwise entitled.

Confidentiality

Records identifying participants will be kept confidential to the extent permitted by applicable laws and regulations and will not be made publicly available. However, auditing departments of Iowa State University, and the Institutional Review Board (a committee that reviews and approves human subject research studies) and representatives from the National Science Foundation (who funded the study) may inspect and/or copy your records for quality assurance and data analysis. These records may contain private information.

To ensure confidentiality to the extent permitted by law, the following measures will be taken: You will be asked to provide basic demographic information, but at no point during the duration of this study will your name be associated with any input you provide. All data will be stored in a password-protected file on a computer requiring administrative credentials in order to access. A backup file requiring the same administrative credentials will be kept on an external hard drive that will remain locked in a filing cabinet when not in use. These data will be kept indefinitely for analysis. If the results are published, your identity will remain confidential.

Questions or problems

You are encouraged to ask questions at any time during this study.

For further information about the study, contact Jack Polifka by email at jpolifka@iastate.edu or Thomas Holme by taholme@iastate.edu.

If you have any questions about the rights of research subjects or research-related injury, please contact IRB Administrator, (515) 294-4566, IRB@iastate.edu, or Director, Office for Responsible Research, (515) 294-3115, 1138 Pearson Hall, Ames, IA 50011.

Participant Signature

Typing your name below indicates that you voluntarily agree to participate in this study, that the study has been explained to you, that you have been given the time to read the document, and that your questions have been satisfactorily answered. Please note, if you are being provided some form of credit by your course instructor for participation, you must also provide your email address for us to forward your participation to the instructor. Only your participation is to be indicated, your individual answers will not be reported to your instructor. Your email will be used only for verification of your participation. Your name and email will not be associated with your responses in any data that is used in the analysis of student interaction with the DLOs of this system.

The online form will ask for the following:

Signature: (If you wish to voluntarily participate in this study.)

Email address for reporting participation only:

APPENDIX N

MEAN, STANDARD DEVIATION, AND STATISTICAL ANALYSIS OF USED TIME IN SECONDS FOR ALL REPRESENTATIONS FOR EACH QUESTION – PILOT STUDY PARTICIPANTS WHO ANSWERED CORRECTLY

Question	CF [a]	LS [b]	D&W [c]	B&S [d]	EPM [e]	Total Time	Friedman Test *p*-value	Pairwise Wilcoxon Test *p*-value [f]
1	1.2 (2.6)	**3.9 (4.7)**	**2.5 (4.3)**	1.1 (2.1)	1.0 (1.9)	13.1 (8.5)	< .0001 [g]	0.0034 [g]
2	2.7 (4.8)	**5.7 (7.3)**	**2.7 (5.5)**	1.1 (2.5)	1.0 (2.8)	18.3 (13.0)	< .0001 [g]	< .0001 [g]
3	0.6 (1.4)	**2.1 (2.6)**	1.4 (2.6)	**1.6 (2.6)**	0.5 (1.2)	9.5 (5.0)	< .0001 [g]	0.3290
4	0.8 (1.8)	**3.0 (4.8)**	2.0 (3.4)	**2.4 (3.2)**	0.9 (2.2)	12.5 (6.7)	< .0001 [g]	> .9999
5	**2.9 (4.4)**	**2.5 (3.5)**	1.5 (3.8)	0.8 (1.8)	0.5 (1.4)	11.7 (7.1)	< .0001 [g]	> .9999
6	0.7 (1.7)	**1.8 (1.8)**	**1.0 (1.8)**	0.6 (1.8)	0.3 (0.9)	7.8 (5.7)	< .0001 [g]	0.0014 [g]
7	1.3 (3.3)	**3.4 (5.2)**	**3.2 (4.6)**	1.4 (2.8)	2.6 (3.3)	15.5 (8.8)	< .0001 [g]	> .9999
8	**4.1 (7.4)**	**5.2 (6.3)**	3.1 (6.0)	1.8 (3.2)	2.0 (5.1)	20.1 (14.8)	< .0001 [g]	0.0032 [g]
9	2.9 (4.2)	**5.7 (8.0)**	5.2 (6.6)	3.3 (5.9)	**7.9 (10.2)**	31.6 (15.5)	< .0001 [g]	0.1000
10	**4.7 (8.0)**	**6.1 (8.9)**	4.7 (9.5)	1.9 (6.2)	3.3 (5.0)	26.2 (18.9)	< .0001 [g]	0.7400
11	**2.8 (5.8)**	**3.9 (6.2)**	1.5 (3.4)	1.8 (2.8)	2.4 (2.5)	16.8 (9.3)	< .0001 [g]	0.1800

[a] CF: chemical formula, [b] LS: Lewis structure, [c] D&W: dash-and-wedge, [d] B&S: ball-and-stick, [e] EPM: electrostatic potential map,
[f] With Bonferroni correction between most and second most used viewed representation, [g] Significant *p*-value

Most Used Representation, **Second Most Used Representation**

APPENDIX O

MEAN, STANDARD DEVIATION, AND STATISTICAL ANALYSIS OF USED TIME IN SECONDS FOR ALL REPRESENTATIONS FOR EACH QUESTION – PILOT STUDY PARTICIPANTS WHO ANSWERED INCORRECTLY

Question	CF [a]	LS [b]	D&W [c]	B&S [d]	EPM [e]	Total Time	Friedman Test *p*-value	Pairwise Wilcoxon Test *p*-value [f]
1	2.3 (4.4)	**4.8 (6.1)**	**3.0 (4.3)**	1.6 (3.1)	2.1 (3.2)	16.9 (7.9)	0.0006 [g]	0.7900
2	2.3 (4.1)	**5.6 (5.8)**	**2.8 (5.9)**	1.1 (2.2)	0.7 (1.3)	17.4 (11.4)	< .0001 [g]	0.0020 [g]
3	1.6 (4.6)	**6.0 (11.9)**	**3.3 (4.7)**	1.1 (2.3)	0.4 (1.2)	15.6 (13.4)	0.0038 [g]	> .9999
4	**3.8 (12.1)**	**6.2 (7.7)**	1.6 (3.7)	2.5 (4.3)	0.5 (1.2)	17.9 (12.6)	0.0001 [g]	0.1800
5	**4.2 (4.4)**	**3.7 (5.4)**	2.4 (4.3)	1.2 (1.9)	1.1 (2.1)	16.0 (8.3)	< .0001 [g]	> .9999
6								
7	2.2 (3.9)	**5.7 (7.6)**	**2.9 (4.3)**	2.0 (2.9)	2.6 (7.3)	18.6 (11.2)	0.0021 [g]	0.2600
8	**5.8 (6.2)**	**5.0 (5.9)**	2.6 (5.0)	1.7 (2.7)	2.0 (4.2)	20.5 (11.6)	< .0001 [g]	> .9999
9	4.2 (5.5)	**7.4 (10.9)**	5.8 (6.5)	3.8 (5.4)	**6.4 (9.7)**	34.5 (17.6)	< .0001 [g]	> .9999
10	3.6 (6.5)	**4.2 (7.5)**	3.6 (4.9)	1.7 (2.9)	**4.1 (3.7)**	21.6 (12.1)	< .0001 [g]	0.7000
11	**3.8 (5.8)**	**5.9 (7.0)**	2.2 (4.2)	1.7 (4.0)	0.8 (1.8)	18.0 (9.2)	< .0001 [g]	> .9999

[a] CF: chemical formula, [b] LS: Lewis structure, [c] D&W: dash-and-wedge, [d] B&S: ball-and-stick, [e] EPM: electrostatic potential map,
[f] With Bonferroni correction between most and second most used viewed representation, [g] Significant *p*-value
Most Used Representation, **Second Most Used Representation**

APPENDIX P

MEAN, STANDARD DEVIATION, AND STATISTICAL ANALYSIS OF USED TIME IN SECONDS FOR ALL REPRESENTATIONS FOR EACH QUESTION – PILOT STUDY PARTICIPANTS WHO ANSWERED SLOW

Question	CF [a]	LS [b]	D&W [c]	B&S [d]	EPM [e]	Total Time	Friedman Test *p*-value	Pairwise Wilcoxon Test *p*-value [f]
1	2.0 (3.7)	**5.4 (5.9)**	**3.7 (5.1)**	1.7 (2.7)	1.9 (2.6)	14.6 (8.4)	< .0001 [g]	0.0427 [g]
2	3.5 (5.2)	**7.0 (7.8)**	**3.7 (6.6)**	1.4 (2.7)	1.3 (2.8)	16.7 (10.0)	< .0001 [g]	< .0001 [g]
3	0.9 (2.0)	**3.0 (5.1)**	**2.3 (3.4)**	2.2 (3.0)	0.7 (1.4)	9.0 (6.0)	< .0001 [g]	> .9999
4	1.4 (4.8)	**4.0 (5.7)**	2.5 (3.7)	**2.7 (3.6)**	1.0 (2.4)	11.5 (7.3)	< .0001 [g]	0.3270
5	**3.7 (5.0)**	**3.3 (4.3)**	2.1 (4.4)	1.1 (2.0)	0.9 (1.7)	11.1 (7.4)	< .0001 [g]	> .9999
6	1.1 (2.1)	**2.0 (2.2)**	**1.4 (2.1)**	0.9 (2.3)	0.5 (1.0)	5.9 (3.0)	< .0001 [g]	> .9999
7	2.3 (4.2)	**5.2 (6.8)**	**4.2 (5.3)**	2.1 (3.3)	3.2 (5.4)	17.0 (8.9)	< .0001 [g]	> .9999
8	**6.5 (7.8)**	**6.2 (6.9)**	3.8 (6.5)	2.2 (3.4)	2.6 (5.6)	21.3 (13.8)	< .0001 [g]	> .9999
9	4.9 (6.0)	**10.6 (12.8)**	8.2 (7.6)	6.0 (7.1)	**10.1 (13.5)**	39.8 (16.0)	< .0001 [g]	> .9999
10	5.3 (8.2)	**6.6 (9.3)**	**5.6 (8.4)**	2.3 (5.5)	4.4 (4.8)	24.2 (15.1)	< .0001 [g]	> .9999
11	**3.9 (6.7)**	**5.8 (7.3)**	2.2 (4.2)	2.1 (3.6)	2.3 (2.7)	16.3 (8.5)	< .0001 [g]	0.0398 [g]

[a] CF: chemical formula, [b] LS: Lewis structure, [c] D&W: dash-and-wedge, [d] B&S: ball-and-stick, [e] EPM: electrostatic potential map, [f] With Bonferroni correction between most and second most used viewed representation, [g] Significant *p*-value

Most Used Representation, **Second Most Used Representation**

APPENDIX Q

MEAN, STANDARD DEVIATION, AND STATISTICAL ANALYSIS OF USED TIME IN SECONDS FOR ALL REPRESENTATIONS FOR EACH QUESTION – PILOT STUDY PARTICIPANTS WHO ANSWERED FAST

Question	CF [a]	LS [b]	D&W [c]	B&S [d]	EPM [e]	Total Time	Friedman Test *p*-value	Pairwise Wilcoxon Test *p*-value [f]
1	0.5 (1.3)	**2.1 (2.1)**	**0.9 (1.8)**	0.4 (1.3)	0.2 (0.9)	4.2 (1.3)	< .0001 [g]	0.0150 [g]
2	1.0 (2.1)	**3.2 (3.0)**	**1.0 (2.2)**	0.5 (1.4)	0.3 (0.8)	5.9 (1.6)	< .0001 [g]	0.0003 [g]
3	0.2 (0.9)	**1.4 (1.6)**	0.5 (1.2)	**0.8 (1.3)**	0.1 (0.6)	3.1 (0.9)	< .0001 [g]	0.0578
4	0.2 (0.8)	**1.1 (1.6)**	0.6 (1.4)	**1.2 (1.6)**	0.2 (0.8)	3.2 (1.0)	< .0001 [g]	> .9999
5	**1.6 (1.7)**	**1.0 (1.5)**	0.2 (1.0)	0.3 (1.0)	0.1 (0.6)	3.3 (0.8)	< .0001 [g]	0.6325
6	0.2 (0.8)	**1.5 (1.3)**	**0.4 (1.0)**	0.2 (0.7)	0.1 (0.5)	2.4 (0.8)	< .0001 [g]	> .9999
7	0.3 (1.0)	**1.7 (2.8)**	**1.5 (2.4)**	0.6 (1.5)	1.5 (2.2)	5.6 (1.9)	< .0001 [g]	> .9999
8	**1.5 (2.3)**	**3.0 (3.1)**	1.2 (2.3)	0.8 (1.5)	0.7 (1.7)	7.1 (2.2)	< .0001 [g]	0.0157 [g]
9	2.7 (3.9)	**3.6 (4.6)**	**3.3 (4.4)**	1.6 (2.8)	4.6 (4.1)	15.8 (5.9)	< .0001 [g]	0.2790
10	1.3 (2.7)	**1.8 (2.9)**	**0.8 (2.0)**	0.7 (1.9)	2.2 (2.7)	6.7 (2.4)	< .0001 [g]	> .9999
11	1.2 (2.2)	**1.4 (2.2)**	**0.7 (1.4)**	1.0 (1.8)	1.5 (1.9)	5.8 (1.6)	< .0001 [g]	> .9999

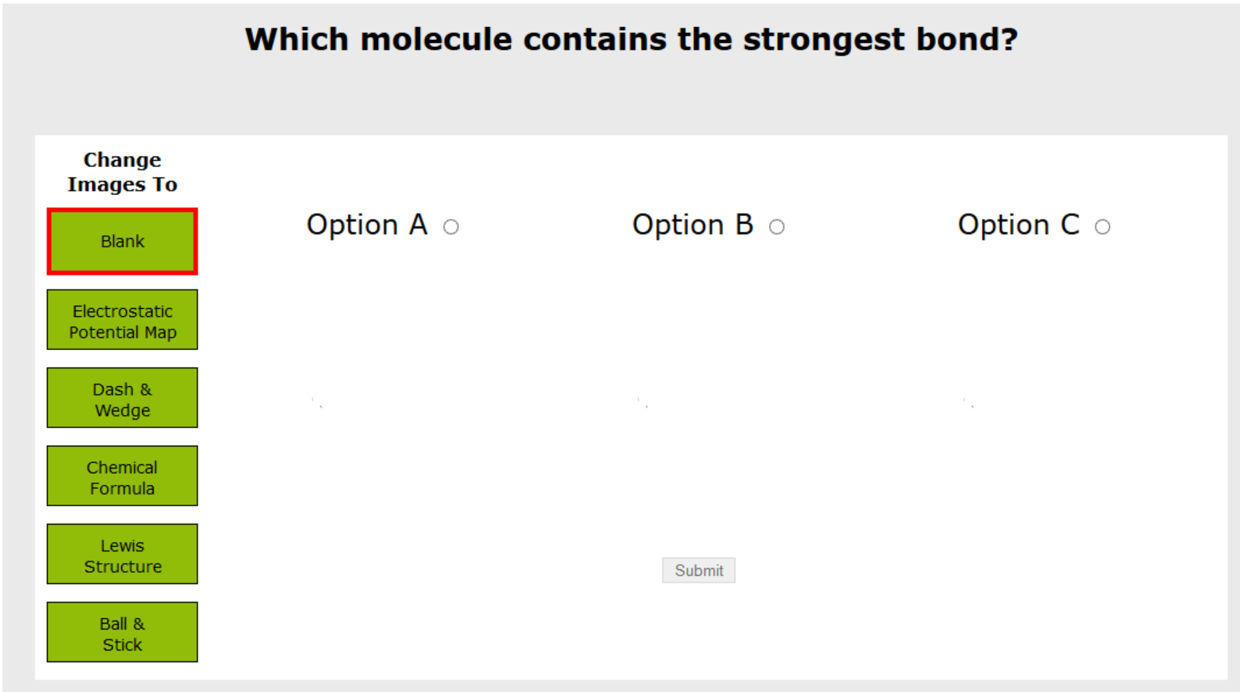
[a] CF: chemical formula, [b] LS: Lewis structure, [c] D&W: dash-and-wedge, [d] B&S: ball-and-stick, [e] EPM: electrostatic potential map, [f] With Bonferroni correction between most and second most used viewed representation, [g] Significant *p*-value

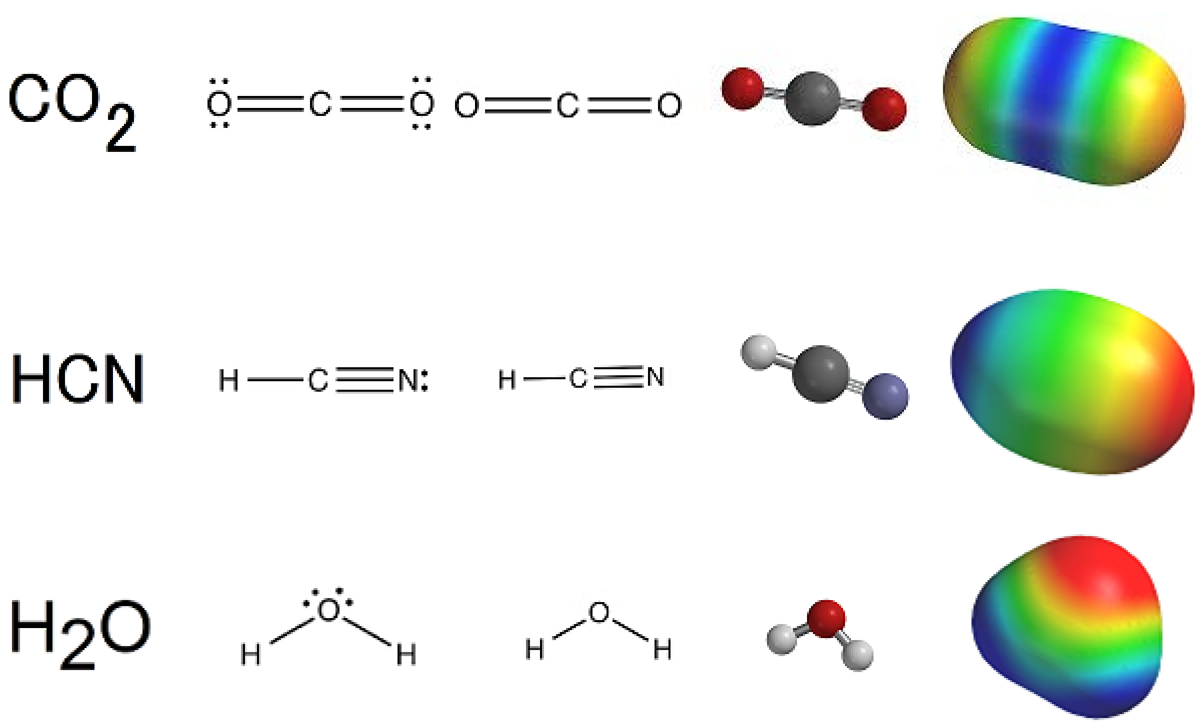
Most Used Representation, **Second Most Used Representation**

APPENDIX R

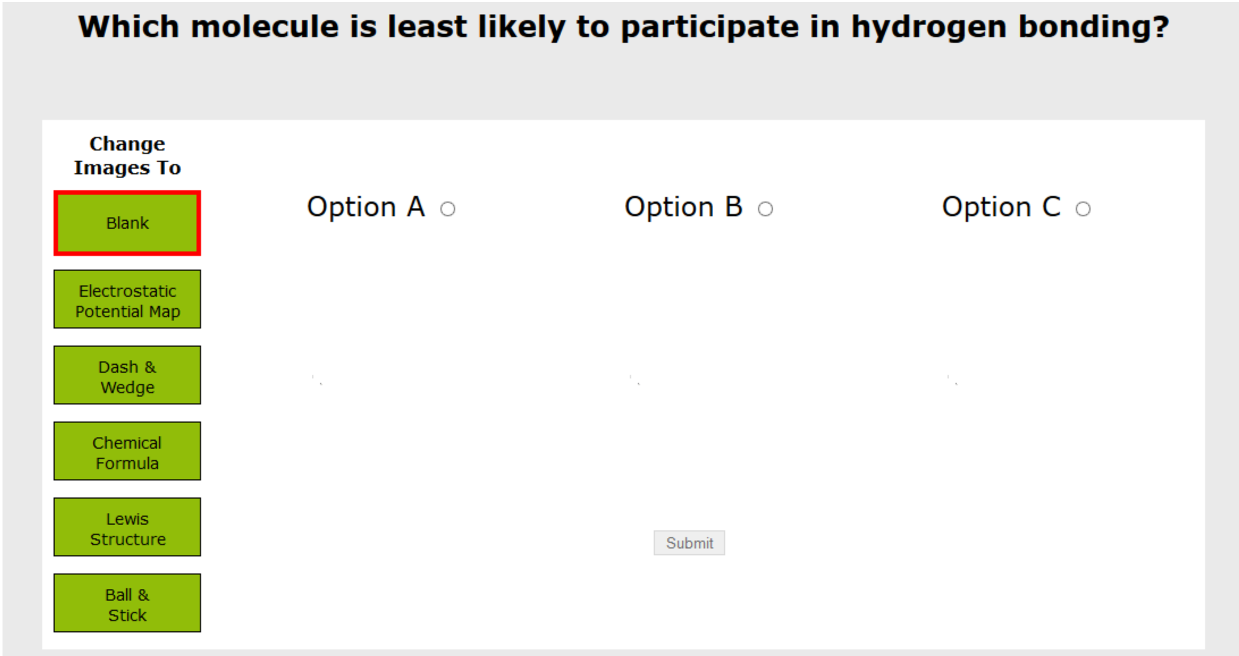
QUESTIONS, ANSWER CHOICES AND REPRESENTATION IMAGES FOR FULL EXPERIMENT

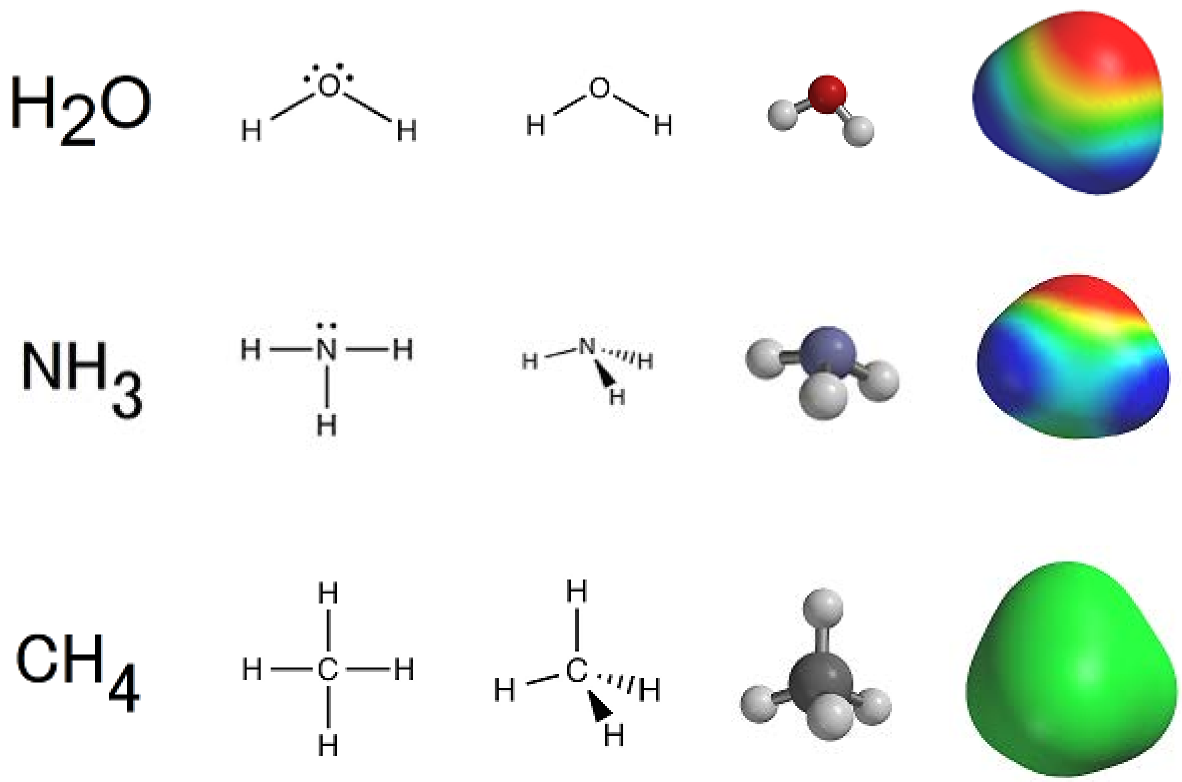
Question 1.)



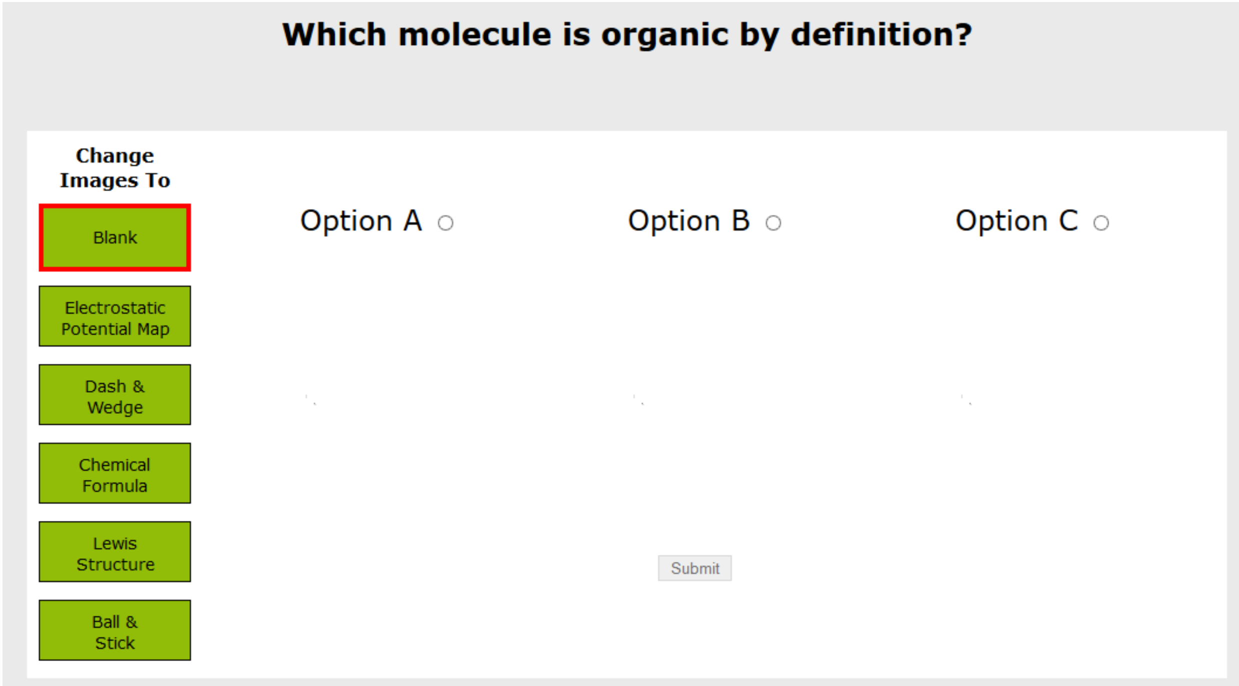


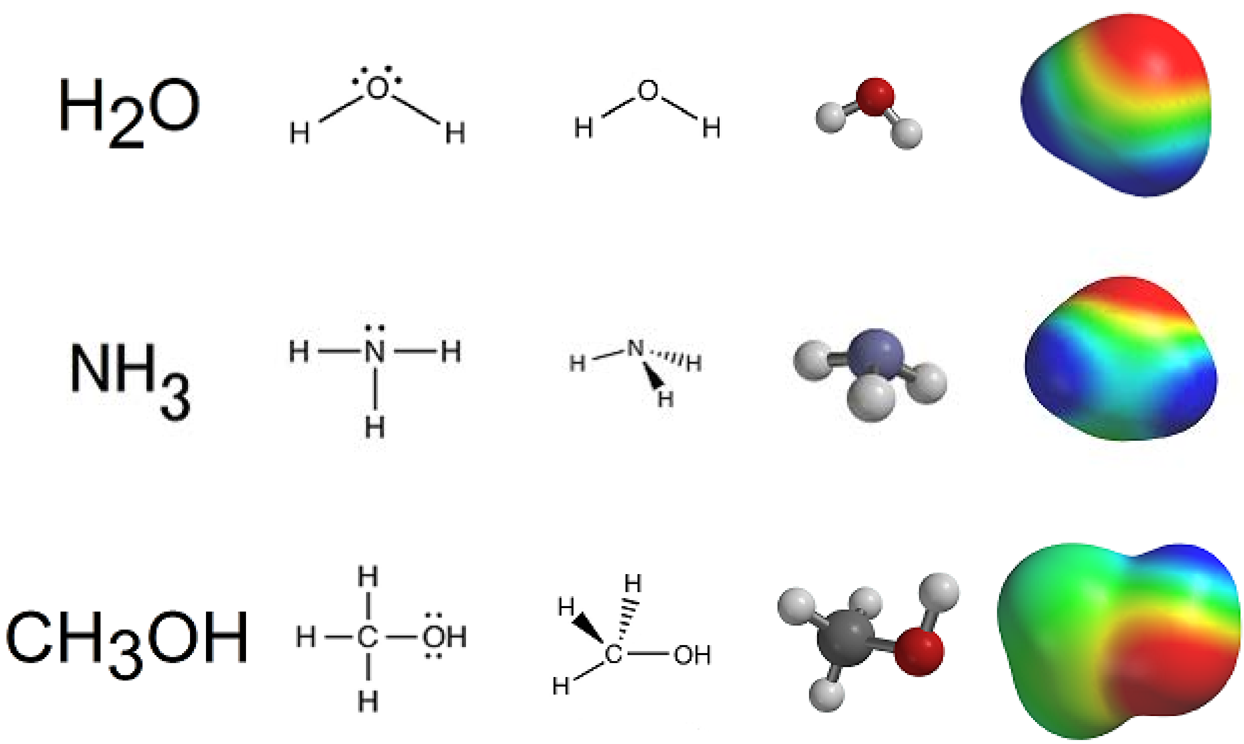
Question 2.)


Which molecule is least likely to participate in hydrogen bonding?
Change Images To
Blank
Electrostatic Potential Map
Dash & Wedge
Chemical Formula
Lewis Structure
Ball & Stick
Option A
Option B
Option C
Submit

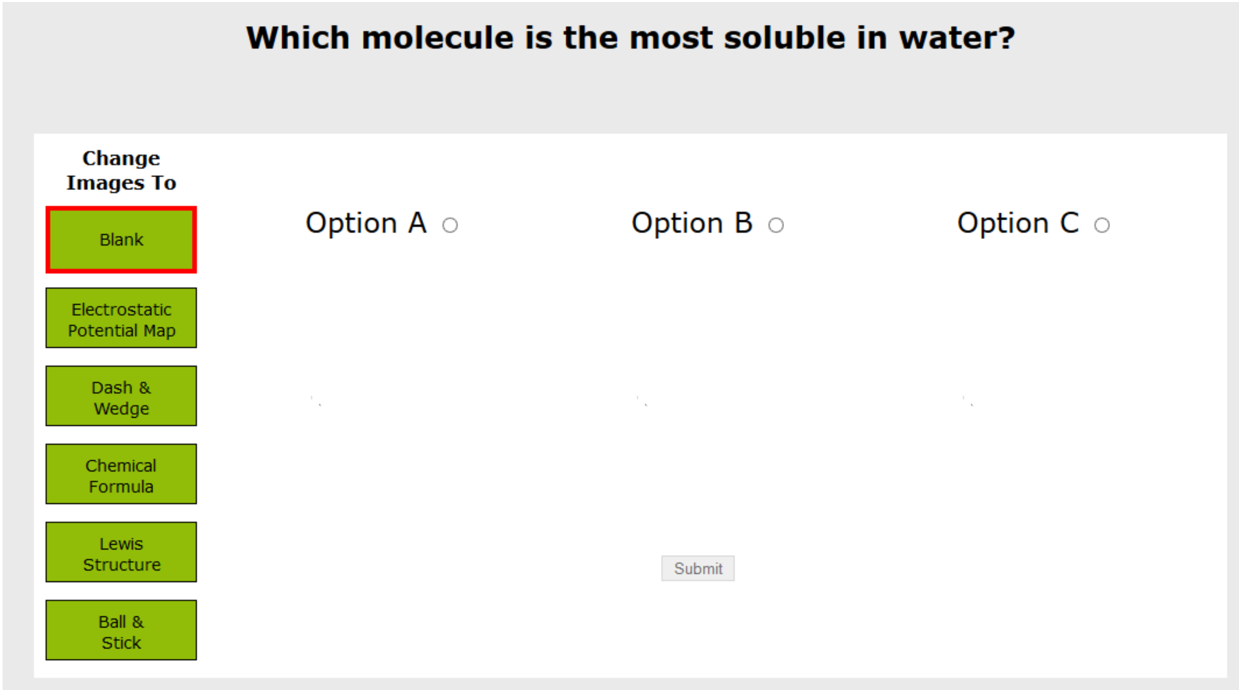

H2O
NH3
CH4

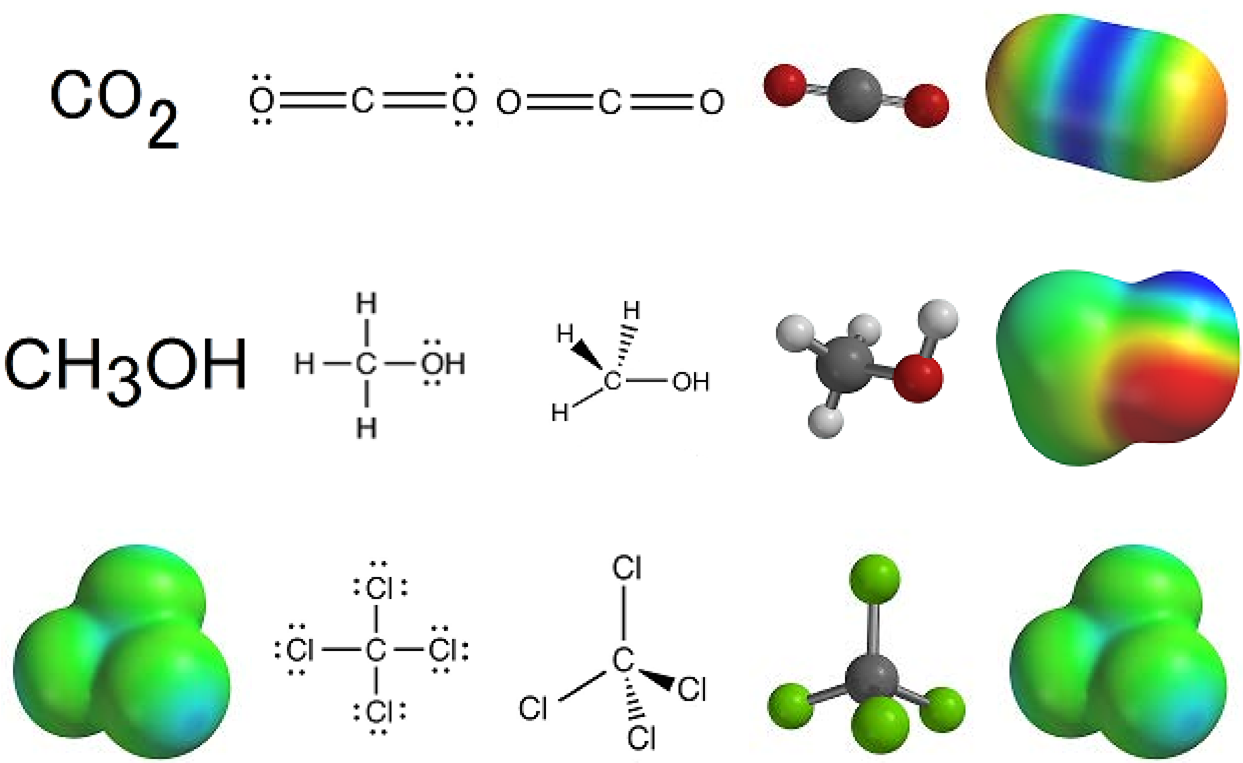
Question 5.)


Which molecule is organic by definition?
Change Images To
Blank
Electrostatic Potential Map
Dash & Wedge
Chemical Formula
Lewis Structure
Ball & Stick
Option A
Option B
Option C
Submit

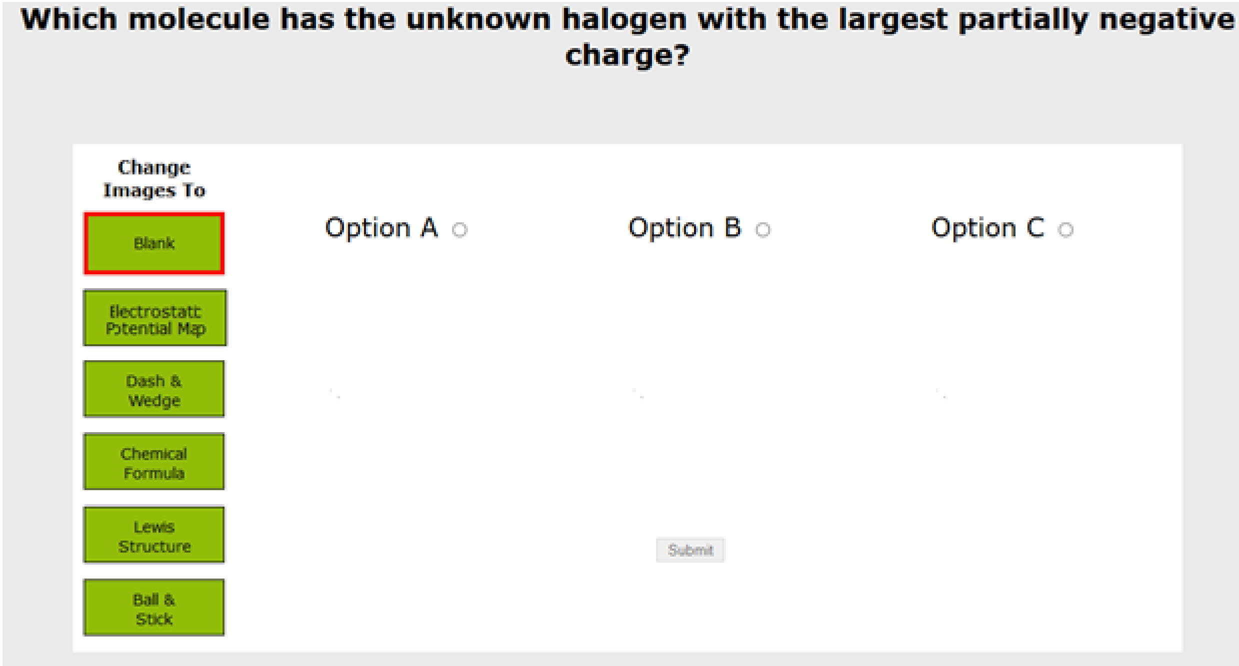

H2O
NH3
CH3OH

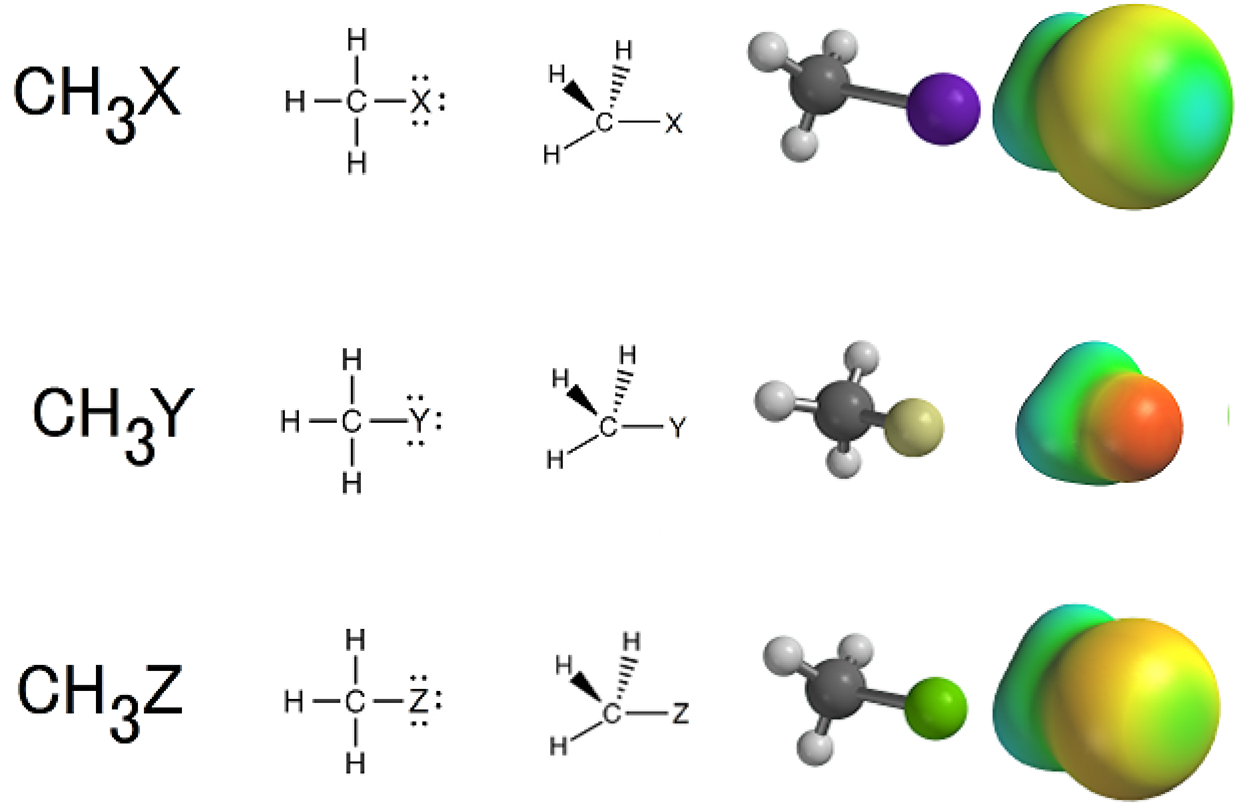
Question 7.)


Which molecule is the most soluble in water?
Change Images To
Blank
Electrostatic Potential Map
Dash & Wedge
Chemical Formula
Lewis Structure
Ball & Stick
Option A
Option B
Option C
Submit

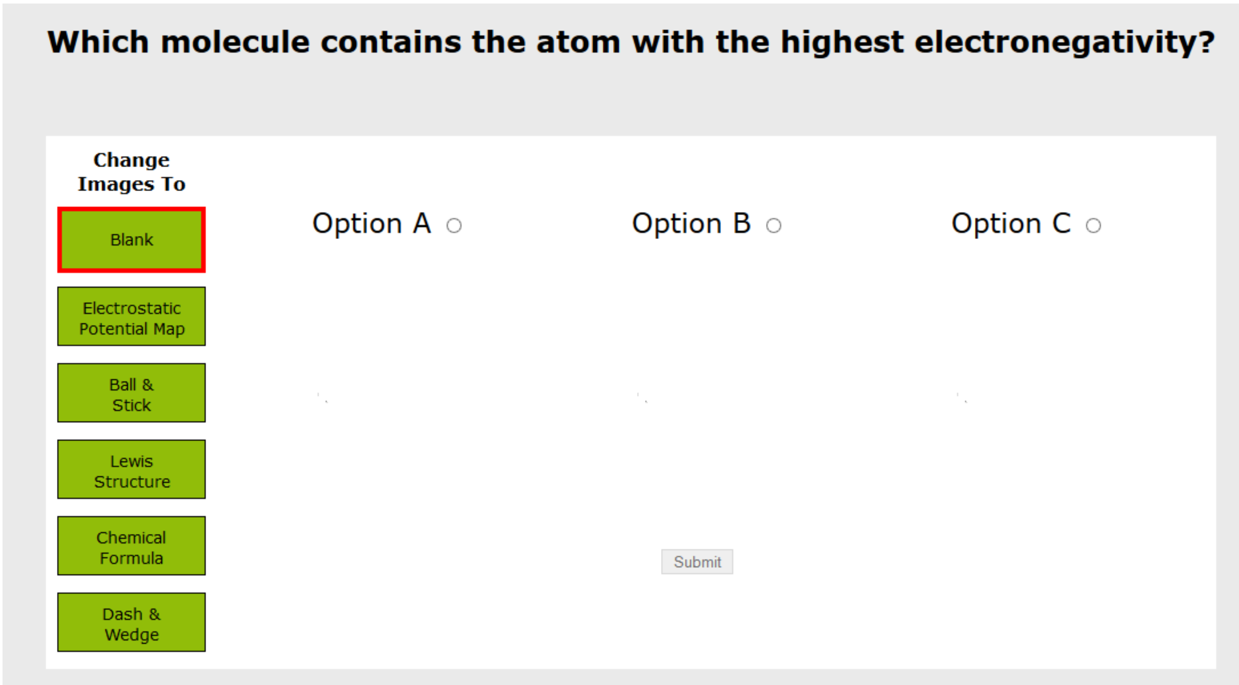

CO2
CH3OH

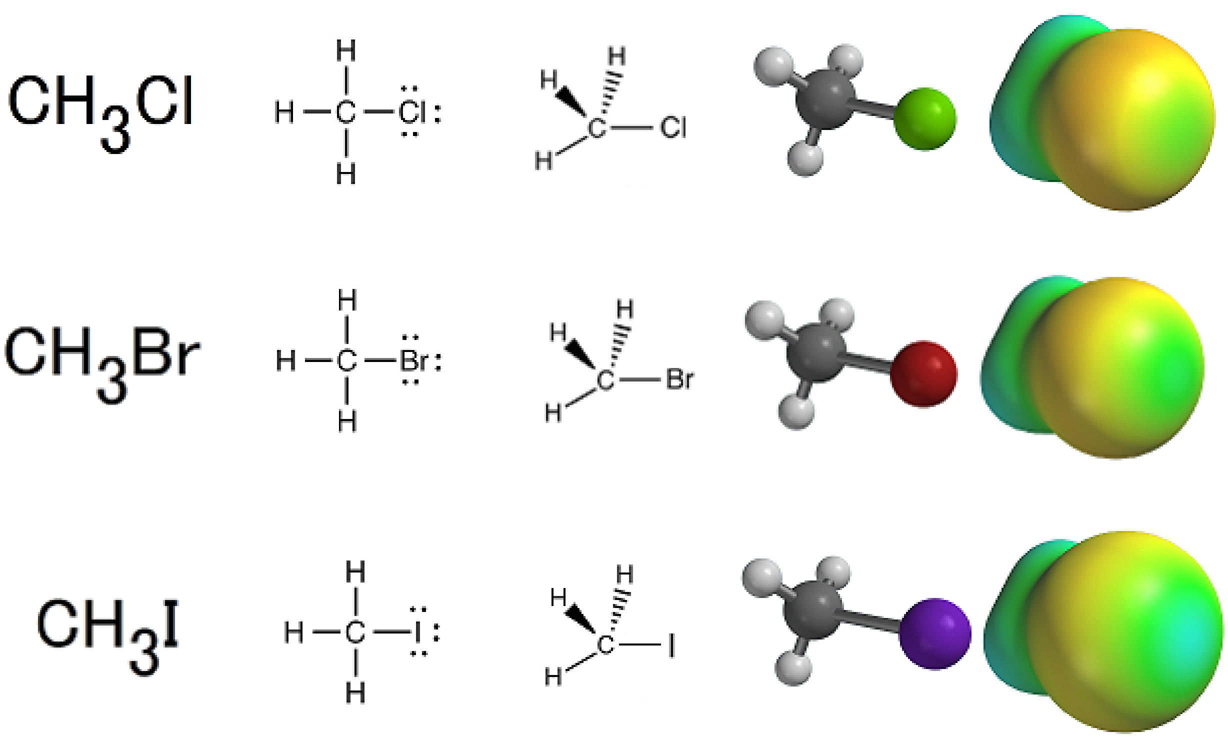
Question 9.)


Which molecule has the unknown halogen with the largest partially negative charge?
Change Images To
Blank
Electrostatic Potential Map
Dash & Wedge
Chemical Formula
Lewis Structure
Ball & Stick
Option A
Option B
Option C
Submit

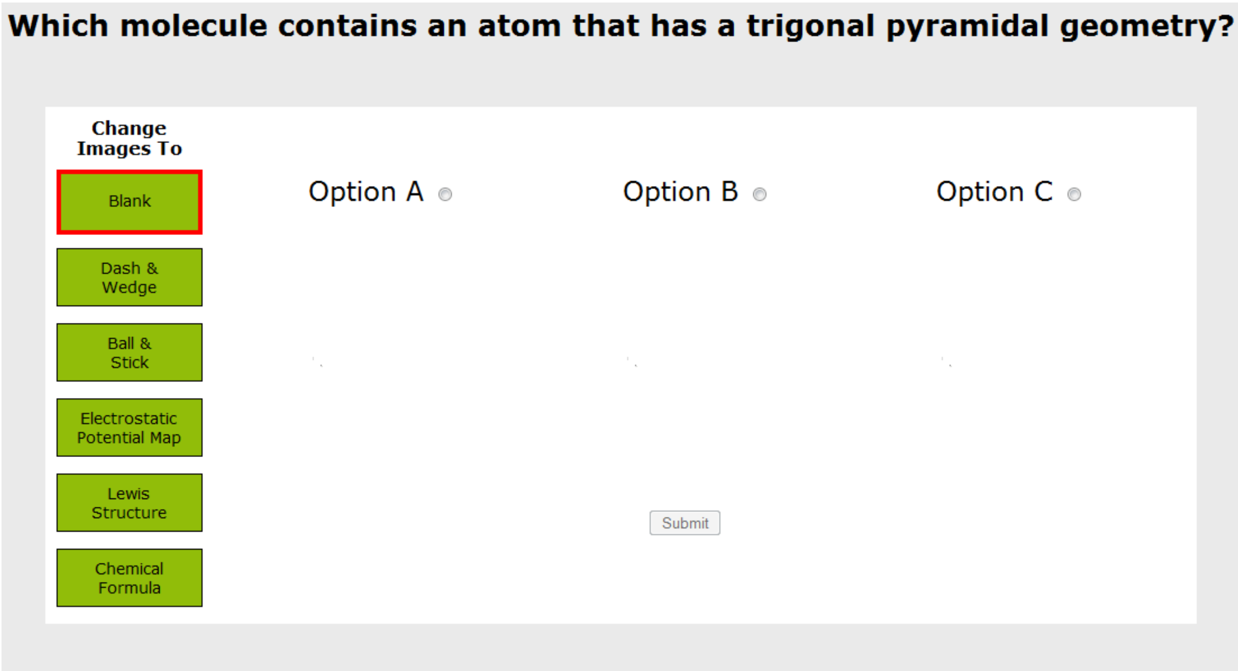

CH3X
CH3Y
CH3Z

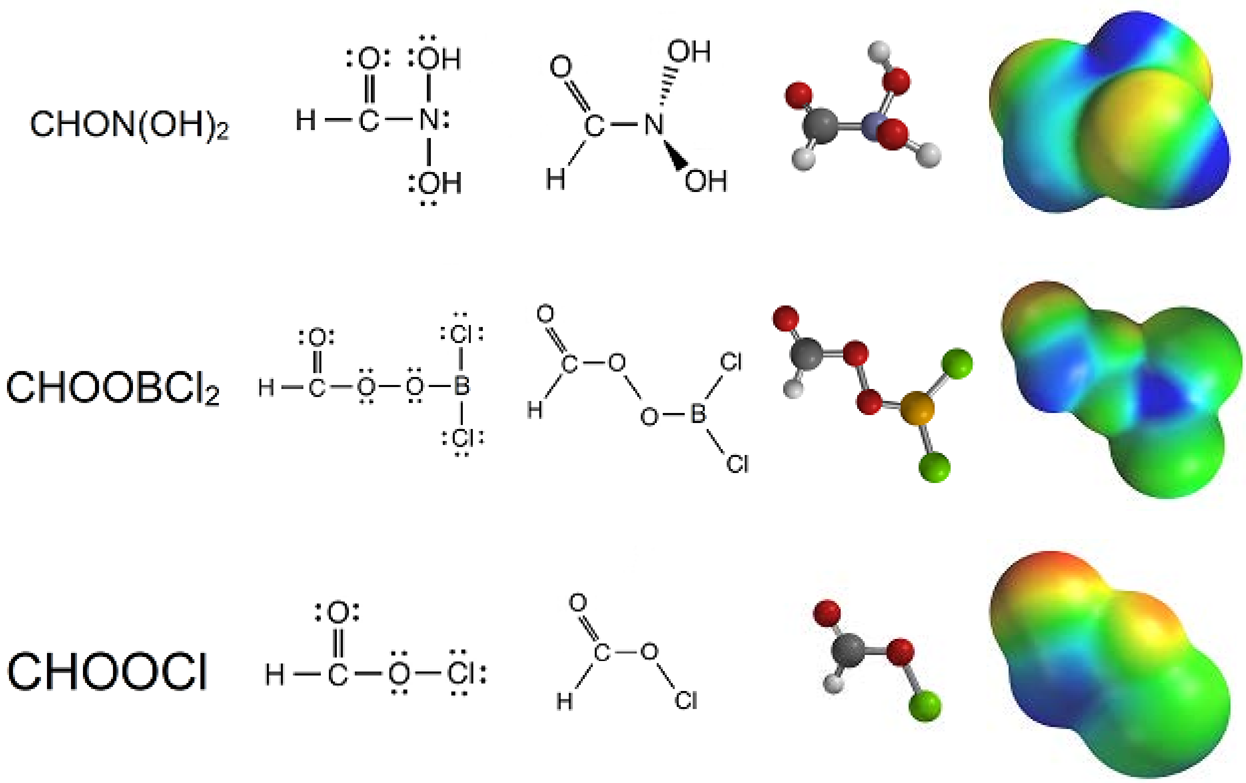
Question 10.)


Which molecule contains the atom with the highest electronegativity?
Change Images To
Blank
Electrostatic Potential Map
Ball & Stick
Lewis Structure
Chemical Formula
Dash & Wedge
Option A
Option B
Option C
Submit

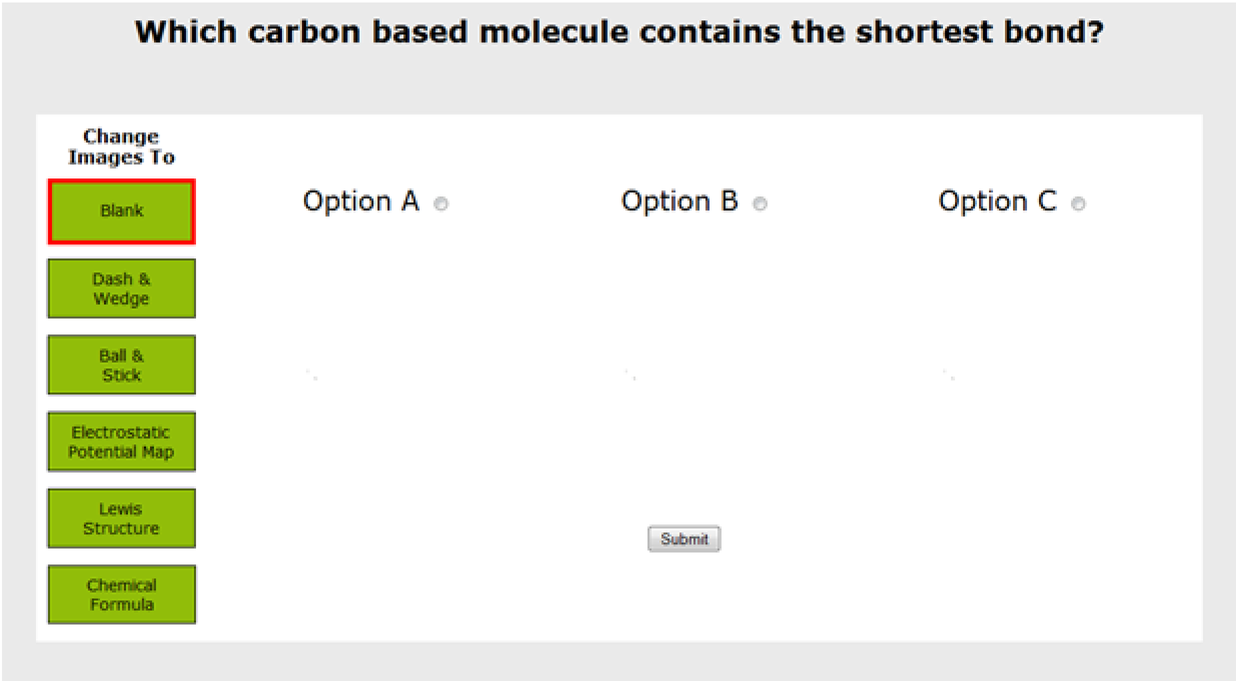

CH3Cl
CH3Br
CH3I

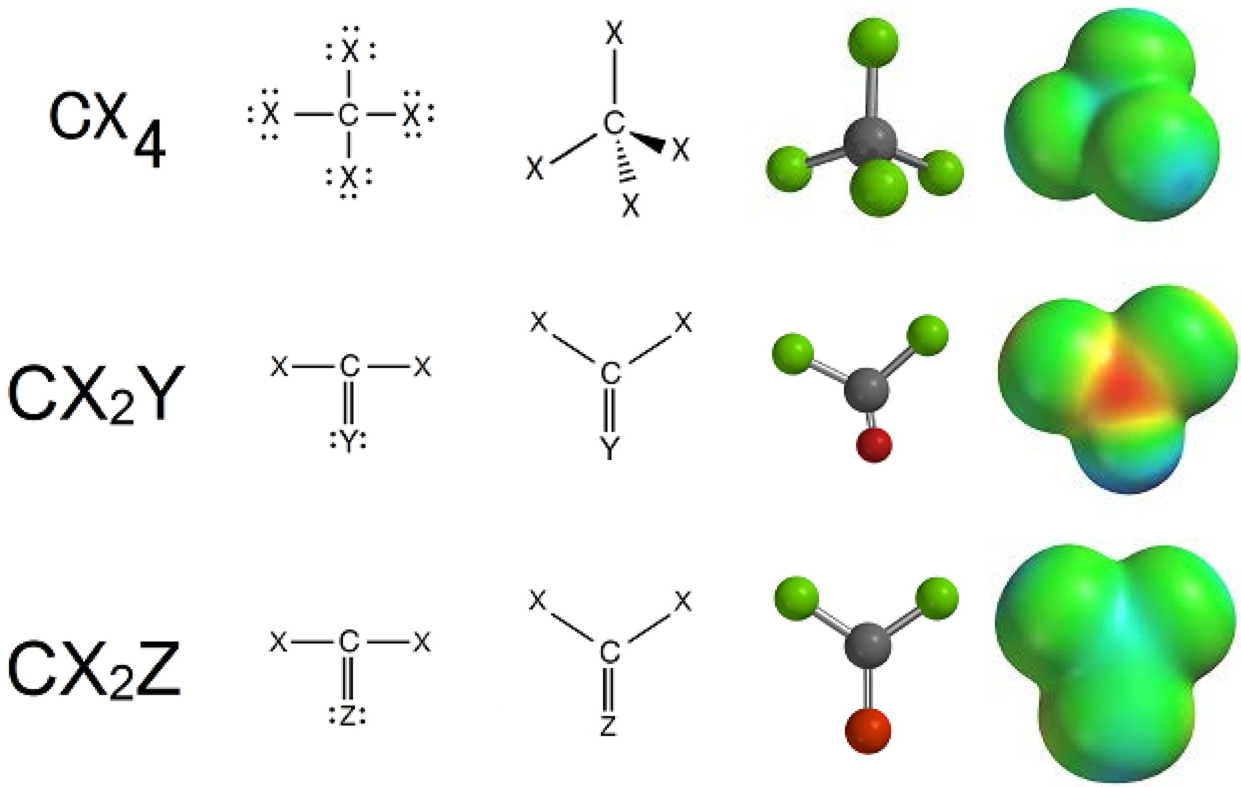
Question 13.)


Which molecule contains an atom that has a trigonal pyramidal geometry?
Change Images To
Blank
Dash & Wedge
Ball & Stick
Electrostatic Potential Map
Lewis Structure
Chemical Formula
Option A
Option B
Option C
Submit

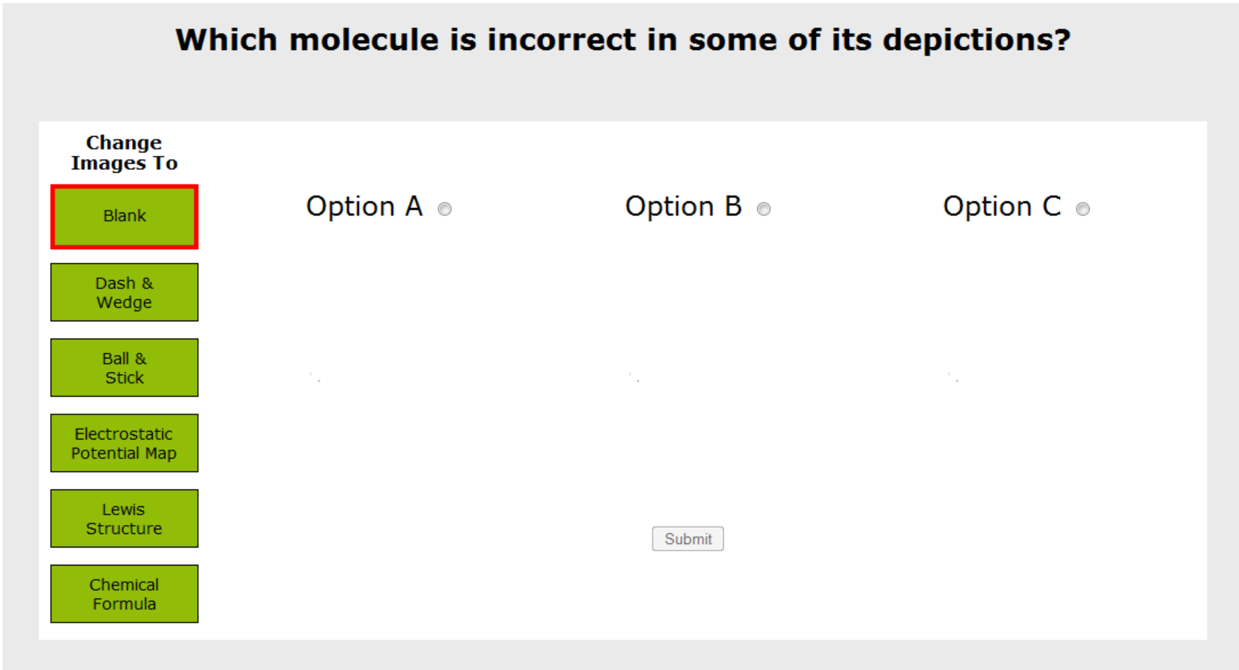

CHON(OH)2
CHOOBCl2
CHOOCl

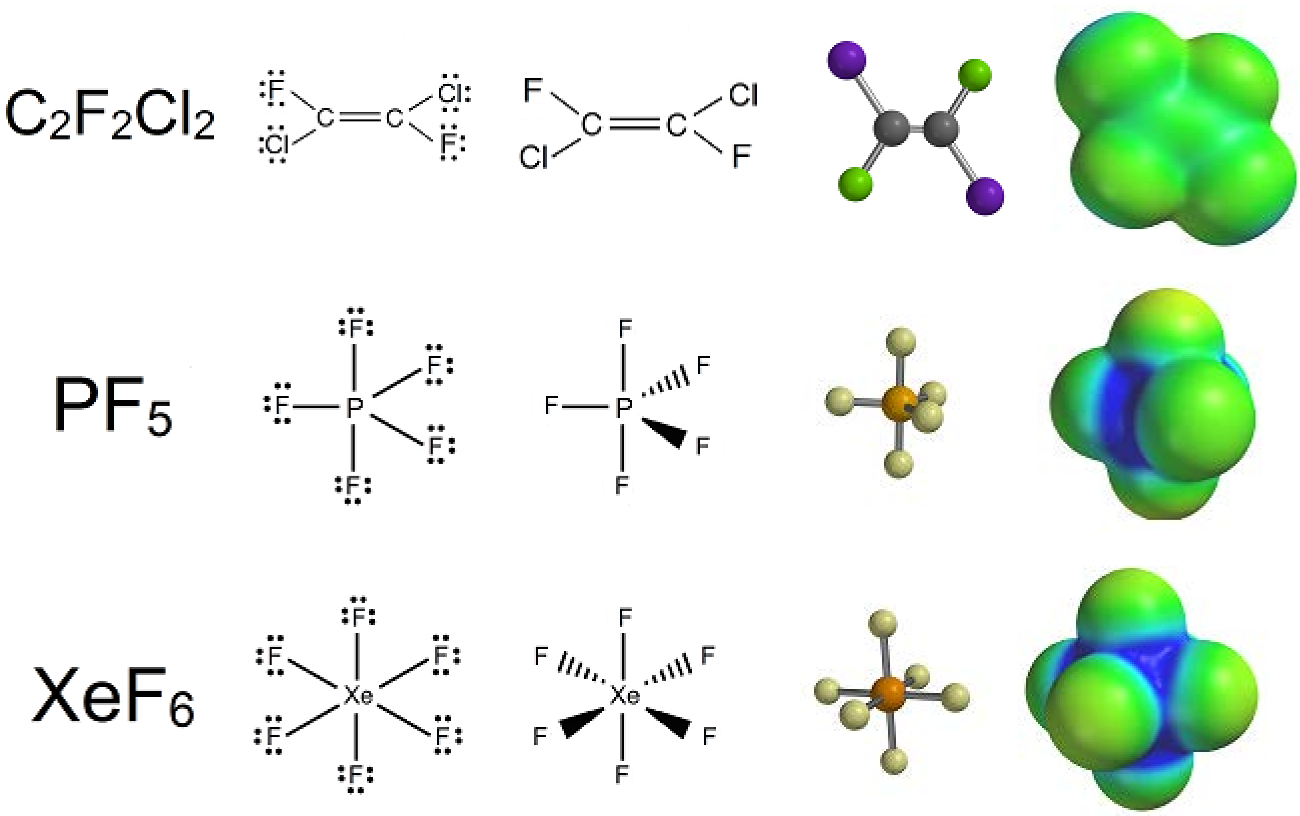
Question 14.)


Which carbon based molecule contains the shortest bond?
Change Images To
Blank
Dash & Wedge
Ball & Stick
Electrostatic Potential Map
Lewis Structure
Chemical Formula
Option A
Option B
Option C
Submit

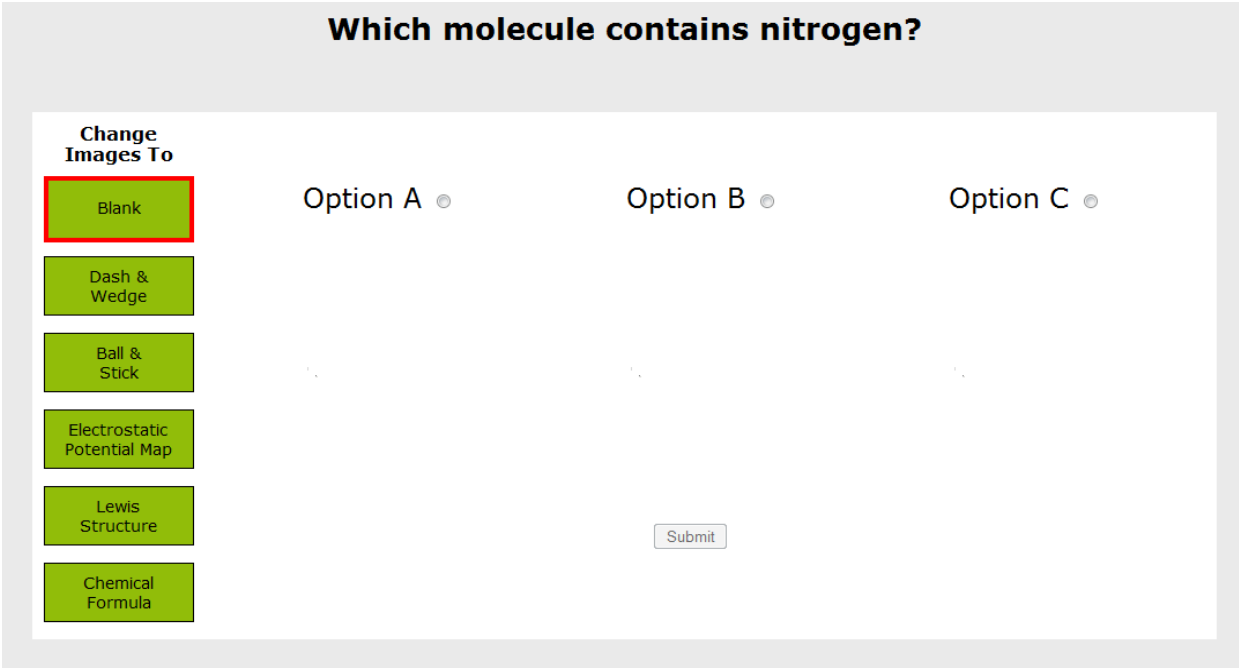

CX4
CX2Y
CX2Z

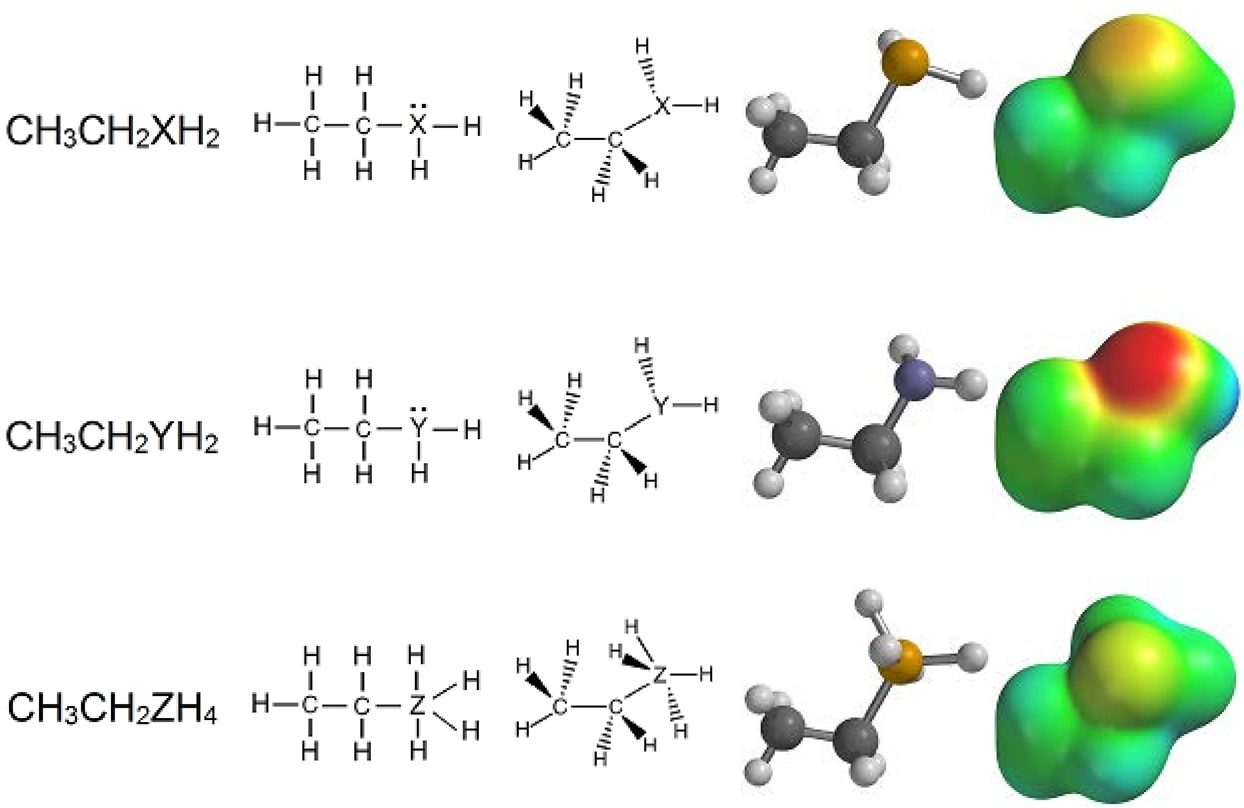
Question 15.)


Which molecule is incorrect in some of its depictions?
Change Images To
Blank
Dash & Wedge
Ball & Stick
Electrostatic Potential Map
Lewis Structure
Chemical Formula
Option A
Option B
Option C
Submit


C2F2Cl2
PF5
XeF6

Question 16.)


Which molecule contains nitrogen?
Change Images To
Blank
Dash & Wedge
Ball & Stick
Electrostatic Potential Map
Lewis Structure
Chemical Formula
Option A
Option B
Option C
Submit


CH3CH2XH2
CH3CH2YH2
CH3CH2ZH4

APPENDIX S

UNIVERSITY A'S PRE-SURVEY RESULTS

Number of University A Students from Based on Whether They Had Previously Used Each

Representation	Yes	No
Chemical formula	188	3
Lewis structure	181	10
Dash-and-wedge model	141	50
Ball-and-stick model	151	40
Electrostatic potential map	59	132

Number of University A Students Based on When They had First Used Each Representation

Representation	Middle School	High School	College	Other
Chemical formula	44	126	15	3
Lewis structure	16	126	39	0
Dash-and-wedge model	4	49	87	0
Ball-and-stick model	19	98	32	2
Electrostatic potential map	3	11	45	0

Frequency of Use of Each Representation Among Study University A Participants

Representation	Never	Rarely	Occasionally	Frequently
Chemical formula	4	32	80	72
Lewis structure	8	35	91	47
Dash-and-wedge model	4	61	49	27
Ball-and-stick model	11	66	61	13
Electrostatic potential map	5	35	14	5

APPENDIX T

UNIVERSITY B'S PRE-SURVEY RESULTS

Number of Students from University B Based on Whether They Had Previously Used Each

Representation	Yes	No
Chemical formula	113	3
Lewis structure	111	5
Dash-and-wedge model	38	78
Ball-and-stick model	94	22
Electrostatic potential map	17	99

Number of University B Students Based on When They had First Used Each Representation

Representation	Middle School	High School	College	Other
Chemical formula	22	67	20	4
Lewis structure	13	74	24	0
Dash-and-wedge model	0	11	27	0
Ball-and-stick model	15	50	26	3
Electrostatic potential map	0	6	11	0

Frequency of Use of Each Representation Among Study University B Participants

Representation	Never	Rarely	Occasionally	Frequently
Chemical formula	1	16	40	56
Lewis structure	4	35	59	13
Dash-and-wedge model	3	19	13	1
Ball-and-stick model	5	51	33	5
Electrostatic potential map	3	11	2	1

APPENDIX U

EMAIL FOR POSSIBLE FULL EXPERIMENT PARTICIPANTS

Hi, my name is Jack Polifka, I am a Masters student in Human Computer Interaction, and work in the Department of Chemistry at Iowa State University. You're being sent this email to inform you about an extra credit opportunity for your {Course name at University} class. For my research study, I'm working with Dr. Thomas Holme and we developed an online assessment for identifying how students use different visual representations in chemistry. The assessment will take approximately 10 minutes of your time. You will receive 5 points of extra credit for participating. Please try to answer the questions to the best of your ability. You will have until {Date of last opportunity to use the VRA} to use the online assessment.

To access the online assessment, use the following:

Username: {Username to access VRA}
URL: {URL to access VRA}

Please use a non-mobile device to use the online assessment.

On the consent page, make sure to fill in the line about your email to receive extra credit if you desire.

Thank you,
Jack Polifka

APPENDIX V

MEAN, STANDARD DEVIATION, AND STATISTICAL ANALYSIS OF USED TIME IN SECONDS FOR ALL REPRESENTATIONS FOR EACH QUESTION – UNIVERSITY A PARTICIPANTS WHO ANSWERED CORRECTLY

Question	CF [a]	LS [b]	D&W [c]	B&S [d]	EPM [e]	Total Time	Friedman Test *p*-value	Pairwise Wilcoxon Test *p*-value [f]
1	0.9 (2.2)	**5.6 (16.4)**	**2.8 (7.2)**	2.5 (17.5)	1.3 (4.9)	16.9 (28.3)	< .0001 [g]	0.0072 [g]
2	**4.7 (8.7)**	**7.1 (11.1)**	3.4 (5.8)	2.8 (10.6)	1.7 (3.3)	25.0 (17.9)	< .0001 [g]	0.1949
5	**5.5 (10.8)**	**4.1 (15.1)**	1.7 (3.6)	1.0 (2.2)	0.8 (1.9)	17.1 (20.0)	< .0001 [g]	0.0901
8 [h]	**8.7 (15.8)**	**4.5 (6.9)**	3.2 (4.6)	2.0 (5.1)	1.8 (4.1)	24.3 (21.1)	< .0001 [g]	0.3031
9	3.9 (5.9)	**8.0 (12.9)**	5.1 (8.9)	4.5 (10.1)	**8.3 (15.1)**	39.1 (32.4)	0.0006	> .9999
10	**6.6 (12.1)**	**12.2 (18.9)**	3.7 (14.0)	1.6 (5.1)	3.0 (7.6)	29.8 (24.4)	< .0001 [g]	> .9999
13	3.1 (6.9)	**17.4 (26.8)**	**11.4 (13.7)**	8.5 (15.1)	2.3 (5.2)	49.8 (40.9)	< .0001 [g]	> .9999
14	2.1 (2.9)	**10.1 (9.4)**	4.3 (4.4)	**4.4 (4.2)**	3.0 (6.2)	28.6 (15.4)	< .0001 [g]	< .0001 [g]
15 [h]	5.7 (9.1)	**12.3 (18.6)**	**10.1 (17.1)**	8.2 (19.9)	4.4 (8.7)	46.3 (56.9)	< .0001 [g]	> .9999
16	**9.4 (18.8)**	**15.5 (24.7)**	5.3 (7.6)	5.6 (6.4)	4.5 (7.3)	44.9 (38.6)	< .0001 [g]	0.0869

[a] CF: chemical formula, [b] LS: Lewis structure, [c] D&W: dash-and-wedge, [d] B&S: ball-and-stick, [e] EPM: electrostatic potential map,
[f] With Bonferroni correction between most and second most used viewed representation, [g] Significant *p*-value, [h] Removed outliner(s)
Most Used Representation, **Second Most Used Representation**

APPENDIX W

MEAN, STANDARD DEVIATION, AND STATISTICAL ANALYSIS OF USED TIME IN SECONDS FOR ALL REPRESENTATIONS FOR EACH QUESTION – UNIVERSITY A PARTICIPANTS WHO ANSWERED INCORRECTLY

Question	CF [a]	LS [b]	D&W [c]	B&S [d]	EPM [e]	Total Time	Friedman Test *p*-value	Pairwise Wilcoxon Test *p*-value [f]
1	1.6 (2.6)	1.9 (3.1)	**2.8 (4.2)**	**2.3 (3.1)**	1.4 (3.0)	14.8 (10.3)	0.0589	> .9999
2	4.8 (11.6)	**8.8 (16.0)**	**5.6 (8.6)**	1.8 (5.2)	1.2 (2.6)	27.0 (22.3)	< .0001 [g]	> .9999
5	**6.2 (10.0)**	**3.3 (11.7)**	2.9 (5.6)	1.5 (4.8)	0.4 (0.9)	19.5 (16.1)	< .0001 [g]	0.1164
8 [h]	**9.0 (15.5)**	**4.8 (9.1)**	2.3 (4.7)	2.6 (10.0)	1.5 (2.2)	28.0 (33.1)	< .0001 [g]	0.0188 [g]
9	4.6 (7.7)	**7.4 (11.3)**	3.6 (6.4)	3.1 (3.7)	**5.2 (7.3)**	32.2 (20.1)	0.0374 [g]	> .9999
10	**4.9 (16.8)**	3.5 (6.7)	1.6 (2.8)	1.3 (2.1)	**4.3 (8.2)**	21.3 (23.6)	< .0001 [g]	> .9999
13	4.9 (15.2)	7.7 (10.3)	**12.4 (17.8)**	**12.7 (23.5)**	2.6 (7.3)	45.4 (38.4)	< .0001 [g]	> .9999
14	1.8 (3.1)	**10.1 (10.0)**	**2.9 (4.5)**	1.4 (2.4)	1.7 (3.9)	22.4 (14.5)	< .0001 [g]	< .0001 [g]
15 [h]	**5.9 (12.4)**	**23.6 (40.9)**	5.7 (9.4)	3.3 (5.4)	2.7 (5.1)	46.3 (52.6)	< .0001 [g]	< .0001 [g]
16	**10.3 (15.6)**	**15.9 (21.9)**	9.4 (17.0)	2.9 (5.5)	1.7 (4.8)	44.3 (36.8)	< .0001 [g]	0.2410

[a] CF: chemical formula, [b] LS: Lewis structure, [c] D&W: dash-and-wedge, [d] B&S: ball-and-stick, [e] EPM: electrostatic potential map,
[f] With Bonferroni correction between most and second most used viewed representation, [g] Significant *p*-value, [h] Removed outliner(s)
Most Used Representation, **Second Most Used Representation**

APPENDIX X

MEAN, STANDARD DEVIATION, AND STATISTICAL ANALYSIS OF USED TIME IN SECONDS FOR ALL REPRESENTATIONS FOR EACH QUESTION – UNIVERSITY B PARTICIPANTS WHO ANSWERED CORRECTLY

Question	CF [a]	LS [b]	D&W [c]	B&S [d]	EPM [e]	Total Time	Friedman Test *p*-value	Pairwise Wilcoxon Test *p*-value [f]
1	1.3 (3.3)	**4.7 (11.1)**	**2.4 (5.5)**	1.5 (2.8)	1.8 (3.4)	16.8 (18.0)	0.0006 [g]	0.4539
2	**7.0 (17.8)**	**6.7 (10.2)**	2.9 (6.9)	2.3 (4.7)	1.3 (2.5)	26.0 (22.7)	0.0010 [g]	> .9999
5	**3.1 (4.9)**	**3.3 (8.2)**	1.9 (4.5)	1.2 (2.4)	0.7 (1.0)	14.3 (11.6)	< .0001 [g]	> .9999
8 [h]	**7.5 (12.7)**	**7.6 (14.6)**	3.1 (4.9)	1.7 (2.7)	2.0 (4.1)	26.3 (20.3)	< .0001 [g]	> .9999
9	2.7 (3.1)	**3.8 (4.6)**	3.6 (8.9)	3.0 (6.1)	**5.8 (5.2)**	28.1 (21.1)	< .0001 [g]	> .9999
10 [h]	**7.7 (14.7)**	**15.2 (25.5)**	2.2 (4.1)	1.5 (3.1)	3.0 (6.4)	36.2 (31.2)	< .0001 [g]	> .9999
13	2.8 (5.4)	4.6 (7.0)	**5.9 (7.2)**	**6.1 (6.1)**	1.9 (3.8)	27.1 (18.2)	< .0001 [g]	> .9999
14	1.9 (2.2)	**6.1 (7.7)**	**2.6 (3.4)**	4.9 (4.4)	3.3 (4.2)	27.9 (23.4)	< .0001 [g]	> .9999
15	**7.6 (8.0)**	**12.4 (18.7)**	4.3 (5.7)	4.1 (5.7)	4.3 (7.0)	37.3 (32.4)	0.0155 [g]	> .9999
16	**8.2 (9.7)**	**21.8 (30.1)**	4.7 (4.8)	5.0 (5.4)	4.1 (4.7)	47.4 (35.1)	0.0006 [g]	0.4539

[a] CF: chemical formula, [b] LS: Lewis structure, [c] D&W: dash-and-wedge, [d] B&S: ball-and-stick, [e] EPM: electrostatic potential map,
[f] With Bonferroni correction between most and second most used viewed representation, [g] Significant *p*-value, [h] Removed outliner
Most Used Representation, **Second Most Used Representation**

APPENDIX Y

MEAN, STANDARD DEVIATION, AND STATISTICAL ANALYSIS OF USED TIME IN SECONDS FOR ALL REPRESENTATIONS FOR EACH QUESTION – UNIVERSITY B PARTICIPANTS WHO ANSWERED INCORRECTLY

Question	CF [a]	LS [b]	D&W [c]	B&S [d]	EPM [e]	Total Time	Friedman Test *p*-value	Pairwise Wilcoxon Test *p*-value [f]
1	**4.9 (10.2)**	**5.5 (13.4)**	3.5 (6.9)	1.6 (2.8)	1.7 (3.6)	20.1 (22.2)	0.1352	> .9999
2	4.2 (8.2)	**5.5 (8.2)**	**5.6 (18.0)**	1.8 (5.0)	0.7 (1.6)	22.5 (24.4)	< .0001 [g]	0.2376
5	**4.4 (8.7)**	**4.5 (13.7)**	1.0 (1.7)	0.8 (2.0)	0.7 (1.1)	16.9 (23.8)	0.0039 [g]	> .9999
8 [h]	**5.2 (7.1)**	**5.0 (7.6)**	2.8 (6.6)	2.9 (9.4)	1.5 (2.9)	21.3 (17.3)	< .0001 [g]	> .9999
9	4.3 (5.3)	**7.7 (8.8)**	3.8 (6.1)	3.4 (4.9)	**6.4 (12.8)**	32.4 (24.5)	0.0136 [g]	> .9999
10 [h]	2.9 (3.8)	**3.1 (6.0)**	1.2 (1.9)	1.3 (2.1)	**4.0 (4.1)**	16.0 (10.2)	< .0001 [g]	0.0412 [g]
13	3.2 (11.8)	**8.5 (15.8)**	**7.3 (16.1)**	7.3 (8.7)	2.7 (8.7)	42.9 (65.2)	< .0001 [g]	> .9999
14	2.7 (3.6)	**6.8 (8.3)**	**3.5 (5.1)**	3.1 (5.0)	1.8 (3.3)	22.3 (15.3)	< .0001 [g]	0.0931
15	3.8 (5.6)	**21.0 (30.7)**	**7.7 (16.0)**	3.5 (7.6)	2.5 (5.6)	43.2 (38.6)	< .0001 [g]	0.0018 [g]
16	**8.4 (8.2)**	**20.5 (36.3)**	5.0 (8.8)	2.9 (4.4)	1.9 (3.2)	43.0 (40.0)	< .0001 [g]	0.0320 [g]

[a] CF: chemical formula, [b] LS: Lewis structure, [c] D&W: dash-and-wedge, [d] B&S: ball-and-stick, [e] EPM: electrostatic potential map,
[f] With Bonferroni correction between most and second most used viewed representation, [g] Significant *p*-value, [h] Removed outliner
Most Used Representation, **Second Most Used Representation**