A COGNITIVE LOAD APPROACH TO LEARNER-CENTERED DESIGN OF DIGITAL INSTRUCTIONAL MEDIA AND SUPPORTING ACCESSIBILITY TOOLS

A thesis

submitted by

Ashley Russell

In partial fulfillment of the requirements

for the degree of

Master of Science

in

Human Factors Engineering

TUFTS UNIVERSITY

August 2011

Advisor: Daniel J. Hannon

Abstract

The development of computer supports to scaffold learning for students with cognitive and emotional disabilities has been largely overlooked in the rise of the digital age. However, with recent legislative changes, demands on many of these students have increased, while instructional materials remain limited to static media (Graham, Harris, MacArthur & Schwartz, 1991; Gribbons, 2008). While certain strides have been made towards providing support for the nearly 23% of the population who struggle with illiteracy or learning impairments primarily through use of the internet, there has been little initiative in the field of human factors (Gibbons, 2008). Consequently, existing design guidelines remain vague and methodology for including universal accessibility in the product development lifecycle for educational technology is limited, at best. Furthermore, the cognitive demands placed on these students are often unmanageable and consequently lead to deteriorating performance.

This thesis argues for a human factors approach to the design, evaluation, and implementation of technology in the classroom. This work includes integrating features that promote universal accessibility in a context that considers both content complexity and information presentation for the greatest utilization of these features. It further suggests considerations of cognitive load theory be applied to curriculum development in order to promote performance gains for varied learners while pursuing active tasks of information acquisition, processing, and output.

Acknowledgements

Firstly, my many thanks to Daniel Hannon. I hugely benefited from Dan's immense expertise in the field of Human Factors and from him showing me how to "trust the process." For his tireless work as my teacher, advisor, and tactful mentor, I am tremendously grateful.

I would also like to thank Ethan Danahy for his daily encouragements, imagination, humor, and eloquent workmanship; I am so glad for the opportunity to have worked with Ethan. My great appreciation also to Chris Rogers for adopting me into the Center and for his amazing energy, support, and vision; I am so pleased to have been able to work on something so dear to my heart in such a wonderful environment.

A big "thank you" to my family for their constant love and encouragement; to my lovely friends who have lent me their ears and thoughts so many times; and, to my fellow grad students James, Maureen, Elsa, and Cheryl, without whose feedback, resourcefulness, and aid, I could have never accomplished this work. And, most importantly, my infinite gratitude to Shashwat Yadav, whose persistence and endless encouragement prompted me to embark on this adventure.

Table of Contents

A COGNITIVE LOAD APPROACH TO LEARNER-CENTERED DESIGN OF DIGITAL INSTRUCTIONAL MEDIA AND SUPPORTING ACCESSIBILITY TOOLS

1. Introduction

In the digital age, technology is "increasingly a means for empowering students, a method for communication and socializing, and a ubiquitous, transparent part of their lives" (Johnson, Smith, Levine, & Haywood, 2010). The recent changes in technology, the adoption of technology into the daily lives of students, and lagging educational systems contribute to an overarching need for computer assimilation in the classroom. Additionally, recent changes in legislation promote an integrated student body, one that includes students with disabilities, which further endorse the need for tools that allow all learners access to information at their own pace and subsequently provide a successful learning environment (Rose, Meyer, & Hitchcock, 2005). Educational technology could be such a tool that embodies the current technical climate and accommodates varied learner needs. However, considerations must be made in the design, evaluation, and implementation of educational technology such that as a tool it is usable, accessible, appropriate, and promotes gains for all students as well as the larger educational system.

Within the realm of educational technology lives a multiplicity of requirements that extend beyond a traditional user interface interaction. Such requirements include an assessment of the student and teacher needs, pedagogical foundation, selection of technology, and additional evaluation for learning outcomes. Furthermore, in order to accommodate varied learner types, the technology needs to be adequately accessible, promoting equitable learning opportunities.

Universal design for learning (UDL) is a set of principles that address inequitable learning opportunities for educational technology. UDL theory, developed by the Center for Applied Special Technology (CAST), suggests that providing multiple means of representation, expression, and engagement allows greatest student access to learning opportunities (CAST, 2011). Cognitive tools, which embody these values, can be provided in a digital environment to increase opportunities for students who may struggle with static informational materials. However, while focus has been directed towards UDL tool development, and considerable gains have been accomplished with the establishment of UDL guidelines, the influence of the instructional design on the utilization of the tools remains largely unexplored.

With regards to the design of instructional material, cognitive load theory suggests that several cognitive mechanisms are involved in learning and understanding instructional material (Sweller, 1994). These cognitive processes allow information to move through the working memory, where current mental activity occurs, into the long-term memory, where information that has been successfully processed is stored (Carlson, Chandler, & Sweller, 2003). In cognitive load theory, cognitive load is the demand placed on the working memory by instructional materials; as cognitive load increases, it is more challenging to manage information processing (Marcus, Cooper, & Sweller, 1996). In cognitive load theory there are three types of cognitive load (intrinsic, extraneous, and germane), which must be managed in order for information to successfully pass from the limited short term memory into the relatively unlimited long term memory and, consequently, for learning to occur. This successful passage of information, from the short term memory to the long term memory supports the retention of learning.

The function of both the working and long term memory systems each have implications for the learning disabled population. The working memory's limited capacity imposes a bottleneck for information processing and human computer interaction. The "dynamic interaction between memory capacity and the time information can be actively maintained also imposes severe constraints on the learning disabled" (Gribbons, 2008). Though it is unclear whether this population has a diminished working memory capacity or whether the underlying cognitive deficiencies place undue burden on otherwise "normal" capacity, there remains the need for design accommodations.

In this thesis, I argue for the integration of cognitive load theory and universal accessibility within the context of human computer interaction. In order to do so, this thesis will commence with a survey of pertinent literature reviewing the current state of educational technology, universal accessibility in system design, and the role of cognitive load theory for a human factors approach to educational technology system design. Subsequent sections will detail a tripartite research focus, which began in the field with an exploratory evaluation of UDL tools, followed by laboratory research exploring cognitive load theory with respect to UDL tools, and lastly returned again to the field evaluating cognitive load theory in practice, again within the classroom environment. Conducting the research in these three successive phases (field observation, lab investigation, and field application) provided a comprehensive, holistic approach to understanding the impact UDL tools in the framework of cognitive load theory.

2. Survey of Literature

2.1 Accessible Educational Technology

2.1.1 Educational technology.

There are different skills, motivations, and objectives for learners using software applications, such as the development of content knowledge in congruence with technological skills and metacognitive strategies (Wallace, Soloway, Krajcik, Bos, Hoffman et al., 1998). These distinguishable user needs require different approaches to designing technology that support users to achieve learning goals. A well-designed software environment can enable students to learn by doing, receive immediate feedback, continually refine understanding, and create new knowledge structures (Phillips, 2005). This role for technology in education extends beyond typical software user needs of adoption ease, performance, and low error rates and must also seamlessly support learner needs of aiding the learning process through acquiring and processing content matter and through expressing knowledge. Thus, in order to aid in the process of learning, considerations must be made that extend beyond traditional human computer interaction concepts.

Van Dam, Becker, and Simpson (2005) further define the role of educational technology to “support adaptability and flexibility, to enable appropriate modes of learning, and to foster interactive involvement of the learner with the educational materials.” Notably, technology affords a means to support these goals that may otherwise be limited by traditional, static media, yet must be appropriately constructed so that it truly satisfies these objectives. Van Dam et al. further note that there is no one size fits all solution for educational technology. Moreover, without accommodating for diverse learning styles, software media can be as limited as fixed media.

Extend the thought of accommodation for diverse learning styles to the current classroom, an environment that now includes an increasingly diverse range of learners and their associated needs. Students that struggle with a range of behavioral, cognitive, and learning impediments work alongside students with a variety of learning styles. This often translates into classrooms that need to accommodate a wide range of interactions between students and instructional material. For example, imagine instructional design that is flexible enough to accommodate a slower reader, who may need consistent feedback and frequent encouragement, as well as a self-motivated learner with advanced information processing and metacognitive skills. In designing educational technology for all, universal accessibility is a pertinent and conscientious design criterion for the development of successful educational technology that can support the needs of all students, regardless of learning style.

2.1.2 Universal Accessibility in the science classroom.

Despite recent legislation established to allow students with disabilities to participate in the classroom, there is still a prominent gap in the science performance of students with and without disabilities (National Assessment of Educational Progress, 2005). While expectations for students with disabilities have increased considerably, the instructional materials used in the science classroom generally remain limited to printed text and paper-and-pencil activities. Often, these static media pose barriers in learning for students with disabilities who often are confronted with a range of reading obstacles. Such fixed media creates barriers to students who might not only struggle with the content, but also struggle with reading, taking notes, organizing information, and analyzing and presenting data. Therefore, the cognitive demands imposed by these activities are often too great for students with disabilities, overburdening the capacity of the short term memory such that knowledge structures become increasing difficult to construct and hampering opportunities for learning to occur.

Additionally, the National Science Education Standards have shifted the emphasis from learning acquisition and teacher presentation to active student observation and interaction (National Science Education Standards, 1996). This only further exacerbates the barriers to Science, Technology, Engineering, and Math (STEM) learning for students with disabilities, as it requires higher-level scientific reasoning and critical thinking. Using higher-order thinking skills with specific content knowledge in problem-solving activities can be challenging for all students, but especially for those with learning disabilities. These students must not only manage the high cognitive load imposed by complex instructional content but must also process the information formats, organize the information, and apply higher order reasoning in order for learning to occur.

The cognitively disabled population has been largely overlooked in the field of system design and improvements for universal accessibility and design practices for software development are, for the most part, omitted from leading human computer interaction research (Gribbons, 2008). Accordingly, interface design methods for working with, and for, the cognitively disabled population are limited; design guidelines are often vague or directed towards specific populations with specific needs, such as the visually impaired. The needs of an integrated student body include accommodation of all learner types and consideration of a wide range in capacity for information acquisition, processing, and subsequent expression of understanding.

In this context, universal accessibility is an approach to meet systemic requirements for the broadest possible user population (Stephanidis, 2001). Here, the user population is derived from the integrated classroom, and embodies users with varying learning types, different levels of cognitive functioning, and a range of motivation, engagement, and attention. When designing for diversity in end user requirements, Stary (2001) proposes considerations for representation of user characteristics, contexts of use, and their mutual interconnectedness.

2.1.2.1 User Characteristics.

Functional user characteristics across learning disabilities can be categorized, interdependently, as: (1) reading (2) memory, (3) metacognitive, and (4) search and navigation (Bohman & Anderson, 2005). Gribbons (2008) proposes that each category encapsulates a broad list of deficiencies and many deficiencies are connected to others. Within the category of reading, for example, one draws on both memory and metacognitive skills. Likewise, search and navigation places load on working and long term memory and needs monitoring from metacognitive functioning. Each category and interdependency suggests particular user requirements.

User requirements, in turn, lead to particular design guidelines. For example, using visual and auditory prompts provides an alternative strategy for users who may be confronted with an initial reading barrier (Gribbons, 2008). Low information density is a complementary strategy for reading, as well as using illustrations to support text, emphasizing connections, and presenting content in sequence. Metacognitive support could be promoted by immediate communication of goals and through the provision of reminders and checklists to support self-monitoring. Effective navigation and search design practices include the use of labels, identified paths, site maps, and use of redundant cues. Guidelines to decrease mental workload, and consequently support memory needs, would include principles such as: maintaining consistency in the design, avoiding split attention with multiple tasks, and supporting mental calculations, decisions, and comparisons. Yet, such mental workload guidelines remain vague and means for attaining low cognitive load, unspecified.

Achieving the appropriate mental workload for instructional and interface design is particularly relevant when considering educational technology, as mental demands in integrated classrooms may often be too high for students with disabilities. The learning disabled are consequently allocating cognitive resources to reading, memory, and metacognitive demands that are often automated by a typically developing student (Gribbons, 2008). This population often is less likely to read due to overlapping difficulties with reading, consequently having less developed domain models; hence, instructional materials can be considerably more demanding. When additional instructional design demands are placed on these students, a consequent situation of mental overload could significantly inhibit the learning process. Presumably, in this context, designers of educational technology must not only be cognizant of the mental demands caused by the interface itself, but additionally by the instructional material.

2.1.2.2 Context of Use.

In this thesis, we will consider the primary context of use for educational technology, the classroom. This context brings considerations in addition to learner characteristics, which include pedagogical theory and impact of the computer on the learning process (Larsen, 1995). Larsen further explains that students with learning disabilities may learn at a slower pace, and consequently need tools to "facilitate, enhance and accelerate" mental processing.

This is congruent with common human computer interaction principles for learner-centered educational technology design. Soloway, Guzdial, & Hay (1994) claim that when designing educational software several additional considerations should be made to traditional user-centered design. These additional steps include a needs analysis for the learners and selection of pedagogy, such as behaviorism, constructivism, or social-constructivism. Upon selection of a pedagogical foundation, the media should match in affordances to the learning objectives. An extension of this principle could be that within the context of learning technology, accessibility tools should afford learning objectives. This raises an important methodological question: how can accessibility tools be designed to afford learning?

When considering the needs of this diverse user group in the context of educational technology, cognitive aids can be used to alleviate some of the mental demands placed on the user both due to the instructional content as well as the interface. These tools can be utilized to communicate objectives, sequences, and organizational structures; they can support self-monitoring, query the user when choices and decisions are required, or aid in calculations and comparisons (Gribbons, 2008). Yet, while principles exist to broadly address functional characteristics of the cognitively disabled, it is unclear on how these tools, when embedded into instructional design, can be best utilized. Consequently, the aforementioned question of how to design these tools to afford learning is a multidimensional problem that must consider the user characteristics, the context, the interface, the content, and the interactions among them.

2.1.2.3 Universal design for learning.

Universal design for learning (UDL) is a set of principles for designing curriculum that provides a flexible approach and promotes equal opportunities for learning for individuals with disabilities (CAST, 2011). The three guiding principles for UDL are to provide multiple ways of representing content, multiple methods of engagement, and multiple ways of expression (Rose & Meyer, 2002; Rose, Meyer, & Hitchcock, 2005). In this context, content presentation could be addressed by providing text in a digital media, which is accompanied by audio, additional graphics or videos, and cognitive aids such as highlighting, hints, and concept maps. Engagement is supported by the provision of intermittent feedback, integrated activities, simulations and games. Tools for expression could include text input, audio record, image capture, and movie record. UDL features can be utilized in several manners to best support the content and to aid in information processing and conceptual understanding. By making instructional goals, strategies, and materials flexible in these ways, potential barriers to learning are lowered and opportunities to learn are increased.

By considering user characteristics for a diverse population, design can be directed in a manner that supports the population. Guidelines that consider characteristic groups promote accessibility for all. In turn, a resulting technology designed for universal accessibility allows the greatest overall gain for the entire user group. UDL theory provides guiding principles to help approach addressing accessibility needs. Yet, a means for how to best employ UDL tools in order to afford learning remain unclear.

Particularly when considering the process of transferring information from the short term to the long term memory, which is essential for learning to occur, there is the contextual need for mapping accessibility to cognitive demands and learning goals. Consequently, UDL tools, when placed in the context of higher cognitive load may result in gains different from when placed in a context of low cognitive load. Furthermore, the application of UDL tools for processing information through the short term memory may impose varying levels of cognitive load dependent on information modalities – acquisition, processing, and output, which are components of instructional design.

2.2 Cognitive Load Theory

2.2.1 Human factors in education.

The field of human factors studies how humans accomplish tasks in a system and how this accomplishment is affected by behavioral and non-behavioral variables (Wickens & Hollands, 2000). Certain fundamental goals in human factors include reduction of error, increase in productivity, improvement of safety, ease of use, and enhancement of comfort while the human in engaged in the system. While traditional human factors stemmed from aviation, the field expanded into many other domains, such as technology, with the development of the computer and computer applications (Meister, 1999). Human computer interaction looks more specifically at these goals for users interfacing with computers, such is the case with educational technology. Yet, consider also the additional system components that may be affected when educational technology is introduced, a few examples being: teacher-student interactions, classroom management, student-student interactions, and classroom layout.

In the domain of education, research has shown much of the variability in cognitive performance and learning can be attributed to specific design features of the learning environment, as opposed to innate biological factors; yet, human factors research is lacking in contributions to improving student learning and performance of educational systems (Smith, 2007). And, if contributions to system improvements mimic improvements in other domains, it can be reasonably assumed that gains can be made by studying the educational system from a human factors perspective. For educational technology, specific human factor gains, which are not met by conventional human computer interaction principles, could be derived in the design of instructional materials. Furthermore, universal design requires additional considerations for an expansion of the traditional user experience to a broader population with a depth in needs requirements.

2.2.2 Components of cognitive load theory.

When considering human factors in the design of instructional materials, as is in the case of educational technology, cognitive load theory (CLT) has many implications for the design of effective learning materials (Paas et al., 2004). Derived from the cognitive processes that are needed for learning and understanding, CLT is based on properties of the short term and long term memory structures. Short term memory structures are described as limited capacity storage that can contain around 7 unrelated items (van Merrienboer & Ayres, 2005). Alternatively, long term memory structures are considered relatively unlimited and hold cognitive schemas, or knowledge structures, that vary in complexity and levels of automation. Numerous studies have shown that instructional formats that ignore the considerations of these concepts of CLT and involve learners in cognitive activities often impose cognitive load in a manner that impedes learning (Leahy & Sweller, 2005; Mayer & Moreno, 2003; Paas, Renkel, & Sweller, 2004; Sweller, 1999, 2003). Therefore, this problem should be amenable to design modifications that can improve performance outcomes in the tradition of human factors.

Within this framework, there are three primary types of cognitive load - intrinsic, extraneous, and germane. Intrinsic cognitive load, as described by

Sweller (1994), is the inherent intellectual complexity of the instructions, which is dependent on the number of elements and interactivity of the elements in the instructional task. For example, the task of learning the location of the bones in the middle ear, which is comprised of the malleus, incus, and stapes has a low intrinsic load; the task of learning the bones in the wrist, of which there are eight, has a higher intrinsic load. Sweller further categorized cognitive load with an extraneous component, which is determined by the manner in which the information is presented, such as text or diagram. Lastly, Paas and Merrienboer (1994) found that by varying the type of instruction, a final type of cognitive load, germane, supports construction of knowledge structures by providing a means of motivation and increased engagement. This can be seen, for example, by providing a variety of problem situations in order to encourage construction of knowledge structures.

Each type of load has varying implications for the design of effective learning materials, and combined are considered additive. It should be noted, though, that both conditions of underload and overload are suspect to cause decreasing performance (Paas et al., 2004). Furthermore, any of the individual load components could independently produce a state of overload for the learner, as well as could the combined, total cognitive load. However, these levels can also be managed within the capacity of the short term memory. For example, in a high state of intrinsic load (complex content), a partially solved problem (extraneous load reduction) would remove the requirement of storing a stipulated number of elements in the working memory. The next sections will further describe the cognitive load theory components and their design implications.

2.2.2.1 Intrinsic load.

Intrinsic load, the complexity inherent to instructional material, can be measured by the amount of element interactivity. Sweller (1994) posits that an element is a single item, or piece of information to be learned, that can be stored in the working memory. Information that can be processed sequentially, and elements that do not interact, are said to have a low level of intrinsic load. Information in which elements interact, and must be processed simultaneously, may be more difficult to learn if there are a large number of elements to process through the working memory. This element interactivity, in turn, would have a higher level of intrinsic load.

As demonstrated by Carlson, et al (2003) varying levels of intrinsic load can be demonstrated by comparing two sets of instructional materials in Table 2.1. In Example 1, the instructional material for learning the prefix “propan” is described as “the prefix for three carbon atoms in a chain”. There are 2 corresponding interacting elements required for learning this prefix, as detailed in the adjacent cell. Alternatively, instructional materials for learning the suffix “amide” have 7 elements, as shown in Example 2. Carlson found that the test performance for learning a set of prefixes was significantly higher than was for learning an equal number of suffixes. Additionally, students reported significantly lower subjective workloads for learning prefixes when compared to learning suffixes.

Table 2.1. *A Comparison of Intrinsic Load Levels in Instructional Materials*

Example Description	**Instructional Material**	**Estimate Number of Interacting Elements**
Example 1: Learn the Prefix Propan- (Low Intrinsic Load)	The Prefix for three carbon atoms in a chain is **propan-**	1. Note number of carbon atoms. 2. Record the name of corresponding prefix.
Example 2: Learn the Suffix -amide (High Intrinsic Load)	The Suffix used if the oxygen structure in a carbon atom and oxygen atom joined by a carbon-oxygen double bond and to the same carbon atom a single bonded nitrogen atom which has two hydrogen atoms attached to it by two single bonds is **–amide**	1. Note the carbon atom in the molecular model. 2. Note the oxygen atom in the molecular model. 3. Ensure that the oxygen and carbon atoms are joined by a double bond. 4. Note the nitrogen atom in the molecular model. 5. Ensure that the carbon atom in Step 1 is joined to the nitrogen atom by a single bond. 6. Note that there are two hydrogen atoms in the model. 7. Ensure that the nitrogen atom in Step 4 is joined to both hydrogen atoms by single bonds.

(Adapted from Carlson et al., 2003)

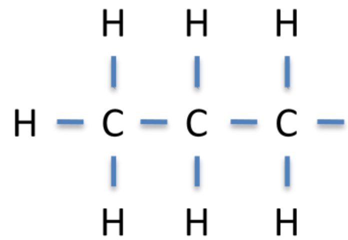
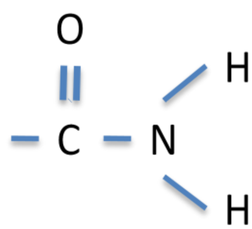
Since intrinsic load is the inherent complexity of the instructional material, it cannot be modified and are impacted by individual prior knowledge structures. However, shown by Carlson et al. (2003), intrinsic load can be measured by the estimate number of interacting elements. Consequently, it can be inferred that instructional material with high intrinsic load and high levels of element interactivity are more difficult to learn. Furthermore, instructional material with too many interacting elements may inhibit learning, as the working memory is placed in a condition of cognitive overload. Consequently, information would fail to be processed into the long term memory.

2.2.2.2 Extraneous load.

Whereas intrinsic load is dependent on the nature of the instructional material, extraneous load is determined by information presentation, or the way the material is formatted (Sweller, 1994). Consequently, the method of presentation varies the level of extraneous load. Some methods of presentation that reduce extraneous load are the inclusion of a worked example or partial solution and the use of multimodal sources such as visual information with accompanying spoken explanatory text (van Merrienboer & Ayres, 2005). However, overall cognitive load is comprised of intrinsic, extraneous, and germane load. Consider the situation where both intrinsic and extraneous load are high, due to a high number of element interactivity and poor instructional design, respectively; the overall cognitive load may, accordingly, be excessive. Conversely, if the number of elements interacting is low, the design of the instruction may have no consequence on learning outcomes.

Varying levels of extraneous load can be seen in Table 2.2, which compares text format and diagrammatic formats for the same tasks. In both formats, the learner must identify the number of carbon atoms and note the corresponding prefix name. For Example 1, the type of instructional format has little consequence, as there are few interacting elements. Alternatively, in Example 2, the diagrammatic format visually organizes the information required to learn the suffix. As the intrinsic load is higher in example 2, the learner will perform better with the low extraneous, or diagrammatic, format.

Table 2.2. *A Comparison of Extraneous Load Levels in Instructional Materials*

Example Description	**Instructional Material: Text Format (High Extraneous Load)**	**Instructional Material: Diagrammatic Format (Low Extraneous Load)**
Example 1: Learn the Prefix Propan- (Low Intrinsic Load)	The Prefix for three carbon atoms in a chain is **propan-**	**Propan-** 
Example 2: Learn the Suffix -amide (High Intrinsic Load)	The Suffix used if the oxygen structure in a carbon atom and oxygen atom joined by a carbon-oxygen double bond and to the same carbon atom a single bonded nitrogen atom which has two hydrogen atoms attached to it by two single bonds is – **amide**	**-amide** 

(Adapted from Carlson et al., 2003)

Carlson et al. (2003) tested the effects of cognitive load on learning by considering levels of high and low element interactivity with text and diagram information presentation. As the limited working memory imposes restrictions on information processing, the researchers found that in cases of lower intrinsic load (or low intellectual complexity of the task) there were no performance differences between text and diagram information presentation formats. However, for tasks with high intrinsic load, the diagram format provided an advantage, which consequently improved performance outcomes. Furthermore, students reported lower mental load for high intrinsic tasks when provided a diagram as compared to the text instructions.

Extraneous load has particular relevance in multimedia presentation, a medium that is well supported with digital presentation of instructional material, and certain principles are particularly relevant in the design of multimedia applications. There are several principles that aid in the reduction of extraneous cognitive load as described by Sweller, van Merrienboer, & Paas (1998). The split attention effect, for one, promotes replacing multiple sources of information with an integrated information source in order to remove the need to mentally integrate multiple source information. The modality effect utilizes both auditory and visual channels for information processing by varying information sources, which distributes load across multiple channels. A third example, the redundancy effect, states removal of redundant information reduces unnecessary processing and consequently reduces cognitive load. Applying these principles to e-learning interfaces similarly promote improved learning outcomes (Mayer, 2003).

2.2.2.3 Germane load.

Germane (or generative) processing can be described as the manner that learners engage in cognitive processing, which is dependent on the learner's prior knowledge and motivation as well as the prompts and supports within the instructional material (Sweller, 2005; DeLeeuw & Mayer, 2008). The germane processing is a complex activity that includes mentally organizing information and relating new information to prior knowledge; thus, information is transferred from the short term memory to the long term memory through the creation of knowledge structures known as schemata. DeLeeuw & Mayer (2008) measured germane load by categorizing responses as high-transfer or low-transfer, as seen in Table 2.3, with correct answers indicated by asterisks. A high-transfer response would include relating solutions to prior knowledge and considering multiple cause-effect relationships. A low transfer response, on the other hand, would not relate information or draw on prior knowledge. The researchers found that low-transfer students would perceive the instructional materials to be more difficult than the high-transfer students.

Table 2.3. *A Comparison of Germane Load Levels in Student Performance*

High-transfer answer	**Low-transfer answer**
Q: Suppose you switch on an electric motor but nothing happens. What could have gone wrong? A: The wire loop could not be making a connection. The battery could be dead.* A wire could have been damaged and not carrying the charge.* Something could be wrong with the switch. Something could be jamming the wire loop preventing it from spinning.* Magnets could have become misaligned.*	Q: Suppose you switch on an electric motor but nothing happens. What could have gone wrong? A: If you switch on an electric motor and nothing happens then the wire loops might not be working. The rotation of the wires is what makes the electric motor work, so if the motor isn't working then the wires of the magnet is not working properly.

(Adapted from DeLeeuw & Mayer, 2008)

2.2.3.4 Summative Nature of Cognitive Load

According to cognitive load theory, each three cognitive load types are added to describe total cognitive load. For the design of effective instructional materials, extraneous load should be kept low, germane should be kept high, and each should be considered in amount with respect to the intrinsic load imposed by the material (Paas et al., 2004). The total cognitive load should be maintained within the limitations imposed by the working memory and the ability to automate schemata.

Studies in cognitive load theory show the importance of considering cognitive load on instructional design and learning outcomes. Hollender, Hofmann, Deneke, & Schmitz (2010) present the need for the integration of cognitive load theory with human computer interaction as the complexity of the electronic learning environments increases and as technology provides a new medium for instructional material. CLT for the e-learning environments, which leverage a multimedia platform, can arguably have many implications for the design of effective learning materials and interactions.

2.3 Cognitive Load Theory for Accessible Educational Technology

2.3.1 Statement of the problem.

In the field of human factors, universal accessibility is largely omitted from leading human computer interaction research and textbooks. Additionally, contributions of human factors for improving student learning and performance of educational systems is lacking. Jonassen (2003) claims that achievement of higher levels of thinking can be gained by reduction of cognitive load to visualize and represent problems via computer applications. Paas et al. (2004) further postulate that conditions of cognitive overload or underload may cause a reduction in learning. Moreover, there is an identified need for methods in design and evaluation of educational software systems considering CLT (Hollender et al., 2010).

Integrating UDL tools provides a means to design for student accessibility pertaining to information input, processing, and output. This poses the need for consideration of universal accessibility for each of these information transfer modes. For information input, this includes alternate forms of multimedia presentation. Pertaining to information processing, cognitive aids include highlighting, concept maps, and embedded glossaries. For information output, multiple methods for knowledge expression by means of text, audio, and image are relevant features. By considering cognitive load demands in content presentation for each of the aforementioned information modalities (input, processing, and output), these accessibility features could have a positive effect on performance outcomes and UDL feature utilization.

By considering accessibility and human computer interaction principles in the design, the widest population of students is reached. By also including CLT as a component for the design of the accessible educational technology, one can seek to understand the relationship between content, information mode, learner, and accessibility with respect to mental demands in varying instructional designs. This also drives exploration of the role that cognitive supports play in the reduction of cognitive load for students utilizing the software features. In order to promote technology design that is accessible and appropriate, educational multimedia applications should incorporate human factors design methods and consider design alternatives by employing principles of UDL in congruence with cognitive load theory.

However, while design for all may be an ideal approach to interface design, and certainly provisions should be made to accommodate this goal state, there may be elements of cognitive load that account for performance variability and UDL feature utilization. By employing CLT in accessible interface design, certain tradeoffs may come in to play. Therefore, designing an interface with considerations for both UDL features and CLT may provide a clarification to some, and an extension to other, accessible design principles for educational technology.

2.3.2 Research objectives.

Though educational technology promises gains in learning performance, human factors should be considered in the design of this electronic form of instructional material. To reach a varied user base, considerations for universal accessibility promote the largest potential gain to the whole by the provision of supports for information input, processing, and output tasks. Cognitive load theory provides design principles that can be leveraged in the instructional presentation modalities to promote learning gains. However, while several principles in universal accessibility and CLT, such as the split attention effect complement each other, several others conflict. For example, universal accessibility promotes redundancy while cognitive load theories admonish expending cognitive resources on extraneous information presentation. Such conflicting principles pose the following research question: What impact do UDL tools have on cognitive load?

Moreover, consider cognitive load with varying cognitive capabilities, as is the case when designing for universal accessibility. Kalyuga, Changler, & Sweller (2000) note that levels of user expertise should be taken into account when designing multimedia presentation for instruction. Cognitive load imposed by processing learning material could, in fact, depend on learner knowledge (Kalyuga & Sweller, 2004). Yet, what remains unanswered is whether learner expertise, which may have an effect on cognitive load, shows performance variations. In the integrated classroom, consider varying levels of expertise in users to span the diverse population. This diversity in "learning expertise" further posits the research question: What effect do UDL tools have on task performance with varying levels of cognitive load?

2.3.3 Research phases.

This research was conducted in three main phases, as detailed in Figure 2.1, in order to explore the research objectives: (1) the impact of UDL tools on cognitive load and (2) the effect of UDL tools on task performance with varying levels of cognitive load. In the first phase, an exploratory analysis was conducted through the design and implementation of 3 interface designs across two Boston public school classrooms. Each design was iterative and included feedback from teachers and students pertaining traditional user-centered criteria: navigation, ease of use, consistency, and utility of features. The primary purpose of this phase was to provide initial exposure of the software to the students, assess classroom management needs, and test initial uses of the UDL tools.

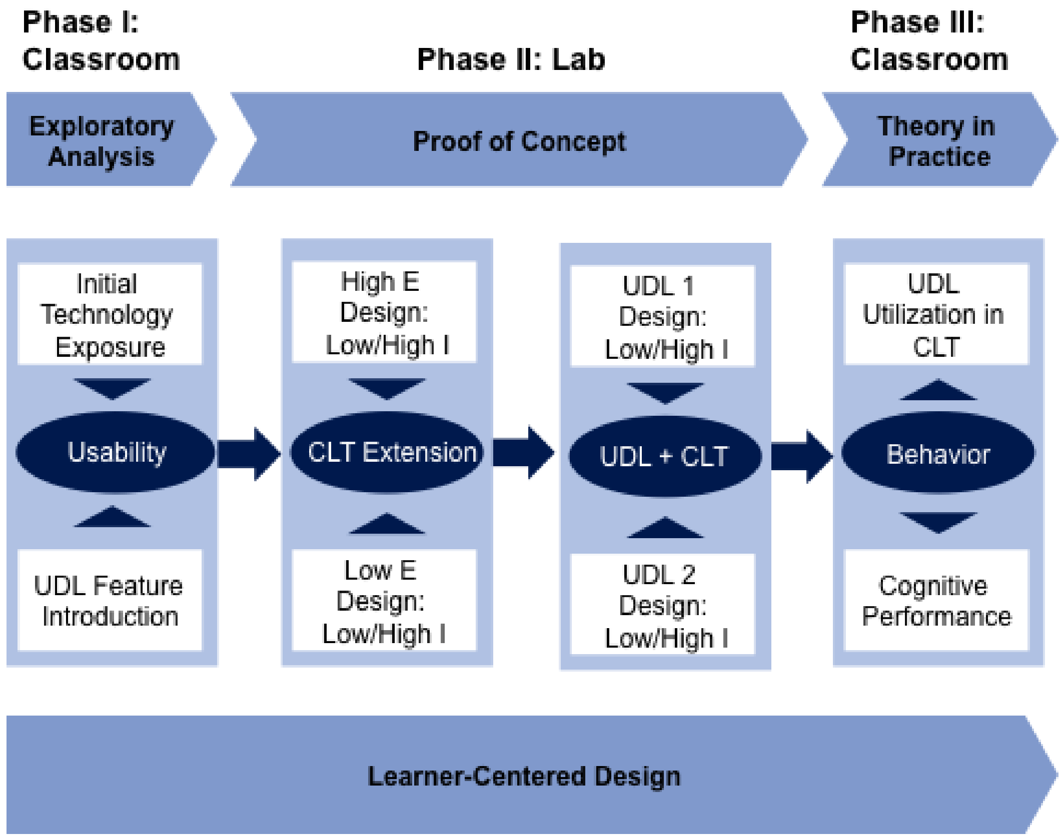


Figure 2.1. Research phases

The second phase was conducted in the laboratory for the purpose of extending ideas of CLT into curriculum content derived from the classroom and testing theory that relates UDL tools to extraneous cognitive load. This phase consisted of two main implementations. The initial part of the second phase, designs "High E" and "Low E", representing high and low extraneous loads via instructional format, were a replication of the science instructional experiments conducted by Carlson et al. (2003), however conducted within a digital environment as opposed to paper-and-pencil media. These designs tested varying intrinsic (I) and extraneous (E) load for several science tasks, including an extension of CLT into new curricula. The successive designs, UDL 1 and UDL 2, tested two UDL tools, concept map and highlighting, on the experimental tasks from the initial designs.

The third phase was conducted to look at UDL usage in the classroom relative to varying levels of intrinsic and extraneous load. In doing so, the concept was explored through real time data tracking of students in the actual context of use. This final phase allowed for investigation of the integration of UDL tools with CLT on student performance in the classroom.

2.3.4 Assumptions.

This research was targeted to run in conjunction with NSF grant #0930896 "Improving STEM learning through interactive RoboBooks." All classroom research was conducted at Boston Arts Academy and Fenway High School in Boston, MA. It should be noted that data was collected in an opportunistic manner throughout the iterative design process, particular curriculum decisions were driven by the syllabus, and the specific UDL tools were defined prior to the outset of this research.

3. UDLs in the Classroom

3.1 Objectives

3.1.1 Phase I description.

RoboBooks, a digital notebook environment developed by the Center for Engineering Education and Outreach (CEEO) at Tufts University, supports students in scientific investigations through interactive presentation of material and provision of dynamic tools for documentation and journaling of student work. The self-guided environment extends overall opportunities for learning by means of scaffolding, interactivity, and real time feedback (Danahy, Goswamy, & Rogers, 2008). As an educational platform, RoboBooks provides a means for interactive, constructivist, inquiry-based curriculum delivery. Moreover, the manipulable interface permits the inclusion of UDL tools. The primary goal of Phase I was to integrate UDL tools into the RoboBooks interface and provide students with a usable, accessible interface.

A human factors approach was taken in order to achieve these objectives and will be detailed in this section. Learner-centered design principles defined by Soloway et al. (1994) provided a method to integrate UDL tools in the RoboBooks platform. Such principles included eliciting teacher and student input and feedback to inform the design, ensure usability, and ultimately promote improved opportunity for student achievement. As such, an initial exploration of the software in the classroom and an iterative implementation for developing the UDL tools in the RoboBooks software platform was defined in a three-session design process, as shown in Figure 3.1. This thesis will peripherally cover the session objectives and findings from phase I in order to illustrate the user-centered design practices followed; however, detailed analysis will focus on the final session, particularly with navigation tracking and associated indications for the subsequent research phases.

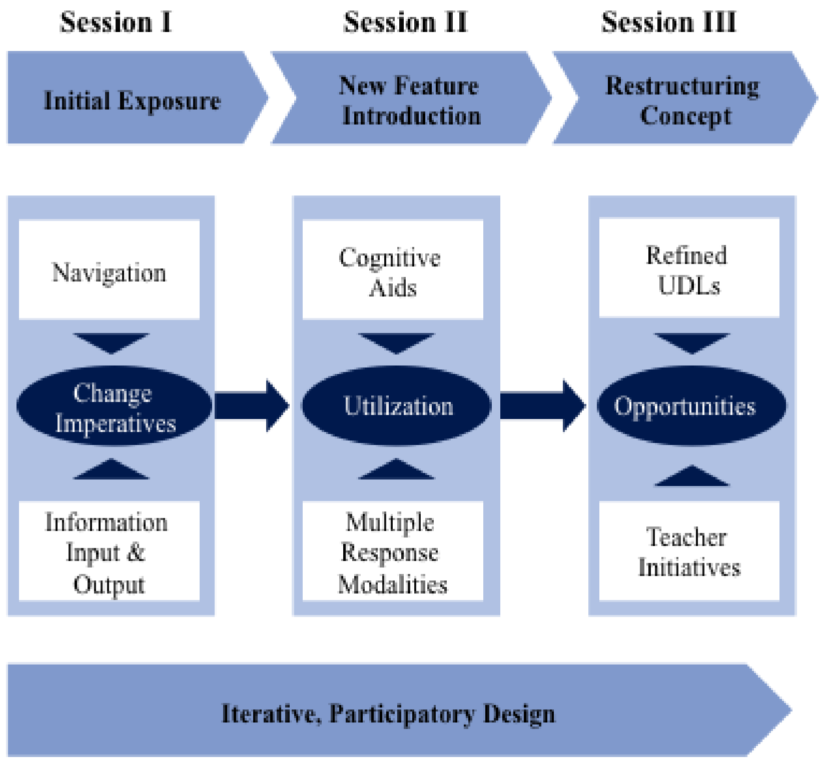


Figure 3.1. *Phase I design plan*

3.1.2 Session objectives.

The design plan was created to establish specific objectives for each session. Throughout each phase of the pilot testing, student and teacher feedback was evaluated in order to determine the appropriate design changes for the subsequent phases. Comprehensively, the goal of the first session was to introduce RoboBooks to the classroom. The second session objective was to initially introduce UDL tools in the RoboBooks platform. The third session objectives were to refine the UDL tools from session 2 and include teacher-led initiatives. This design model allowed for a gradual introduction of the RoboBooks platform to the classroom and the refinement of new RoboBook features.

3.2 Apparatus

3.2.1 Hardware.

All classroom sessions were conducted with 13" MacBook laptop computers operating with OS X version 10.6.7. The classroom tests required input from the built-in track pad to navigate the RoboBooks software and input from the built-in keyboard. Use of the built-in web camera and microphone were also used by students to respond to questions in alternate modalities in RoboBooks.

3.2.2 Software.

3.2.2.1 Pedagogical foundation.

Learning, according to Piaget, involves the construction of new knowledge from existing knowledge through the manipulation of artifacts and direct observation of consequent behavior (Duckworth, 1973). The constructivist theory concept has been extended by Papert to constructionism, the pedagogical foundation from which RoboBooks was derived. Constructionism is a learning theory that posits active construction using tangible objects to explore the world produces improved learning outcomes. As an interactive notebook, RoboBooks provides a manipulable interface, live connection to hardware, and interactivity with real-time data access (Danahy et al., 2008). Multimedia elements such as text, images, audio, and movies allow students to explore content in a variety of ways. RoboBooks support students in scientific investigations through interactive presentation of material and dynamic tools for documentation and journaling the students' work. Furthermore, the exploration, manipulation, and creation provided by interacting with RoboBooks promote the construction and expression of knowledge schemata in complex problem-solving arenas.

3.2.2.2 Specifications.

The RoboBooks software version used in this research is driven by a LabView engine. The user interfaces with an HTML front panel hosted within the Firefox web browser. Navigation through the software is enabled by use of the computer's mouse pad to access previous or successive pages, as indicated on the screen. Upon selection of a master RoboBook, which contain lesson instructions, activities, and prompts, a unique student book is created and stores all student responses.

3.2.3 Data collection.

3.2.3.1 User surveys.

User surveys were provided to the students upon completion of each session; see Appendices A, B, and C for complete surveys. These surveys were developed to elicit feedback pertaining to navigation, transparency in functionality, and consistency throughout the interface, traditional user-centered deign practice, as well as determine utility of the UDL tools. The survey results were used to inform the design of subsequent sessions.

3.2.3.2 Pre & post tests.

In session 3, cognitive gains were assessed by provision of pre and post content tests. The teacher derived the test questions, which were used to determine if learning gains occurred. The questions were the same in the pre and post treatment tests.

3.2.3.3 Navigation tracking.

Navigation tracking was developed in order to provide detailed feedback pertaining to student behavior in the software. This tracking script assigns a time stamp in HH:MM:SS for all button presses, which included movement between pages and selection of several of the UDL features. Consequently, navigation tracking can be used to determine relevant time to task information as well as feature utilization.

3.2.4 Environment.

The exploratory analysis was conducted in one class at Fenway High School and two classes at Boston Arts Academy (BAA). The Fenway class was a yearlong, 9th grade Physics course. The 8 students in the class were all male and each categorized in the high incidence disabilities group. Each student was provided his own laptop computer. The two BAA classes were semester long, 10th grade Chemistry courses. Of the 23 students, 22% are categorized in the high incidence disabilities group, 8 were male and 15 were female. Due to resource constraints, in most cases students were paired at a laptop, unless student absences resulted in single computer use.

3.3 Session 1

3.3.1 Session objectives.

The primary objective for Session I was to familiarize students with navigating the RoboBooks software. This initial session allowed RoboBooks to be tested with each classroom setup. Additionally, the researchers collected preliminary observations and initial impressions from students and teachers pertaining to navigation & information presentation.

3.3.2 Methods.

3.3.2.1 Instructional materials.

The RoboBooks for Session 1 were scheduled in congruence with introductory subject material in the classroom. Lab safety and measurement topics were presented to the students in the chemistry and physics classrooms, respectively. The content was presented in the RoboBook in multiple combinations of text, supplementary audio, videos, simulations, games, and activities. Activities were presented in the digital platform, which required input from the students in their personal workbook. For all sessions, the content topic and curriculum was selected and verified by the teacher prior to classroom testing. (Note: The chemistry teacher had over 20 years of experience; the physics teacher had 3 years.) Refinements and edits to the content were made from initial feedback.

3.3.2.2 Interface.

Initial user characteristic guidelines, as established by Gribbons (2008) pertaining to reading, memory, metacognitive and search and navigation, informed the design of the user interface. These guidelines were leveraged to structure the information, ensure consistent navigation, and promote overall usability for the population demographic. Each of the interface elements had specific design criteria determined from the user characteristics and session objectives.

As shown in Figures 3.2 and 3.3, several constructs were considered in the design of the interface. Figure 3.2 highlights the main navigation elements, which were defined to include a left-hand chapter bar, help and home buttons, and a lower page forward and backward buttons. Figure 3.3 exemplifies information presentation by means of information presentation as text, accompanying text to speech and supporting video. Each of the remaining constructs, response options, content, and layout, were similarly structured to afford the functionality and learning objectives.

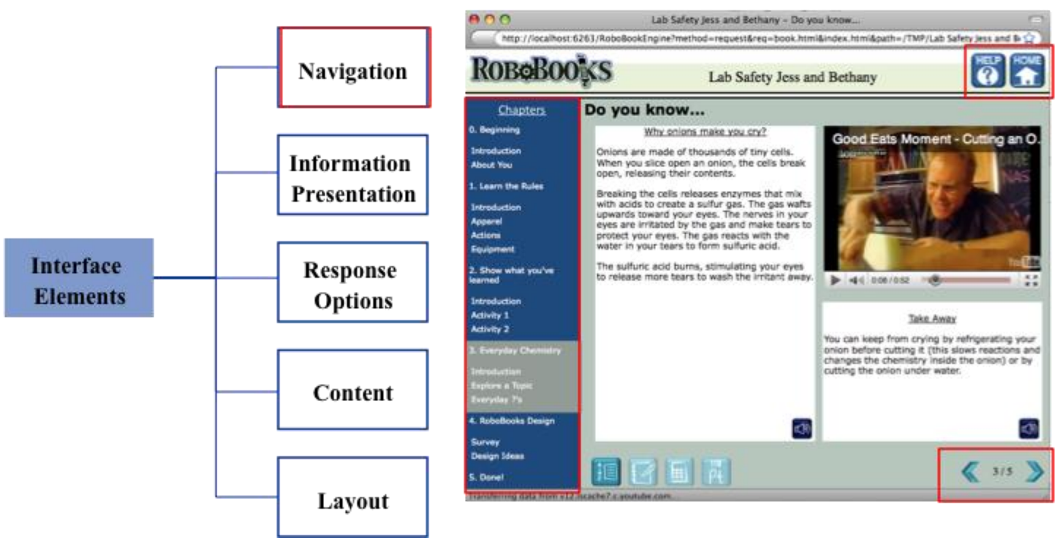


Figure 3.2. *Screenshot of navigation elements for Session 1.*

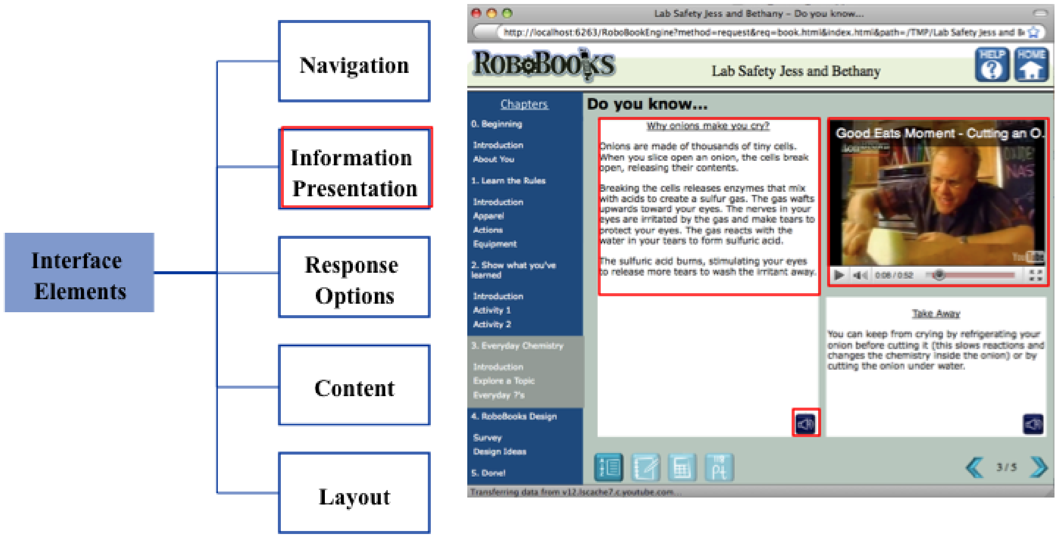


Figure 3.3. *Screenshot of information presentation elements for Session I.*

3.3.2.3 Student survey.

Each student, or student pair, worked through the digital workbook content by interacting with the interface. Upon completion of the instructional content, students were presented with survey questions in the RoboBook, as demonstrated in Figure 3.4. A full list of survey questions can be found in Appendix A. These questions were developed to elicit pertinent feedback pertaining to initial impressions and usability of the interface and inform subsequent design changes.

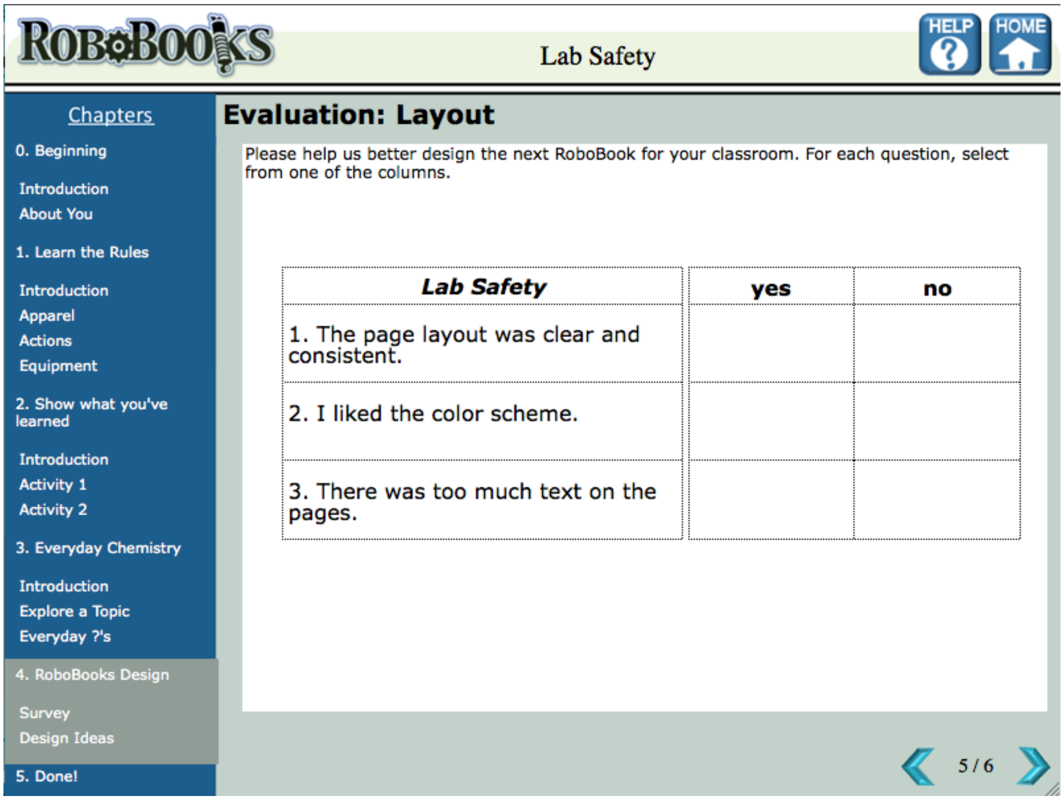


Figure 3.4. *RoboBooks survey example screenshot, Session 1.*

3.3.3 Subject information.

27 students completed the RoboBooks for their class in Session I. Of the combined class responses from BAA & Fenway students, 17 were from male students and 10 were female. Student ages ranged from 15-17. 19 of the students were provided the Chemistry RoboBook and were paired at the computers; 8 students were provided the Physics RoboBook and had individual computers.

3.3.4 Results.

Figure 3.5 summarizes the findings of survey responses to each the design constructs. 13 of the 17 elements were found to be acceptable to over 80% of the students, and were considered established successfully in the design. As indicated, reading was the least preferred method of information acquisition. Less than 43% of the respondents liked typing as a response method, while, 80% liked the audio record option, a feature that required several subsequent steps to operate at this phase. The journal function was liked by 75 % of the students and 51% found the aesthetics of the interface appealing.

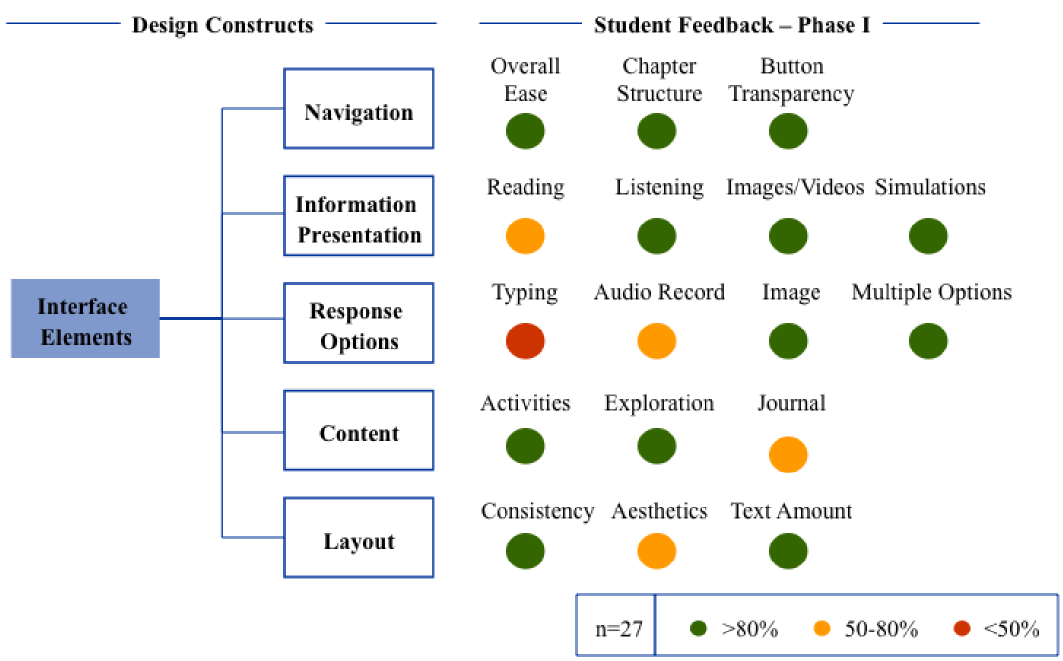


Figure 3.5. *Summary of Session I feedback on interface elements.*

3.3.5 Discussion.

The first session, primarily an introduction to RoboBooks included the goal of providing multiple means of information presentation (representation) and required the use of several response types (expression), such as audio, text, and image. Results indicate that students preferred to choose a method of response, as opposed to being required to answer a question with a mandatory response type. It was also found that the aesthetics of the interface should be improved.

These findings determined the design objectives for the following class sessions. Consequently, response features were redesigned to allow for option selection. A new color palate and templates were created to promote visual engagement. And, the audio feature redesigned to streamline the record process and engage students with this response modality.

3.4 Session 2

3.4.1 Session objectives.

Objectives for the second phase included the addition of UDL tools, creating activities with means for pacing, and collecting preliminary data pertaining to usability with students. The second phase was designed to provide the students with multiple response options for all answer prompts and a new color scheme. Teacher checkpoints (incorporating group discussions within the material) were included in order to test a more integrated classroom style, as well as the addition of regular, automated assessment and feedback to the students that indicated a level of subject mastery as they progressed.

3.4.2 Methods

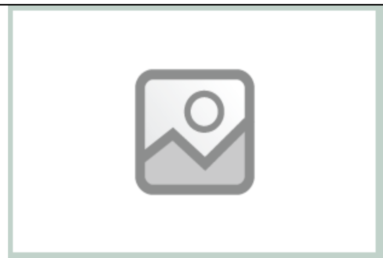
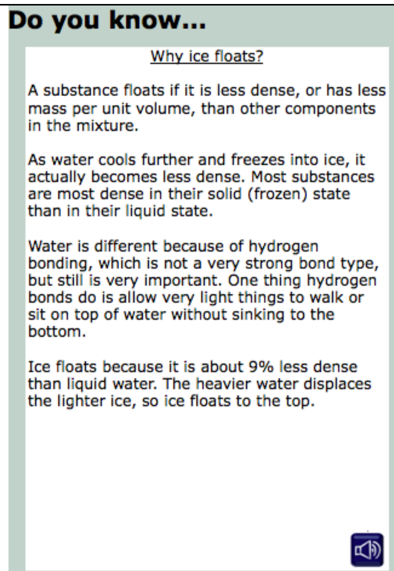
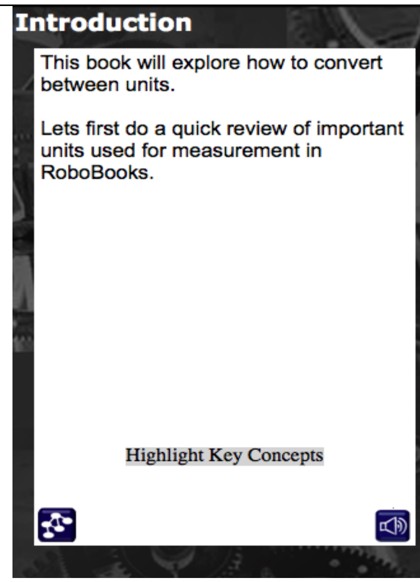
3.4.2.1 Instructional materials.

The content for session 2 was selected by and developed with the teachers. For chemistry, the topic was bonding; for physics, the topic was dimensional analysis. This session spanned two days. As in session one, a variety of instructional presentation and activities were presented in the RoboBook and were followed by a user survey. Additionally, intermittent checkpoints were established so that the teachers could assess the progress of the class.

3.4.2.2 Interface.

Several design changes were made in the second session, as shown in Table 3.1. One major change was to develop a multiple response modality, which permitted selection of response modality by the student for each question prompt. The color scheme was modified, font size increased, and highlighting, concept maps, and hints were also introduced in this design. Text font size was increased, as paired computer use imposed new design requirements for readability of the text.

Table 3.1. Sample screenshots of major design changes in Session II

<table>
<tr><th></th><th>Session I Interface</th><th>Session II Interface</th></tr>
<tr><td>Response Modality</td><td></td><td></td></tr>
<tr><td>Aesthetics and UDL tools</td><td>
</td><td>
</td></tr>
</table>

3.4.2.3 Student survey.

Upon completion of the content RoboBook, students were presented with survey questions, as detailed in Appendix B. The intent of the survey was to elicit responses from the students on usability of the interface. The questions were presented in alternating inflections, such as a positive inflection: “I would imagine that most people would learn to use RoboBooks very quickly” compared to a negative inflection: “Learning how to use all the functions in RoboBooks is difficult”. All responses were recorded in the RoboBook.

3.4.3 Subject information.

34 students completed the RoboBook in Session 2 between BAA and Fenway classes. Of the participants, 22 were from male students and 12 were female. Student ages ranged from 15-17. 26 of the students were provided the Chemistry RoboBook, and were paired at the computers; 8 were given the physics RoboBook. 10 of the responses were from paired groups and 14 of the responses were from individual students.

3.4.4 Results and discussion.

Figure 3.5 presents the mean responses for the students on the survey questions from Appendix B. Positive inflection statements over 80% indicated an acceptable level of agreement by the students for the interface. Of the positive inflection statements, 10 of the 11 questions resulted in mean student responses over 80%; only ease of information location to complete activities failed to reach 80% accordance. Notably, changes in aesthetics now resulted in an acceptable level. Of the negative inflection statements, less than 20% agreement (or 80% disagreement) indicated an acceptable level of response to the statement by the students for the interface, as was found in 5 of the 9 questions. Of the negative inflection statements, more than 20% agreement with a statement was found with the following statements: too much inconsistency, content did not meet expectations, too many functions on a page, don’t feel comfortable using RoboBooks.

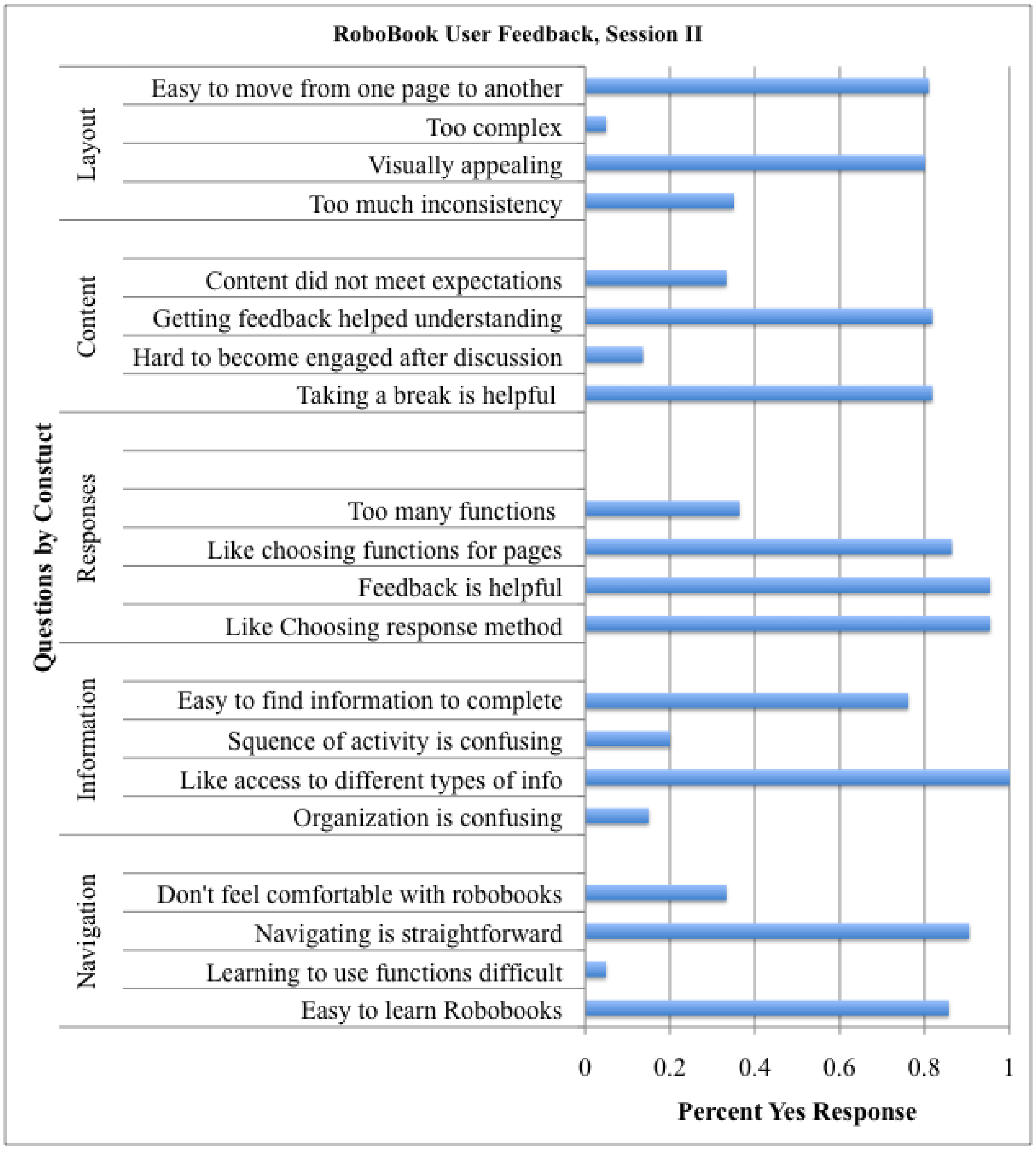


Figure 3.5. *Summary of Session I feedback on interface elements.*

Outcomes from the second session showed that overall usability had an acceptance level greater than 80% for 70% of the questions. Results also yielded design improvement criteria for session 3 to reduce multiple answer questions, accordingly reducing inconsistency in templates and a surplus in functions that accompanied multiple answer response questions. Several pre-session meetings with teachers were established in order to promote content improvements.

3.5 Session 3

3.5.1 Session objectives.

The objectives of the third phase were to integrate teacher-led curriculum, test the refined UDL tools, and pilot learning outcome evaluations in the RoboBooks platform. Several design features derived from pre-session meetings with teachers and included: activity timers, multiple choice questions that provided new content representation upon an incorrect answer selection, and mathematical calculation supports created specifically for activity tasks. Additionally, navigation tracking was collected and analyzed for this session.

3.5.2 Methods.

3.5.2.1 Instructional materials.

The content for session 3 was selected by and developed with the teachers and the session spanned three days. As in the previous sessions, a variety of instructional presentation and activities were presented in the RoboBook, as demonstrated in Figure 3.6. For the physics class, the lesson covered the topic of acceleration; the chemistry class was provided digital content for the topics of chemical reactions and chemical bonding.

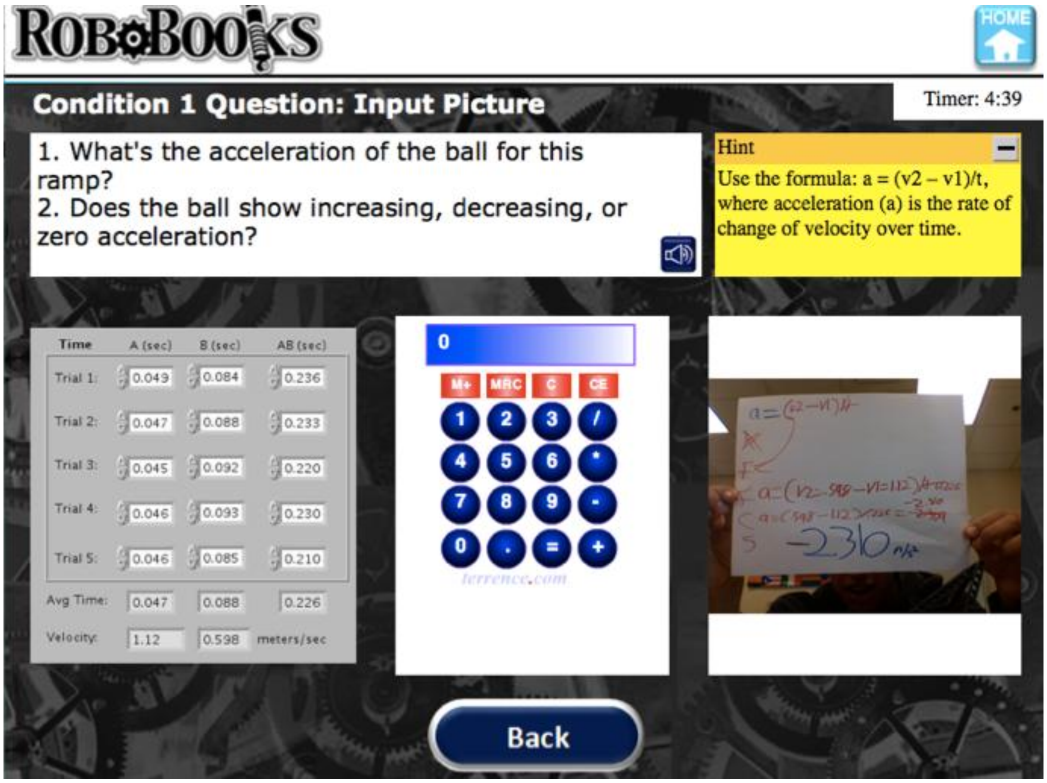
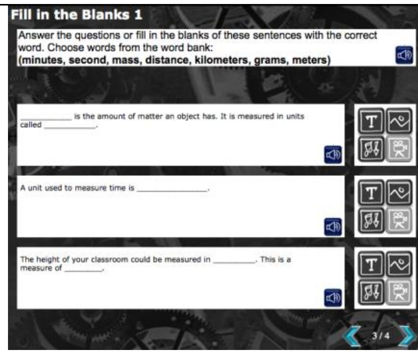
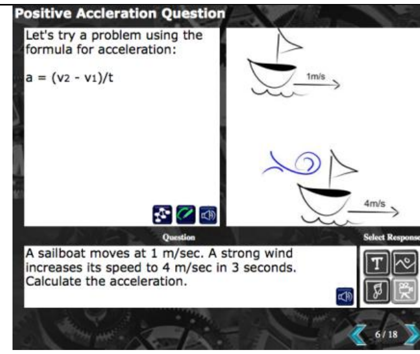
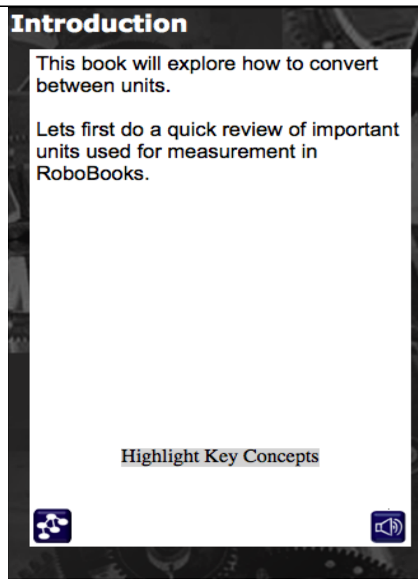
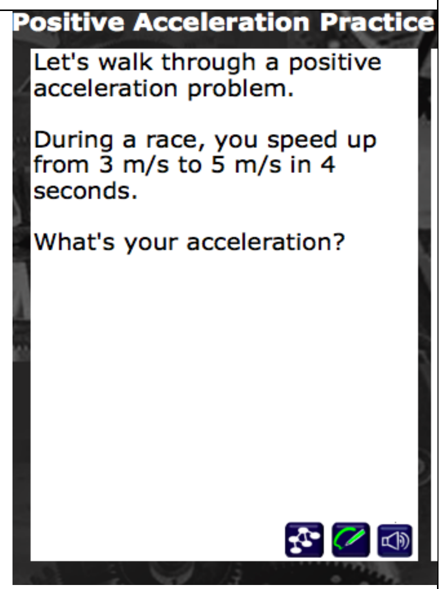


Figure 3.6. *Example activity page with student response displayed*

3.5.2.2 Interface.

Major design changes were derived from Session II feedback and are shown in Table 3.2. Pages with excessive button functions were redesigned, such that supporting information could be provided yet minimize redundant button occurrence. Button placement was revised, to provide consistency in supplementary function and congruency across pages.

Table 3.2. Sample screenshots of major design changes in Session III

	Session II Interface	Session III Interface
Page Functions		
Layout of UDL tools		

3.5.2.3 Student surveys.

Upon completion of the content RoboBook, students were presented with feature survey questions, as detailed in Appendix C. The intent of the survey was to elicit responses from the students on usefulness of the features. The 4-point Likert scale from which they selected their responses ranged from most negative (1) to most positive (4), with the intervals evenly spaced. The evaluation questions looks at the 22 items designed to gather information on the uses of, and responses to, various RoboBook features, tools, and resources. Each class was provided different content in the RoboBook, but all features (except for the periodic table which was only relevant in the Chemistry books) were provided in the RoboBooks for both classes. Additionally, a pretest and congruent posttest was administered to the students in order to assess cognitive gains. All responses were recorded in the RoboBook.

3.5.3 Subject information.

A total of 28 students completed the RoboBooks for Session III, of which 17 were males and 11 were females. The student ages ranged from 13-16 and 41% of the students had a high incidence disability. 20 of the students were paired on computers and 8 of the students individually operated a computer.

3.5.4 Results and discussion.

3.5.4.1 Usability evaluation.

Students were surveyed upon completion of the RoboBooks activities for both high school classes. The RoboBook activities were held over 3 class periods. Table 3.3 shows a summary the students' responses. All items were found to be significant, meaning that the students were generally united in their responses to each of the items. There were no differences between schools, aside from the preference of the calculator and understanding the science content, for which students from the Fenway class reported higher values.

Table 3.3. Student response options to user survey in Session III

Item	Mean	SD
The option to record audio for your responses?	2.25	1.14
The option to type for your responses?	3.39	0.86
The option to take pictures for your responses?	2.96	1.14
The option to make a stop motion movie for your responses?	3.32	0.91
The videos about science concepts?	3.52	0.59
The text read out loud?	2.79	1.22
The activity timer?	2.11	1.03
The questions with right or wrong feedback?	3.39	.88
The teacher RoboBook projected on the whiteboard?	3.52	0.79
The movies that went with a text passage?	3.43	0.84
The images that went with a text passage?	3.41	0.75
The concept map?	2.79	1.26
The highlighting?	2.60	1.20
The hints?	3.25	1.00
The calculator?	2.75	1.32
The averaging tool?	2.96	1.21
The games and simulations?	2.54	1.17
The science content that was presented in the RoboBook?	3.29	3.10
Did RoboBooks help you understand the science information?	3.11	0.79
Would you like to use RoboBooks in the future?	3.29	0.71
Overall, did you enjoy learning about science when using RoboBooks in class?	3.29	0.66
Did you enjoy working on science activities in a digital environment?	3.44	0.64

The mean values in the user survey tend toward the positive, with only 2 of the 22 items (9%) below the 2.5 threshold for moving into more positive judgments. 59% of the items generated mean responses in excess of 3.0. The narrow standard deviations provide additional indication that the students were strongly united in their assessments of the many features of RoboBooks.

3.5.4.2 Cognitive gains.

Table 3.4 shows pre-post data gathered from the survey items designed to explore the learning gains for the Physics RoboBook for session 3. In Questions 1-3, the students were asked to evaluate simulations of cars in motion over time in order to determine states of acceleration, deceleration, and constant velocity. Questions 3-6 were used to further probe understanding of the acceleration formula in terms of criteria and constraints. Each question was scored such that the received a 0 for an incorrect response, a 1 for a partially correct response, and a 2 for a completely correct response. The mean and standard deviation (SD) data is from the Fenway classroom (n=8). (Note: due to complications in facilitating paired student cognitive testing in the RoboBooks, the cognitive gains were not derived for the chemistry class.)

Table 3.4. Learning Gains

	Mean: Pre	**SD: Pre**	**Mean: Post**	**SD: Post**
Question 1	0.25	0.46	0.83	0.75
Question 2	0.25	0.46	1.17	0.75
Question 3	0.25	0.46	0.33	0.52
Question 4	1.00	1.07	1.83	0.41
Question 5	1.00	1.07	1.00	1.10
Question 6	0.25	0.71	1.00	1.10
Total	3.33		6.17	

This preliminary evaluation of cognitive outcomes shows pre to post test score improvement for all questions except question 5, where there was no mean change. Comparing mean values indicates an overall positive trend in performance outcomes. Though not used to characterize the student population, from a preliminary evaluative standpoint using the RoboBooks environment appears to have potential for achieving learning gains for students with disabilities.

3.5.4.1 Navigation behavior.

The navigation tracking data for students was used to map behavior patterns of the students throughout a particular RoboBook. For example, Figure 3.7 shows the progression of the Fenway students through the Session 3 Robobook, by page in the digital workbook (y-axis) and over time throughout one class period (x-axis). The grey areas to the left of the graph show major tasks within the RoboBook for that day. Each students' behavior within the books is shown by the individual color lines. An arrow indicates improvements in pre to posttest cognitive gains (green arrow) or now change in test performance (yellow arrow).

This visual representation of tracking data shows behavior that trends similarly (as seen in pages 1 – 9), a convergence point (pages 10 – 12), and divergence afterwards. These similar and diverging patterns indicate sections of the RoboBook driving behavior and likewise sections that are determined by individual behavior. This additionally implies that there is transparency in the interface, such that students do have the opportunity to interact with the interface without it dictating their full span of behavior. Furthermore, there appears to be evidence of task driven behavior, when sections of convergence and divergence are mapped to type of activity occurring on specified pages in the book.

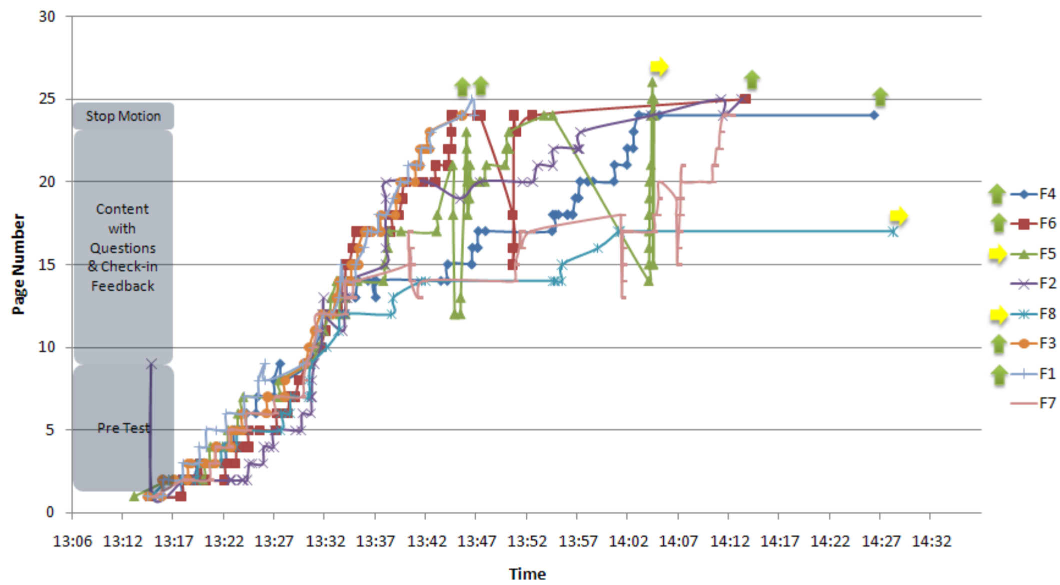


Figure 3.7. *Fenway navigation data for Day 1, Session 3*

Alternatively to trends of diverging behavior from Figure 3.7, Figure 3.8 shows similar behavior for students in one of the BAA classes. On this particular day, the teacher was projecting the RoboBook at the front of the classroom as the students followed along at their own computers. This graph exemplifies the effects of the teacher minimizing variability in student navigation as classroom management techniques reduced students' individual exploration. In later tasks, denoted with the grey boxes, it should be noticed that divergence again occurred as the teacher permitted self paced work at this point.

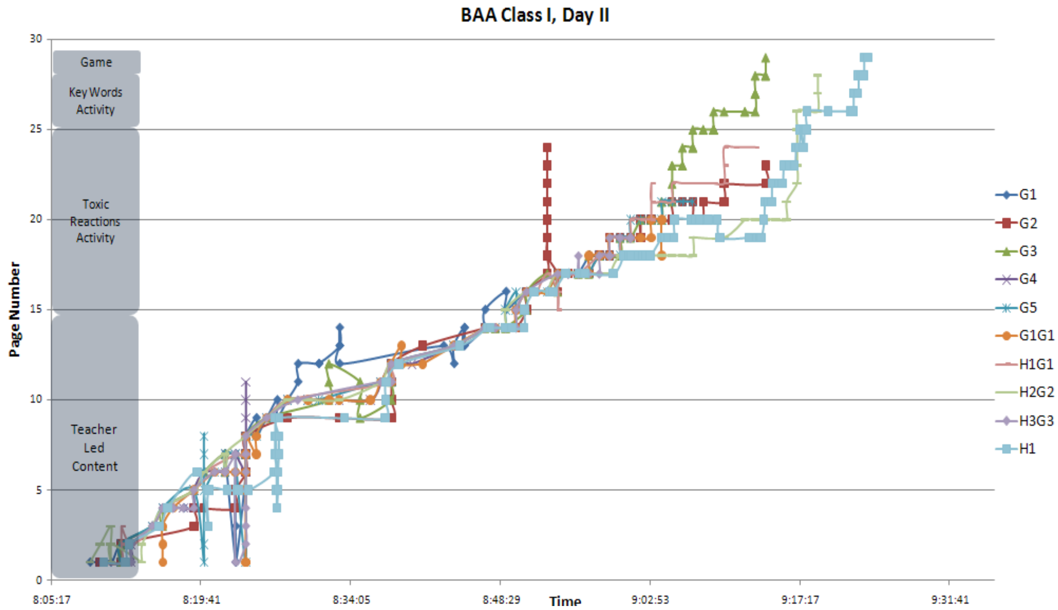


Figure 3.8. *BAA navigation data for Day I1, Session 3*

Upon noticing possible task driven behavior in the navigation data, further investigation of time to page was conducted. Figure 3.9 looks at 5 students' task times by page for the second day RoboBook at BAA. In this graph, time on page for instructional presentation, answering a question with multiple modalities, answering an intermittent check-in question with multiple choice, and time interacting with a simulation, are shown. Notably, the different tasks result in varying behavior amongst the students. It can be inferred that decision-making tasks, depicted by the spikes in the graph, require more time to complete.

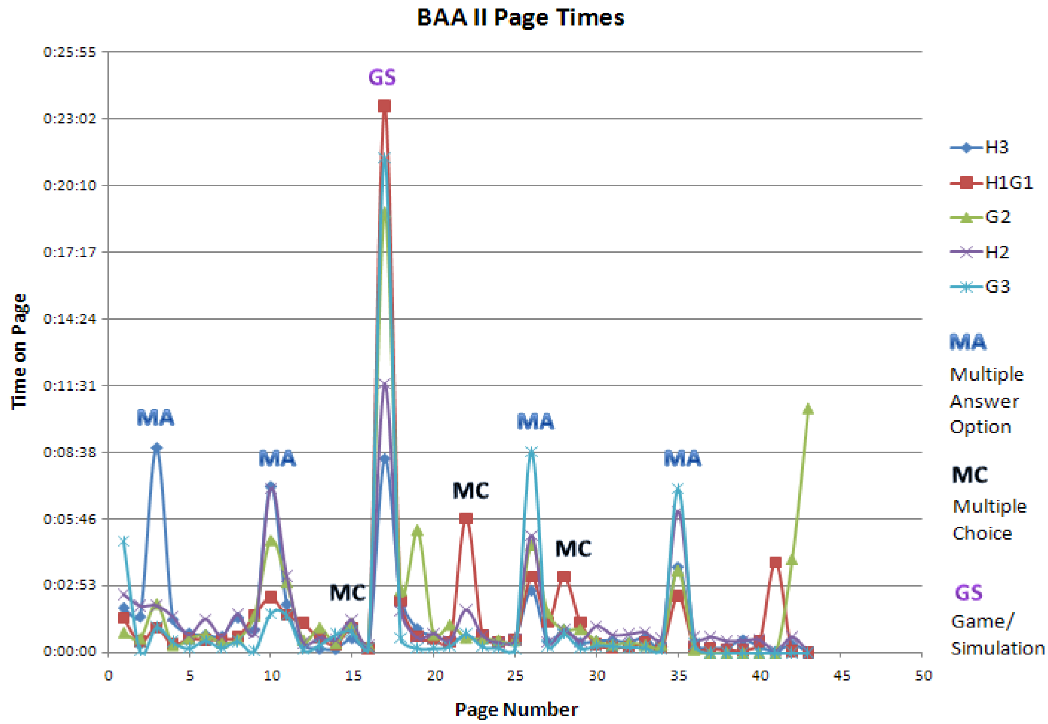


Figure 3.9. *BAA time on page navigation*

These task times were further investigated by considering time to task for the three main portions of the RoboBook, labeled as Tasks 1, 2, and 3. Student pairing are separately considered with regards to the task times in Figure 3.10. As coded, H is high incidence disability student groups, HG is a high incidence and general education student pairing, and G is a general education student pairing. Notably, in Task 1 and 3, the H group means are longer than the other groupings. While these tasks independent measures, there was not indication of significance between tasks times. However, the mean group performance difference by time approaches significance in Task 1, suggesting a possible difference in performance dependent on user characteristics.

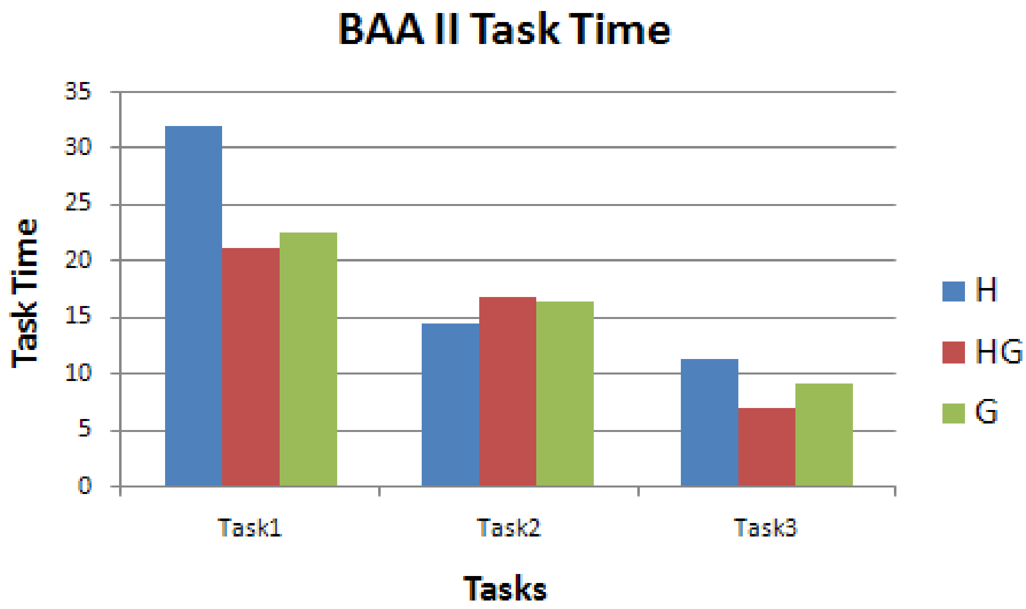


Figure 3.10. *BAA navigation and task times, by grouping*

Additionally, navigation behavior with inclusion of UDL tool usage was investigated on an individual student basis, as exemplified in Figures 3.11 and 3.12. In the first graph, the navigation through the book is tracked, with instances of text input, audio input, picture input, video input, concept map use, and feedback questions indicated by time and page. This student showed pre to post test cognitive gains. Notice the use of a variety of input modalities option use of the concept map several times. This could indicate higher engagement and cognitive aids potentially promoting learning gains. Alternatively, the student in Figure 3.12 predominately chose to answer questions with text only and only utilized the concept map feature once. This student did not exhibit cognitive gains in the pre to post test performance. It should be noted that these trajectories are anecdotal and represent the variance in student behavior, however implications of utilization and performance gains required further investigation.

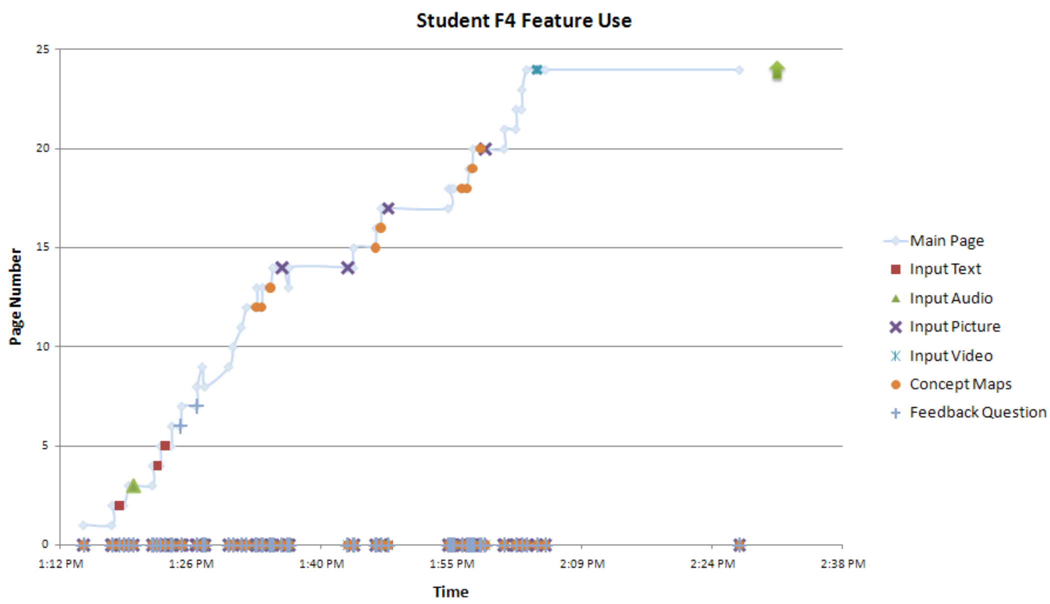


Figure 3.11. *Feature utilization for Fenway student with improved performance*

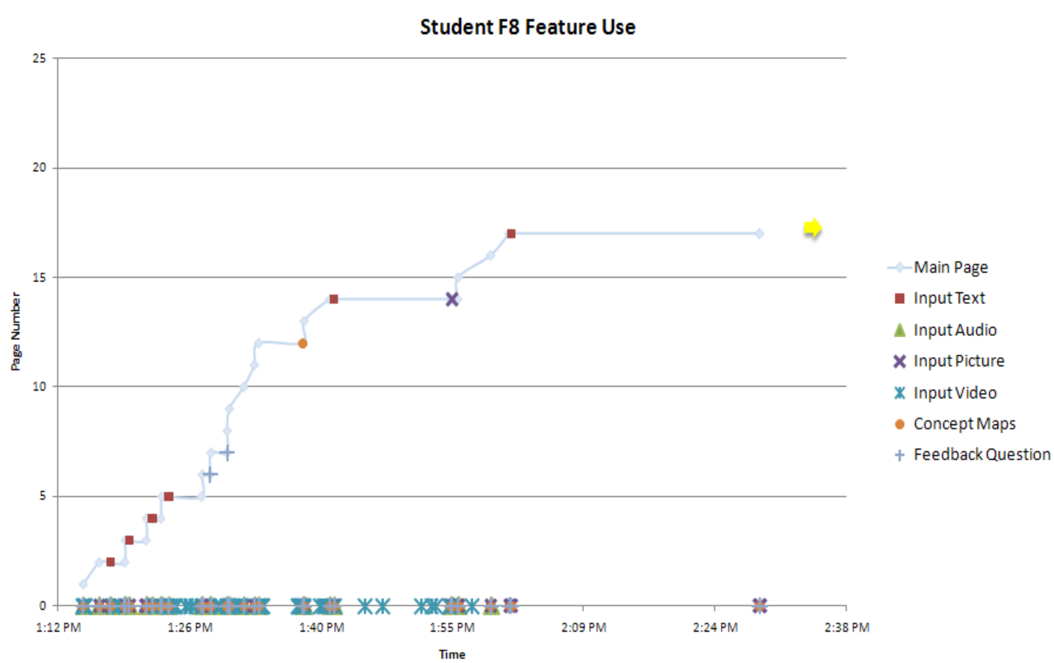


Figure 3.11. *Feature utilization for Fenway student with no performance gains*

3.6 General Discussion

Conducting design driven software development, with a learner-centered focus and actual classroom testing, provided the opportunity to solicit student and teacher feedback to inform the design of educational technology. Each implementation determined direct design objectives for subsequent phases and investigated focused research objectives. Ultimately, by conducting learner-centered design in the classroom, the researchers were able to make design decisions to better serve student and teacher needs while developing the educational technology. This approach allowed for the informed development of UDL tools in the RoboBooks software. Preliminary cognitive and usability feedback provided insight into the utility of the digital platform as a testing environment as well as the user evaluation of the software.

Moreover, the third phase outcomes also included data tracking that could map the utility of UDL tools with certain activity objectives. These data allude to task dependencies in UDL tool utilization, a focus for the next phase. While the data presented in this section should be considered an exploratory analysis and statistical significance was not found in these initial investigations, there is suggestive evidence shown in the navigation tracking that (1) the behavior of learners may be distinguishable by student characteristics, (2) task demands may drive performance outcomes differentially and (3) the utilization of UDL features may be task dependent. These observations can be leveraged to further explore feature utilization in a manner such that determinations for how to best employ UDL features can be determined. The next section will consider this question within the context of a theoretical framework and with a formal design of experiments in attempt to better understand the relationship between the learner, the UDL tools, and the instructional materials.

4. Cognitive Load Theory and UDLs in the Lab

4.1 Objectives

4.1.1 Experiment description.

Using cognitive load theory as a conceptual foundation to explore UDL principles, the laboratory experiments were designed to further investigate the behavior of students performing different tasks that employ UDL tools, see Figure 2.1 “Phase II”. In order to do so, the research on CLT conducted by Carlson et al. (2003) on learning and understanding science instructional material was first replicated in the RoboBooks digital platform. The intent of the experiment replication was to determine if intrinsic and extraneous load manipulations could be successfully conducted in the digital workbook in congruence with CLT as performed by Carlson. The replication, conducted in an initial laboratory session, included an extension into new content area, and the development of an additional task. The intent of the extension task was to determine if the principles of CLT were consistent across varying science tasks. In the first lab session, two designs, which have varying levels of extraneous load, were created to test three tasks; each with subtasks having low and high intrinsic load.

A second session was conducted to compare extraneous load manipulations conducted in the first session with an additional UDL component. In this session, two additional designs, in which UDL tools were employed, utilized the second and third tasks from the first session. These designs explored two UDL tools, highlight and concept map, under varying levels of intrinsic cognitive load. The objective of this session was to determine if the UDL tools in the second set of designs produced similar behavior when compared to the varying levels of extraneous load in the first session designs.

4.1.2 Research questions.

The research questions explored in the laboratory were derived from observations in the classroom as described in section 3.4. The preliminary classroom exploratory analysis suggested that the utilization of UDL features might be task dependent. By considering UDLs within the CLT framework, it can be reasoned that tasks have varying levels of cognitive load components, which would in turn have implications for UDL utilization. Specifically, this would suggest that UDL features may be best employed for tasks that have high intrinsic load and high extraneous load in order to reduce cognitive load in a manner similar to established extraneous load reduction techniques. In accordance with this implication, the following hypotheses were investigated:

H1a: Test scores are higher when intrinsic load is low as compared to when intrinsic load is high for science instructional materials conveyed through digital media
H1b: Task times are shorter when intrinsic load is low as compared to when intrinsic load is high for science instructional materials conveyed through digital media
H1c: Subjective workload is lower when intrinsic load is low as compared to when intrinsic load is high for science instructional materials conveyed through digital media

H2a: Test scores are higher when intrinsic load is high and extraneous load is reduced as compared to when both intrinsic and extraneous loads are high for science instructional materials conveyed through digital media

H2b: Task times are shorter when intrinsic load is high and extraneous load is reduced as compared to when both intrinsic and extraneous loads are high for science instructional materials conveyed through digital media
H2c: Subjective workload is lower when intrinsic load is high and extraneous load is reduced as compared to when both intrinsic and extraneous loads are high for science instructional materials conveyed through digital media

H3a: Test scores are higher when intrinsic load is high and UDL tools are employed as compared to when intrinsic load is high and no UDL tools are employed for science instructional materials conveyed through digital media
H3b: Task times are shorter when intrinsic load is high and UDL tools are employed as compared to when intrinsic load is high and no UDL tools are employed for science instructional materials conveyed through digital media
H3c: Subjective workload is lower when intrinsic load is high and UDL tools are employed as compared to when intrinsic load is high and no UDL tools are employed for science instructional materials conveyed through digital media

4.2 Apparatus

4.2.1 Hardware.

All lab experiments were conducted on 13" MacBook laptop computers with OS X Snow Leopard. An Apple Mouse was attached via the USB port. The experiment required both input from the external mouse to navigate the RoboBooks software and input from the built-in keyboard and web camera to respond to questions in the experiment.

4.2.2 Software.

The RoboBooks software is driven by a LabView engine. The user interacts with an HTML front panel hosted within the Firefox web browser by a LabView server. In the web browser, the toolbar and navigation bars were disabled to reduce extraneous window features. Participants navigated through the software by using the mouse to access previous or successive pages, as indicated on the screen. During the test portions of the experiment and between subtasks, access to prior pages was disabled by removal of the "back" button.

For each design, a master project was presented in the RoboBook "shelves", as shown in Figure 4.1. Each participant was prompted to select one of 4 master books, which correspond to each of the 4 instructional formats. Upon selection of the RoboBook, a unique participant book was created, as shown in "Modified Projects". This modifiable book stored all responses from the participant.

Figure 4.1. *RoboBook "shelves" with master & participant books*

4.2.3 Navigation tracking.

Upon creation of a participant book, a navigation tracking script recorded all actions from the participant within their unique book.   These actions include selection of forward or back buttons on each page of the workbook. Upon selection of a button, a corresponding time stamp, HH:MM:SS, was recorded for each page. This allowed the derivation of precise tasks times, as defined by movement between pages in the digital workbook.

4.2.4 Data collection.

Three primary dependent measures were collected to assess tasks within the four experiment conditions. Task time and test scores were used to quantify performance; subjective workload rating (SWR) was used to quantify mental workload, as perceived by the participant. These dependent measures were used to compare the effect of the four independent variables: high intrinsic load, low intrinsic load, high extraneous load, and either low extraneous load or UDL presentation.

4.2.4.1 Test scores.

For Tests 1 & 2, further described in section 4.3, scores were assigned per the metric defined by Carlson. Accordingly, 1 mark was assigned for each correct prefix and 1 mark for each correct suffix. A student could receive a total of 12 points for the prefix portion of Test 1, and 12 points for the suffix portion of Test 1. For Test 2, participants could receive a possible total of 6 points for the prefix and the suffix components of the test. Tests 3 & 4 were similarly constructed, such that the participant could independently receive up to 10 points for Test 3 and 5 points for Test 4 for identifying reactions and balancing equations. Per the hypotheses H1a and H2a, it is expected that test scores should be higher when intrinsic load is low and when extraneous load is reduced in high intrinsic conditions, when compared to low intrinsic load and low extraneous load, respectively.

4.2.4.2 Task times.

Task time was derived from the navigation tracking, which assigned a time stamp to each “page turn.” By taking the difference in time stamps, a page time was calculated for each of the participants. By summing the page times, a time per subtask was derived. Per the hypotheses H1b and H2b, it is expected that task times should be shorter when intrinsic load is low and when extraneous load is reduced in high intrinsic conditions, when compared to low intrinsic load and low extraneous load, respectively.

4.2.4.3 Subjective workload rating.

Upon completion of each subtask, the participant was asked to rate their workload on a 7-point Likert scale, where 1 = extremely low mental effort, 2 = very low mental effort, 3 = low mental effort, 4 = neither low nor high mental effort, 5 = high mental effort, 6 = very high mental effort, and 7 = extremely high mental effort, as shown in Figure 4.2. This scale has been researched in several studies, which have shown that students can effectively report imposed mental effort by assigning a numerical value to describe their cognitive workload (Carlson et al., 2003). It has also been shown that this subjective measure correlates highly with objective measures of mental load (Gopher & Braune, 1984). Upon completing a SWR scale, the participant continued to the next subtask or main task description. Per hypotheses H1c and H2c, it is expected that workload should be perceived to be less when intrinsic load is low and when extraneous load is reduced in high intrinsic conditions, when compared to low intrinsic load and low extraneous load, respectively.

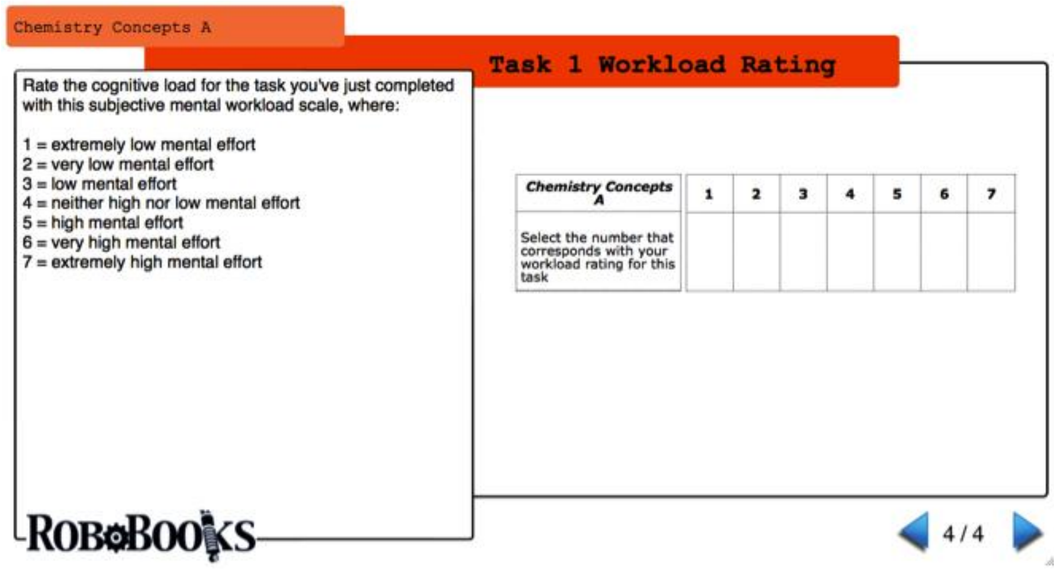


Figure 4.2 *Subjective workload rating scale in RoboBooks*

4.2.5 Environment

The apparatus, as previously described, was the same for each of the 4 conditions. All participants in the first session were assigned a master book by the researcher. Participants in session 1 were randomly assigned either the "Chemistry Concepts A" or the "Chemistry Concepts B" master book; session 2 participants were likewise assigned the C or D master books. All experiments were run in Anderson 001, Human Factors Lab, at Tufts University, to provide a noise and distraction free environment to the participants.

4.3 Session 1: CLT Replication and Extension

4.3.1 Session objectives.

As determined by Carlson et al. (2003), extraneous load reduction can show performance gains for test scores in situations of high intrinsic load. Learners report higher subjective workload in instructional material with high extraneous and high intrinsic load when compared to the same instructional material with reduced extraneous load format. The researchers found that diagrams lowered extraneous load and facilitated learning and understanding for science instructional material only when instructional material is high in element interactivity. In low levels of element interactivity, the instructional format did not impact performance, and it can be inferred that cognitive load is at a manageable state.

Carlson performed two experiments in order to derive these CLT concepts. In the first, 24 ungraded high school students were asked to perform a molecular modeling task. Half of the students were provided instructional materials in a text version, the other half in a diagrammatic version. Each of the students was first asked to perform a task with a low level of element interactivity, or low intrinsic load task, as described in their instructional materials. The students were then asked to perform a high intrinsic load task. The low intrinsic load task involved modeling 10 basic molecules, each modeling involving 3-5 interacting elements, as described in Appendix D. The high intrinsic load task involved modeling 2 complex molecules, which had 16 and 19 interacting elements. For these tasks, times and subjective workload ratings were collected.

The second experiment was conducted with 28 male, high school students. These students were asked to study instructional material for naming prefix and suffixes in carbon compounds. The prefixes had low intrinsic load (2 interacting elements), while the suffixes have high intrinsic load (6-8 interacting elements), as detailed in Appendix E. After studying the instructional materials, the students were given two tests. In the first test the students were allowed access to the instructional material; in the second test the students were not permitted to use the instructional materials. As in the first experiment, the students were assigned either a text or a diagrammatic condition for studying their instructional materials.

The objectives for the first session for the experiment conducted in the laboratory included a replication of the two aforementioned experiments defined by Carlson, albeit in the RoboBooks platform, as opposed to a paper-pencil medium. Additionally, to test applicability of these findings in additional science instruction, an extension of these experiments created with the inclusion of a third task. This task was selected for its relevance to the chemistry classroom from which the initial exploratory analysis was conducted. The task included a lower intrinsic load component, identifying types of chemical reactions (4 or 5 interacting elements), and a higher intrinsic load component, balancing equations (10 - 16 interacting elements), as detailed in Appendix F.

4.3.2 Methods.

4.3.2.1 Task descriptions.

The participants were asked to complete the three tasks (molecular modeling, naming compounds, and chemical equations) successively in the RoboBooks environment. The first task was comprised of two subtasks, while Tasks 2 & 3 were comprised of four subtasks, in the order shown in Table 4.1. The first two subtasks comprised of learning instructional material with low element interactivity (Subtask 1), which will forthwith be referred to as "Low I" to represent low intrinsic load. Similarly learning instructional material with high element interactivity (Subtask 2) will be referred to as "High I". The Test subtasks, described as "open" or "closed" book tests, represent a first test where the instructional materials are allowed, and a subsequent test where instructional materials are not available.

Table 4.1. Task and corresponding subtask descriptions

	Subtask 1 (Low I)	Subtask 2 (High I)	Subtask 3	Subtask 4
Task 1: Molecular Modeling	Model Simple Molecules	Model Complex Molecules	N/a	N/a
Task 2: Naming Compounds	Learn Prefixes	Learn Suffixes	Test 1: "Open Book"	Test 2: "Closed Book"
Task 3: Chemical Equations	Learn Chemical Reaction Types	Learn to Balance Equations	Test 3: "Open Book"	Test 4: "Closed Book"

Preceding each primary task, the participants were provided a pre-task explanation page that explained the components of the task, in congruence with Carlson's methodology. This explanation would describe the task in terms of subtasks and describe any particular components of the task that would otherwise not be included in the instructional material. This includes descriptions of the modeling pieces or definitions of key words that might be used in the instructions, as shown in the example in Figure 4.3.

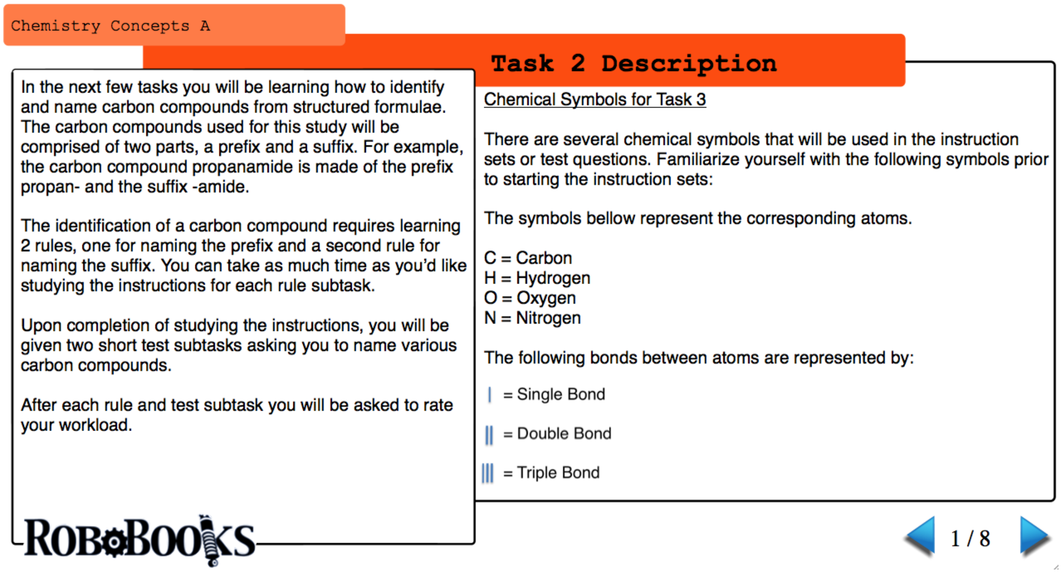


Figure 4.3. *Pre-task Explanation Sheet for Task 2: Naming Compounds*

In each of 4 conditions, the same sequence of tasks, subtasks, and tests were presented to the participants. However, in Session 2, only Tasks 2 (Naming Compounds) and 3 (Chemical Equations) were required of the participants. In each design, all participants were provided the same pre-task instructions, subjective workload rating scales, and test questions. The instructional format for the Low I and High I subtasks, for each of the 3 tasks, varied with the objective for each of the 4 designs.

4.3.2.2 Instructional materials

The instructional materials and test questions were derived from samples provided by Carlson et al. (2003) for each of the instructional tasks and test subcomponents, as well as estimate element interactivity to describe each instructional task, for the Modeling and Naming Tasks (1 & 2). Consequently, the amount of instruction, the content material, and the test questions should be highly congruent with the experiments conducted by Carlson. Notably, there may be some differences in the replication with respect to the exact molecules selected for modeling and the carbon compounds tested. However, each molecule and compound was particularly selected to meet the element interactivity requirements and should be well representative of the original tasks.

The instructional materials for the Equations Task (3) was derived from the RoboBooks curricula developed by the teacher and researcher in Phase I, Session 3. This task consists of identifying 5 types of chemical reactions (lower intrinsic load) and learning 3 main steps to balancing equations (higher intrinsic load). The researcher employed Carlson's method of estimating element interactivity to determine levels of low and high intrinsic load, as detailed in Appendix F. An additional principle to reduce extraneous load was employed for these two learning subtasks. As defined by Sweller et al. (1998), worked examples reduce extraneous cognitive load in a similar manner to a diagram. As the chemical equation tasks require understanding and application of mathematical calculations, a worked example was employed to create the low extraneous load condition for this task.

4.3.2.3 Content tests

Two tests followed the completion of each of the learning instructional material subtasks presented in the Naming and Equations Tasks (2 & 3). These tests assessed the information presented in the preceding two subtasks. In the first test, access to the instructional materials was provided via buttons at the bottom of the page, as shown in Figure 4.4 and the resulting pop-up shown in Figure 4.5. In the second test, the participants were not provided access to the instructional materials.

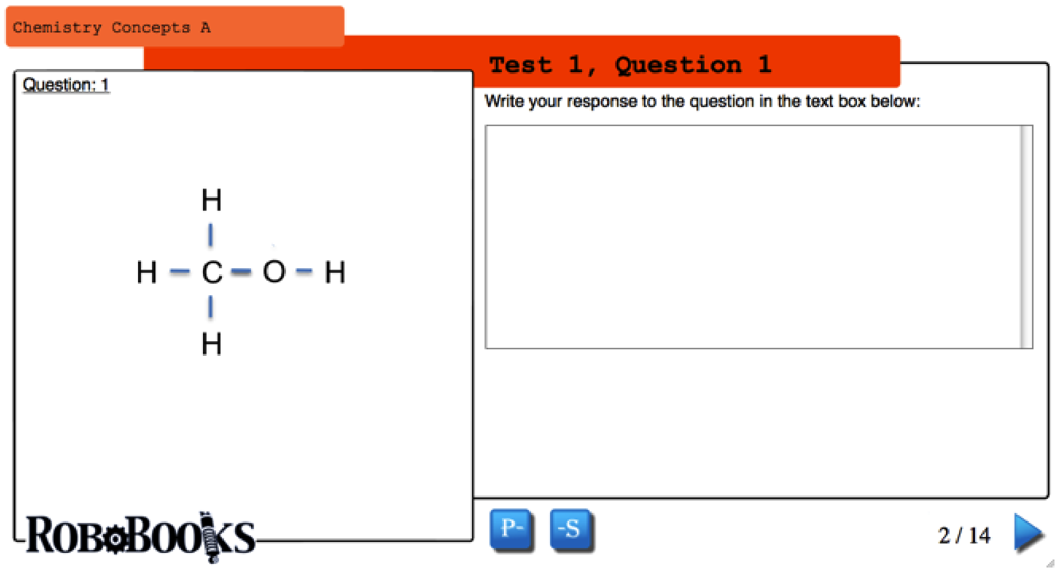


Figure 4.4. *Test 1 with instructional material buttons example screenshot*

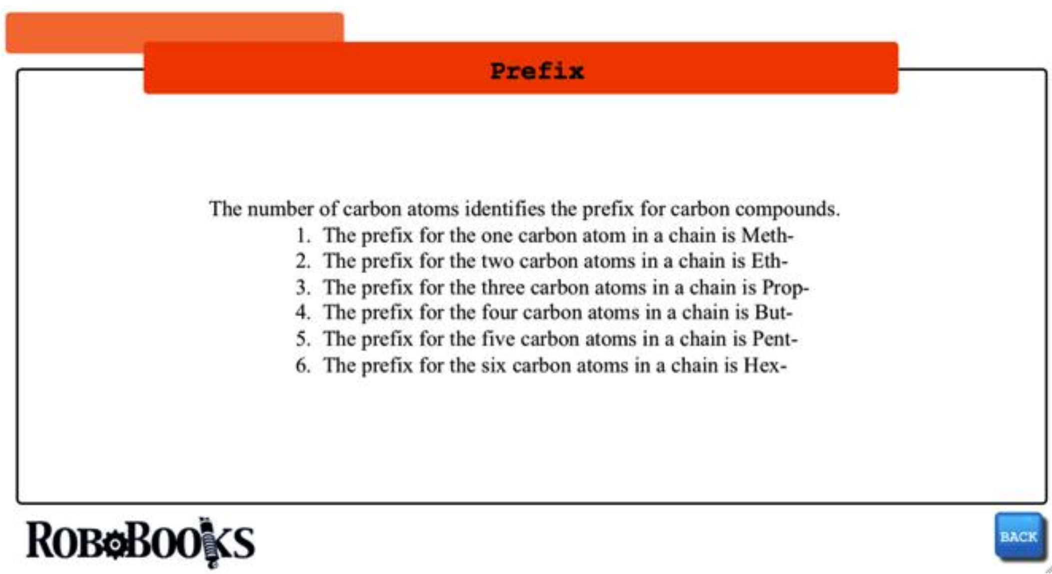


Figure 4.5. *Test 2 "Prefix" instructional materials example screenshot*

4.3.2.4 Interface.

4.3.2.4.1 Condition A: "High E"

The first design (A) was developed to depict the text instructional format as described by Carlson et al. (2003). This design included each of the tasks and subtasks, as described earlier in Table 4.1, with only the text instructions. This design exhibits varying levels of intrinsic load, but a fixed high level of extraneous load (van Merrienboer & Ayres, 2005; Sweller, van Merrienboer, & Paas, 1998). The design will be referred to as "High E" to represent the high level of extraneous load. Figure 4.6 exemplifies the text instructional materials for the low intrinsic subtask (simple models) for the Modeling Task. Figure 4.7 is provided to compare this to the high intrinsic subtask (complex models).

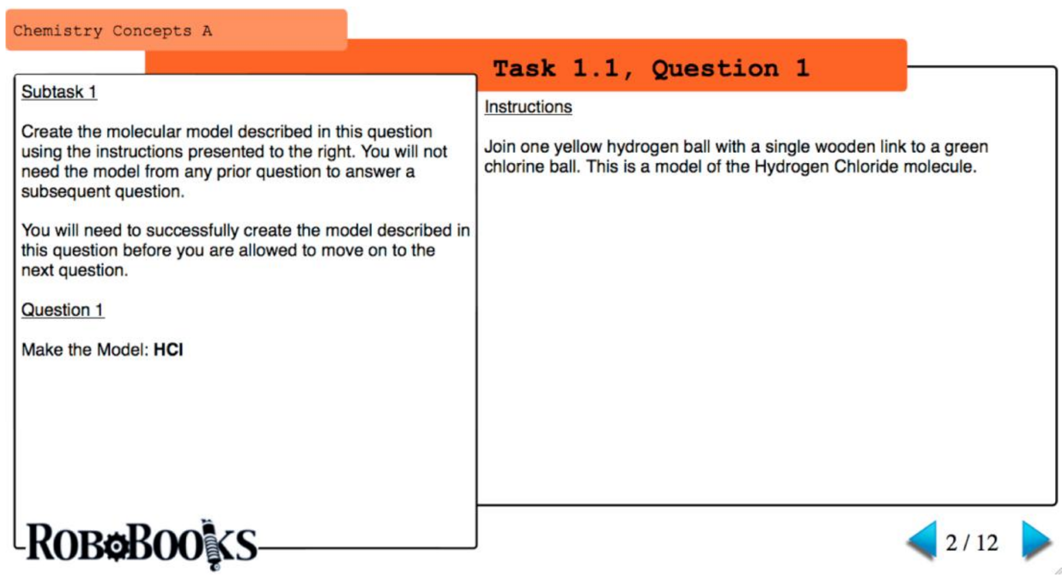


Figure 4.6. *High E Design: Modeling Task (Low I Subtask) example screenshot*

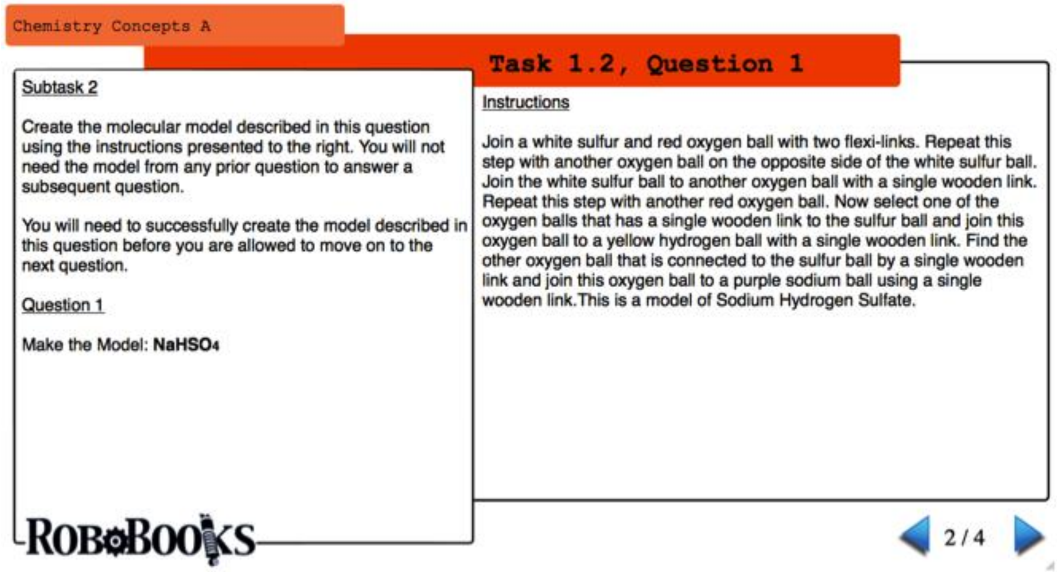


Figure 4.7. *High E Design: Modeling Task, (High I Subtask) example screenshot*

4.3.2.4.1 Condition B

The second design was similarly developed to show the diagrammatic instructional format for Tasks 1 & 2 and the worked example instructional format for Task 3 in a digital medium. This design included each of the tasks and subtasks, as previously described in Table 4.1, with diagrammatic or worked example instructions. This design also has varying levels of intrinsic load, but a fixed level of extraneous load. Diagrams and worked examples are forms of extraneous load reduction; as this design employs extraneous load reduction it will be referred to as “Low E” (van Merrienboer & Ayres, 2005; Sweller, van Merrienboer, & Paas, 1998). Figure 4.8 exemplifies the diagrammatic instructional materials for the low intrinsic subtask (prefix) for the Naming Compound task. Figure 4.9 compares this to a high intrinsic (suffix) subtask.

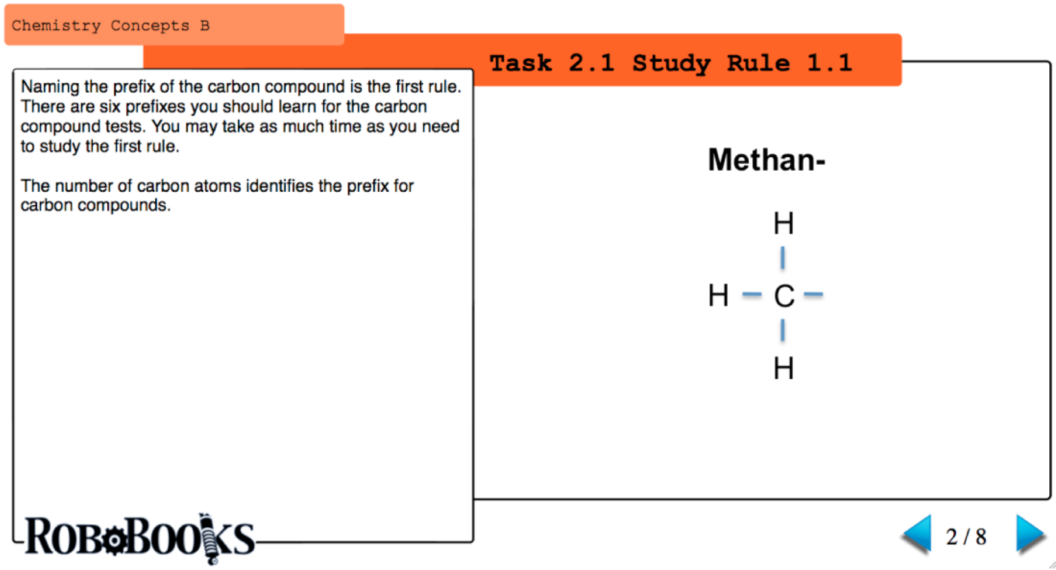


Figure 4.8. *Low E Design: Naming Task (Low I Subtask) example screenshot*

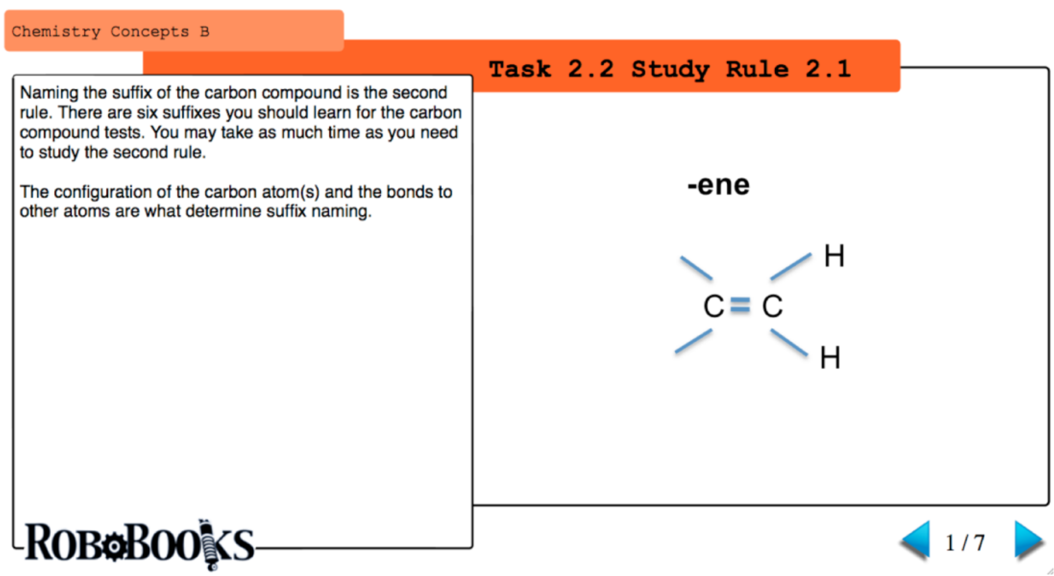


Figure 4.9. *Low E Design: Naming Task, (High I Subtask) example screenshot*

4.3.3 Subject information.

The participants were 20 students from Tufts University. Of the 20 subjects, 8 were male and 12 were female. Participant ages ranged from 19-35, and 6 of the participants were graduate students. The participants had varying amounts chemistry background and were categorized into novice and expert categories, accordingly, though performance means were aggregated. The number of experts and novices in each condition was balanced. 10 of the students were randomly assigned the "High E" Design condition and 10 were assigned the "Low E" Design condition.

4.3.4 Task 1: Molecular modeling.

4.3.4.1 Results.

The measures of interest in Task 1 were task time and subjective workload rating (SWR). (Note: as these subtasks were not followed by a subsequent content test, test score is not applicable for Task 1). The high extraneous load instructional format in the High E Design was compared to the low extraneous instructional format in Low E Design, as the between subject factors. Subtask 1 (Low I), the task with low intrinsic load, was compared to the high intrinsic load Subtask 2 (High I), as the within subject factors. Task time was compared in the between subject variable. SWR was compared in the between and within subject variables.

Table 4.2 shows mean seconds to task completion and average SWR (M) for each of the conditions, with associated standard deviations (SD). It should be noted that the direction of the means show a reduction in time to task and workload between High E and Low E Designs for both subtasks. Further comparisons show an increase in SWR between Low I and High I subtasks.

Table 4.2. Means and standard deviations for task time and SWR in Task 1

Group	Max	High E		Low E	
		M	SD	M	SD
Subtask 1: Simple Model					
Task Time	551	340.7	103.9	263.0	89.9
SWR	7	2.5	0.7	1.7	0.5
Subtask 2: Complex Model					
Task Time	444	295.4	81.7	212.1	60.1
SWR	7	3.7	1.0	2.4	1.1

Note. n=10. Max = maximum.

Task time data was subjected to a 2 (instructional format [High E] vs. [Low E]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I]) ANOVA, with repeated measures on the second factor. There was a significant main effect for the design, $F(1, 18) = 293.1$, $MSE = 3086913.6$, indicating that the diagram group (Low E) overall took less time than the text group (High E). (Note: the 0.05 level of significance is used throughout this paper). Results also showed a significant main effect for the subtask, $F(1, 18) = 5.7$, $MSE = 23136.1$, suggesting that overall Subtask 2 (High I) took longer to complete than Subtask 1 (Low I).

Subjective workload data was also analyzed in a 2 (instructional format [High E] vs. [Low E]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I]) ANOVA. Results also showed a significant main effect for the design, $F (1, 18) = 174.9$, $MSE = 255.0$, indicating that overall the diagrammatic group (Low E) reported less mental workload than the text based group (High E). Results showed a significant main effect for the subtask, $F(1, 18) = 33.5$, $MSE = 9.0$, suggesting that overall subjects reported higher mental workload for Subtask 2 (High I). There was also a significant format x level of element interactivity interaction,

$F(1, 18) = 9.1$, $MSE = 13.2$, suggesting that in high intrinsic load subtask (High I), the design of instructional material impacted the workload.

4.3.4.2 Discussion.

The findings for Task 1 indicate that diagrammatic instructional format, as presented in the Low E Design, resulted in the faster construction of high element interactive molecular modeling (High I). Overall, learners required less time to construct the high intrinsic load models presented in a diagrammatic form (Low E) and reported them to have lesser mental workload than the text based counterpart (High E). These findings were congruent with Carlson et al. (2003) and further support the following claims: (1) when compared to low element interactivity, high intrinsic load leads to decreased performance and higher subjective workload, (2) extraneous load reduction can promote performance gains and reduction in workload for tasks with high element interactivity. Notably, these findings were maintained when the media shifted from paper-and-pencil to digital workbook tasks.

Though mean differences were less distinguishable for Subtask 1, one finding by Carlson et al. (2003) was not supported in this experiment. There was no significant design x task interaction for task times, which Carlson supported the claim that greater improvement was found in the diagrammatic format (Low E) for Subtask 2 (High I) than for Subtask 1 (Low I). This may be due to a slightly smaller sample size in this experiment (n=10), as the means do trend in a similar manner to the Carlson data (n=12).

4.3.5 Task 2: Naming compounds.

4.3.5.1 Results.

The measures of interest in the Naming Task (2) were test scores, task time, and SWR, as the two learning subtasks were followed by subsequent content tests. The high extraneous load instructional format in the High E design was compared to the low extraneous load instructional format in the Low E design as the between subject factors. Subtask 1 (Low I), the task with low intrinsic load, data was compared to the Subtask 2 (High I), high intrinsic load, data as the within subject factors.

Table 4.3 shows the average test scores, mean seconds to task completion, and average SWR (M) for each of the conditions, with associated standard deviations (SD). Test 1 mean scores for both subtasks and Test 2 mean scores for the first subtask appear to decrease between High E and Low E designs; however, Test 2 mean scores increase between High E and Low E designs. It should be noted that the direction of the means show a reduction in time to task between High E and Low E designs for both subtasks. The mean SWR was relatively the same for Subtask 1 between High E and Low E, and higher in High E for Subtask 2. Further comparisons also show an increase in task time and SWR, and a decrease in scores, between Subtask 1(low intrinsic load) and Subtask 2 (high intrinsic load) in both designs.

Table 4.3. Means and standard deviations for scores, task time, & SWR in Task 2

Group	Max	High E		Low E	
		M	SD	M	SD
Subtask 1: Learn Prefix					
Test 1 Score	12	11.9	0.3	10.9	1.2
Test 2 Score	6	5.4	0.7	5.3	1.3
Task Time	217	71.8	56.8	60.0	29.1
SWR	7.0	3.5	0.9	3.6	1.0
Subtask 2: Learn Suffix					
Test 1 Score	12	11.0	1.1	10.7	1.5
Test 2 Score	6	3.1	1.9	4.5	1.7
Task Time	651	221.0	179.7	88.1	59.7
SWR	7	5.3	0.8	4.6	1.0

Note. n=10. Max = maximum.

Test score data was subjected to a 2 (instructional format [High E] vs. [Low E]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I]) ANOVA, with repeated measures on the subtask. For Test 1, no significance was found in either the within or between subject conditions. Test 2 results showed a significant main effect for the subtask $F(1, 18) = 21.8$, $MSE = 24.0$ and a significant design x subtask interaction $F(1, 18) = 5.1$, $MSE = 5.6$, showing a score difference between tasks and suggesting that scores for Low E were greater than High E for high intrinsic load conditions.

Task time data was subjected to a 2 (instructional format [High E] vs. [Low E]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I]) ANOVA, with repeated measures on the subtask. Results showed a significant main effect for the design, 4.6, $MSE = 52345.2$, indicating that the diagram group (Low E) overall took less time than the text group (High E). There was also a significant main effect for the subtask, $F(1, 18) = 9.1$, $MSE = 78588.2$, indicating that there were time differences between the learning prefixes and suffixes. There was an additional significant design x subtask interaction, $F(1, 18) = 4.2$, $MSE = 36663.0$, suggesting that in tasks of low extraneous load (Low E) and high intrinsic load (Subtask 2), the performance time decreased.

Subjective workload data was also analyzed in a 2 (instructional format [High E] vs. [Low E]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I]) ANOVA, with repeated measures on the subtask. Results showed a significant main effect for the design, $F (1, 18) = 246.0$, $MSE = 656.1$, indicating that, overall, the diagrammatic group (High E) reported less mental workload than the text based group (Low E). Results also showed a significant main effect for the subtask, $F(1, 18) = 29.9$, $MSE = 19.6$, suggesting that overall subjects reported higher mental workload for subtask 2 (High I).

4.3.5.2 Discussion.

The data analysis shows that learners scored higher in the prefix portion of Test 2 than the suffix portion, and suggests the diagrammatic group performed higher in the high intrinsic load condition than the text-based group. The findings for Task 2 also indicate that diagrammatic instructional format, as presented in Low E, had faster task time performance and decreased workload than High E. The findings also include a decrease in task time for the subtask with high element interactivity when presented in a diagrammatic form. Furthermore, learners reported the subtask to learn prefixes to have lesser cognitive workload than the task to learn suffixes. These findings support the experimental findings of Carlson, replicated in the digital environment.

Carlson found, however, some differences in test scores. While this experiment showed no performance differences in Test 1, Carlson found significance in this measure for the instructional format and format x intrinsic load level interaction, but not in the subtask comparison. Additionally, in Test 2, Carlson did not find a main effect for instructional format or subtask x design interaction, but did find a main effect for subtask. In this experiment, on the other hand, significance was not found in Test 1 scores, but in Test 2 scores, for the subtask and subtask x design interaction.

In Test 1, participants were provided access to the instructional materials, yet for the second test were asked to recall prefix without the instructional materials. By design, working memory capacity is being tested with the removal of the instructional material in Test 2. Alternatively, in Test 1, with provision of the instructional materials, it would be possible for all participants to receive a perfect score. As such, it can be inferred variances in cognitive capacity be greater and more indicative of individual capabilities pertaining in the second test. Consequently, as overall Test 2 scores improved in the Low E condition, with an additional significant interaction effect for the high intrinsic load conditions, these results support the hypotheses.

4.3.6 Task 3: Chemical equations.

4.3.6.1 Results

Task 3 measures were task time, SWR, and test scores, as the two learning subtasks were followed by subsequent content tests, similarly to Task 2. Table 4.4 shows the average test scores, mean seconds to task completion, and average

SWR (M) for each of the conditions, with associated standard deviations (SD). It should be noted that the direction of the means show an improvement in all test scores when comparing the high extraneous load condition to the low extraneous load condition, except for Test 4 score, Subtask 2 and a reduction in time to task between High E and Low E for both subtasks. SWR was lower for Low E in both subtasks. All measures trended such that scores were higher, task times were shorter, and SWR was lower in Subtask 1(low intrinsic load) when compared to Subtask 2 (high intrinsic load) in both designs.

Table 4.4. Means and standard deviations for scores, task time, & SWR in Task 3

		High E		Low E	
Group	Max	M	SD	M	SD
Subtask 1: ID Reactions					
Test 3 Score	10	9.5	1.3	9.7	1.0
Test 4 Score	5	4.5	0.5	4.9	0.3
Task Time	250	93.6	74.5	50.0	21.1
SWR	7	4.1	1.2	3.2	0.79
Subtask 2: Balance Equations					
Test 3 Score	10	7.3	3.0	7.8	1.6
Test 4 Score	5	4.1	0.9	3.9	0.7
Task Time	488	234.1	191.0	185.8	101.5
SWR	7	4.8	0.9	4.3	1.2

Note. n=10. Max = maximum.

For each test, the score data was subjected to a 2 (instructional format [High E] vs. [Low E]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I]) ANOVA, with repeated measures on the subtask. For Test 3, results showed a significant main effect for the subtask $F(1, 18) = 635.7$, $MSE = 2975.6$, indicating an overall score difference between the two tasks. Test 4 results showed a significant main effect for the subtask $F(1, 18) = 16.9$, $MSE = 4.9$, also showing an overall score difference between levels of element interactivity.

Task time data from Task 3 was subjected to a 2 (instructional format [High E] vs. [Low E]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I]) ANOVA, with repeated measures on the subtask. Results showed a significant main effect for the design, $F(1, 18) = 38.7$, $MSE = 793830.6$, indicating that the worked example group (Low E) overall took less time than the text group (High E). There was also a significant main effect for the subtask, $F(1, 18) = 32.4$, $MSE = 190854.2$, suggesting that there were time differences between learning to identify chemical reaction types and learning how balance equations.

Subjective workload data was also analyzed in 2 (instructional format [High E] vs. [Low E]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I]) ANOVA. Results showed a significant main effect for the design, $F (1, 18) = 216.9$, $MSE = 632.0$, indicating that, overall, the worked example group (Low E) reported less mental workload than the text based group (High E). Results also showed a significant main effect for the subtask, $F(1, 18) = 12.9$, $MSE = 7.2$, suggesting that overall subjects reported higher mental workload for subtask 2 (High I) than for subtask 1 (Low I).

4.3.6.2 Discussion.

The data analysis suggests that learners scored higher for learning subtasks with lower levels of element interactivity when compared to higher levels, for both tests. While not statistically significant, the means for Test 3 possibly suggests a higher performance in the worked example group (Low E) condition for the second subtask. While this is not replicated in Test 4, it can be reasoned that the nature of the test tasks for Task 3, which requires mathematical logic, may have a learning effect. This could be occurring due to possible skill development by balancing the 10 equations in Test 3. Hence, after practice in Test 3, which requires initial application of the instructional material, subjects may reach a point of conformity in the second test set of questions. The evidence of extraneous load reduction on performance in Design B is, notably, supported by reduction in overall task time and a lower reported mental workload for this condition. While there were no interaction effects for any of the measures, the means trend in a manner congruent with CLT for Test 3 scores, task times, and SWR.

4.3.7 General discussion.

Session 1 objectives included testing CLT in digital media and extending the theory into new subject content area. All tasks showed evidence that as overall cognitive complexity increased between Subtasks 1 and 2, participant's time to task increased, they associated a higher cognitive workload with higher complexity, and their scores accordingly decreased. Participants also benefitted from the diagrammatic and worked example condition (Low E) as shown in task time improvements. Overall, they associated higher workload to the High E over the Low E condition. And, for Tests 2 and 3, participants overall scored higher in the Low E design. These findings supported Hypothesis 1: as intrinsic load increases (H1a), performance decreases (H1b), and cognitive load increases (H1c).

Several interactions were found that support Hypothesis 2. Evidence that test scores improve when instructions are presented in a diagrammatic form for material with high element interactivity was found in the design x subtask interaction for Test 2 scores (H2a). Similarly, task times in Task 2 showed faster performance with diagrams when presented more complex material (H2b). Also, SWR is lower for high intrinsic, low extraneous load in Task 1 (H2c).

While Hypothesis 2 was not determined significantly for all measures in every task, the means for these measures also trend in the right direction. A Power Analysis determined that for these measures, at Power=0.8, an n=30 would be sufficient to see significance for most of the subtask x design interactions. It should also be noted that either the small sample size or the subject background may drive a high variance in the between condition.

Figure 4.10 depicts the mean differences in SWR for Task 1. This comparison shows diverging mental workload in the high intrinsic (I) load condition, and exemplifies Hypotheses 1 and 2. Satisfying Hypothesis 1c, workload is perceived higher with higher levels of intrinsic load. Satisfying Hypothesis 2c, workload is perceived higher in the High E design, with high extraneous load, than the Low E design, and at high intrinsic load, the difference in workload is greater than in the low intrinsic load subtask.

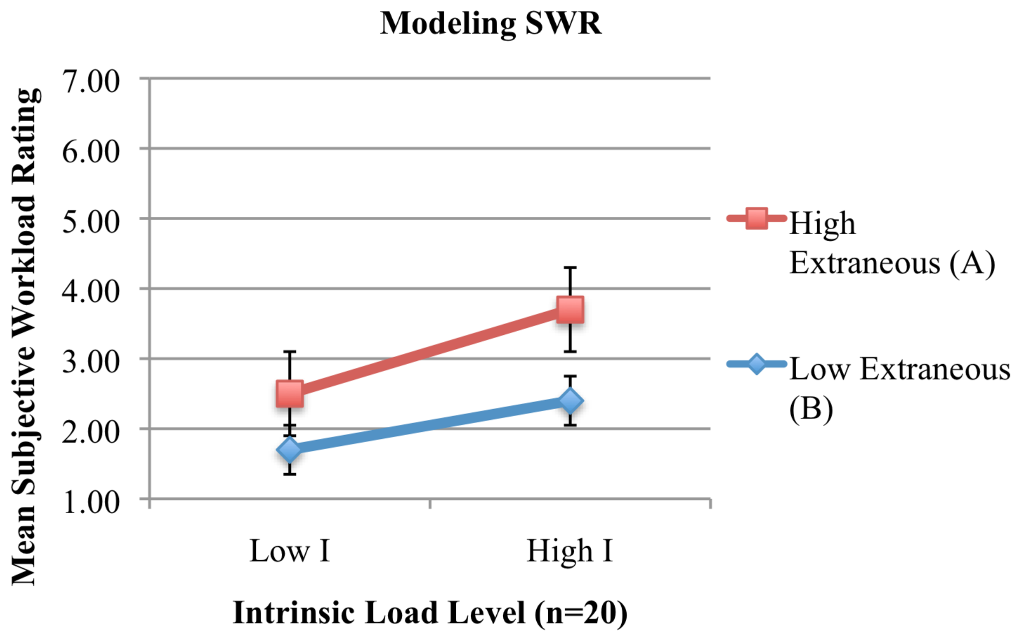


Figure 4.10. *Subjective workload rating mean comparisons for Task 1*

Hypothesis 3 suggests that UDL tools will affect the dependent measures such that behavior should fall between the high extraneous and low extraneous instructional format designs. The more the UDL tool conditions (Designs C and D) progress towards the low extraneous load condition, evidence of cognitive load reduction by the UDL tools may be evident. This will be further explored in the following section.

4.4 Session 2: CLT and UDLs

4.4.1 Objectives.

The main objective of session 2 was to explore UDL tool presentation within the context of CLT. The experiment was designed to use the same tasks and subtasks from the previous session to compare UDL presentation in designs UDL 1 and UDL 2 similarly to condition Low E when compared against high extraneous load in condition High E. In order to do so, the information presentation of the instructional subtasks (1 & 2) in the Compound Naming and Chemical Equation main tasks was modified by the inclusion concept maps and highlighting. It should be noted that the level of element interactivity remains the same between each of the 4 designs.

4.4.2 Methods

4.4.2.1 Task descriptions.

The second and third tasks from session one, as shown in Table 4.5, were replicated in the design in session 2. The Subtasks 1 and 2, where low and high intrinsic load instructional material is presented, were modified to include UDL tool presentation of information. Similarly to session 1, each subtask was preceded by an explanation sheet and followed by a subjective workload rating scale. After the instructional subtasks, two tests were administered within the digital workbook. In the first test, access was provided to the instructional materials via buttons at the bottom of the screen. In the second test, access to instructional materials was restricted by removal of the buttons.

Table 4.5. Task and corresponding subtask descriptions

	Subtask 1	Subtask 2	Subtask 3	Subtask 4
Task 2: Naming Compounds	Learn Prefixes	Learn Suffixes	Test 1: “Open Book”	Test 2: “Closed Book”
Task 3: Chemical Equations	Learn Chemical Reaction Types	Learn to Balance Equations	Test 3: “Open Book”	Test 4: “Closed Book”

4.4.2.2 Instructional materials.

The instructional materials for the two session 2 UDL designs were derived from the previous session, Tasks 2 and 3. Task 2 was replicated from

Carlson et al. (2003). However, for UDL 1 Design, a text-only concept map was added to all of the instructional materials, for both the low intrinsic and the high intrinsic load conditions. Alternatively for UDL 2, a symbolic concept map, consisting of letters to represent atoms and single, double, or three lines to represent bond type, was added to the same instructional materials.

Task 3 was derived from classroom curricula, as per session 1. In UDL Design 1, highlighting was added to the text, or high extraneous load, version of the instructional material (High E). Alternatively, for the UDL 2 design, highlighting was added to the low extraneous load version of the instructional material, which included worked examples.

4.4.2.3 Interface.

4.4.2.3.1 Condition C: UDL 1.

The UDL 1 design was developed to trial the concept map and highlight tools in varying levels of intrinsic load. The design included the Naming Compounds Task and all associated subtasks. Figures 4.10 and 4.11 show the application of the concept map UDL tool for the low intrinsic subtask (learn prefixes) and high intrinsic subtask (learn suffixes), respectively. In this condition, a text version of the concept map was employed, as depicted in the screenshots.

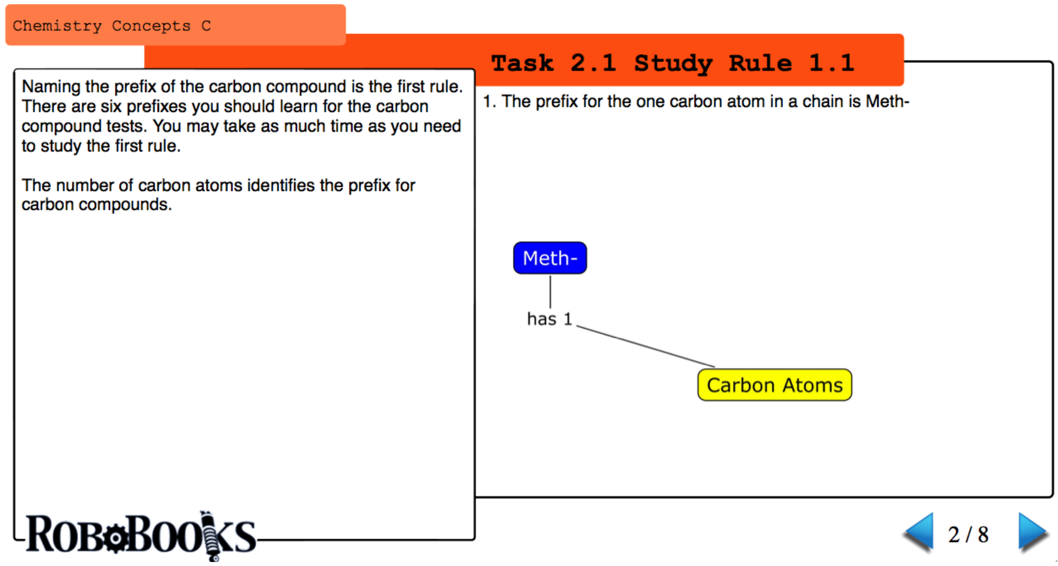


Figure 4.10. *UDL 1 Design: Naming Task 2 (Low I Subtask) example screenshot*

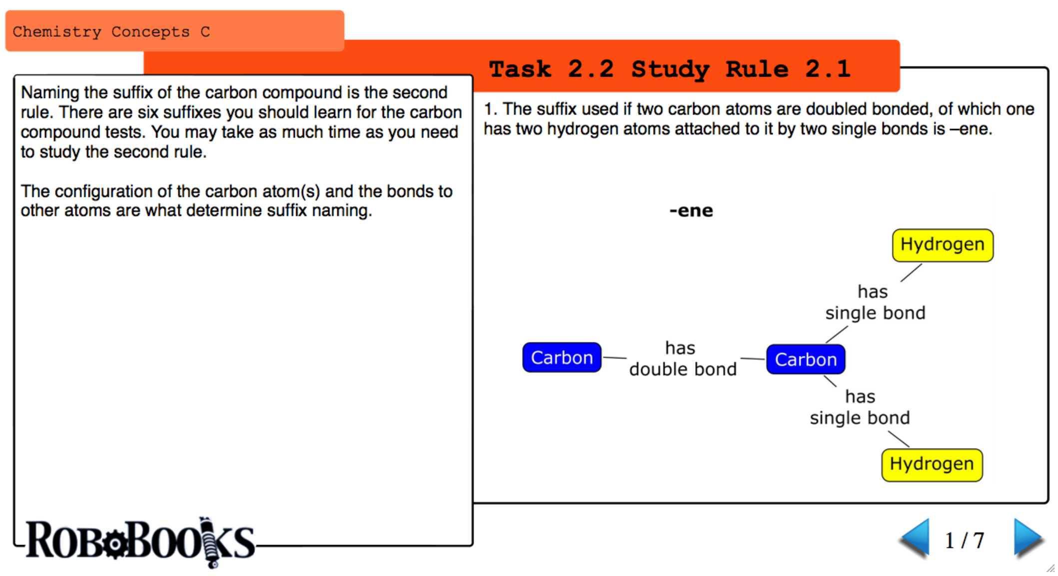


Figure 4.11. *UDL 1 Design: Naming Task 2 (High I Subtask) example screenshot*

The UDL 1 design also included the Chemical Equations task and all associated subtasks. Figures 4.12 and 4.13 show the application of the highlight UDL tool for the low intrinsic subtask (learn prefixes) and high intrinsic subtask (learn suffixes), respectively. Notably, as this task has full text, it should also be considered the high extraneous load base condition.

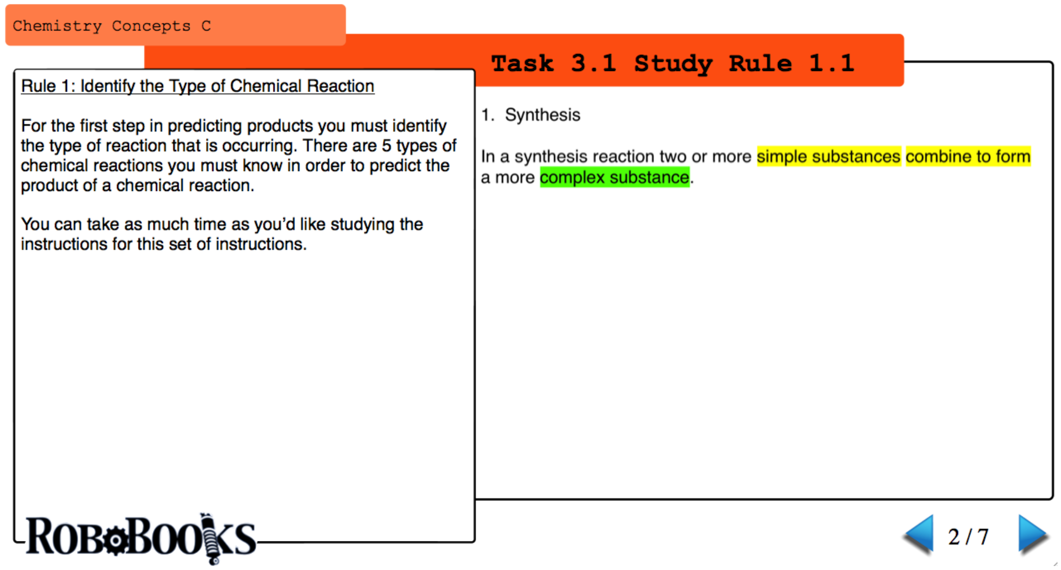


Figure 4.12. *UDL 1 Design: Equations Task 3 (Low I Subtask) example*

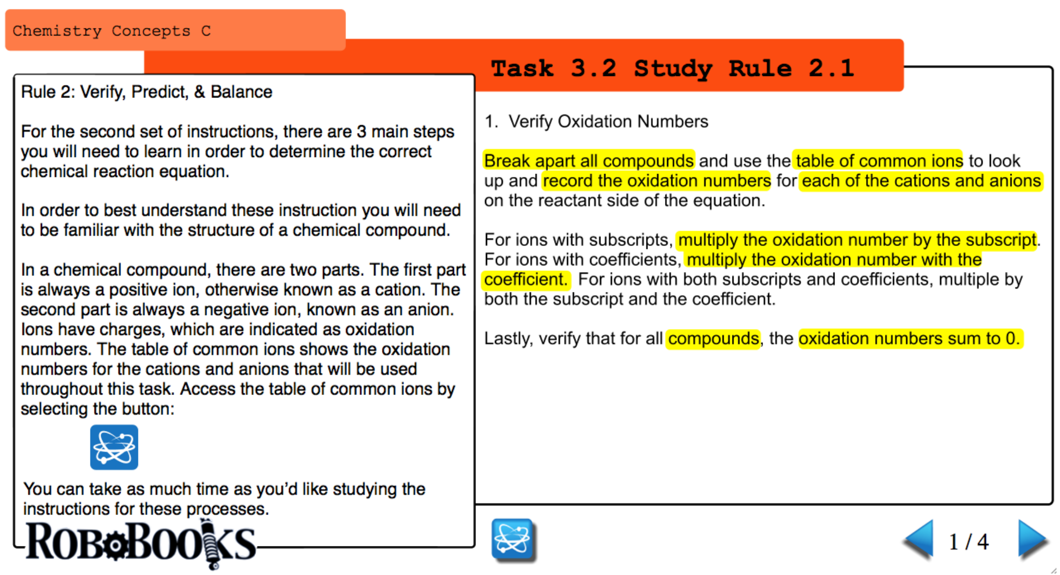


Figure 4.13. *UDL 1 Design: Equations Task 3 (High I Subtask) example*

4.4.2.3.2 Condition D: UDL 2

The UDL 2 design was developed to additionally test the concept map and highlight tools in varying levels of intrinsic load. The design included the Naming Compound Task and all associated subtasks. Figures 4.14 and 4.15 show the application of the concept map UDL tool for the low intrinsic subtask (learn prefixes) and high intrinsic subtask (learn suffixes), respectively. In this condition, a "diagrammatic" version of the concept map was employed.

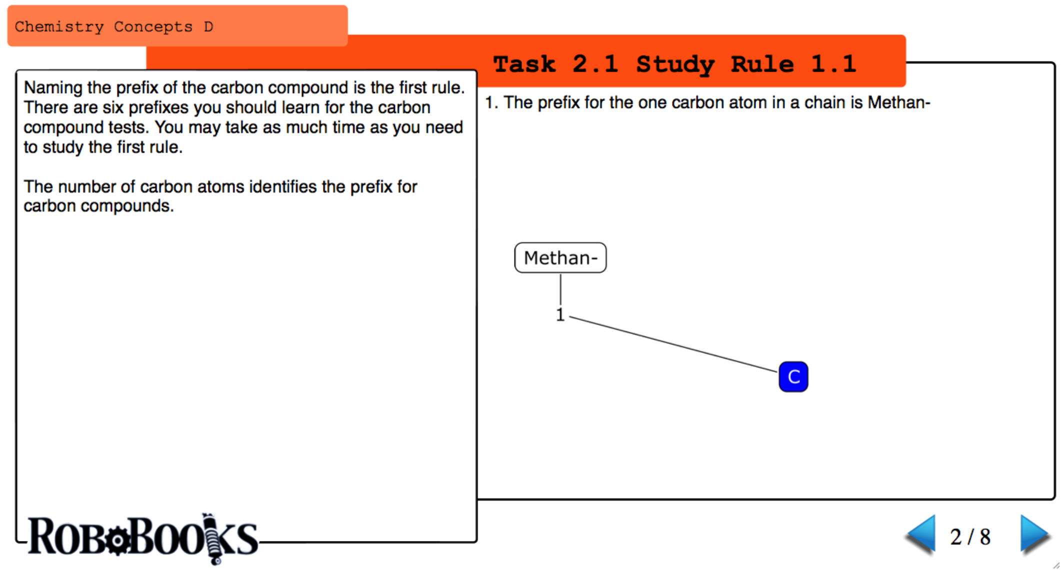


Figure 4.14. *UDL 2 Design: Naming Task 2 (Low I Subtask) example screenshot*

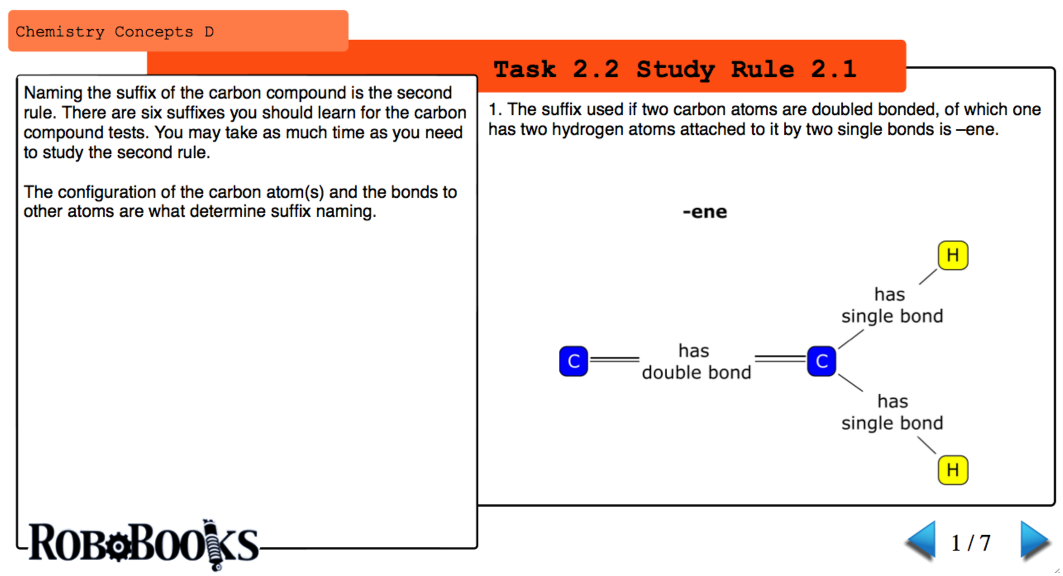


Figure 4.15. *UDL 2 Design: Naming Task 2 (High I Subtask) example screenshot*

The UDL 2 design also included the Chemical Equations Task and all associated subtasks. Figures 4.12 and 4.13 show the application of the highlight

UDL tool for the low intrinsic subtask (learn prefixes) and high intrinsic subtask (learn suffixes), respectively. Notably, as this task additionally has worked examples, it should also be considered a lower extraneous load base condition.

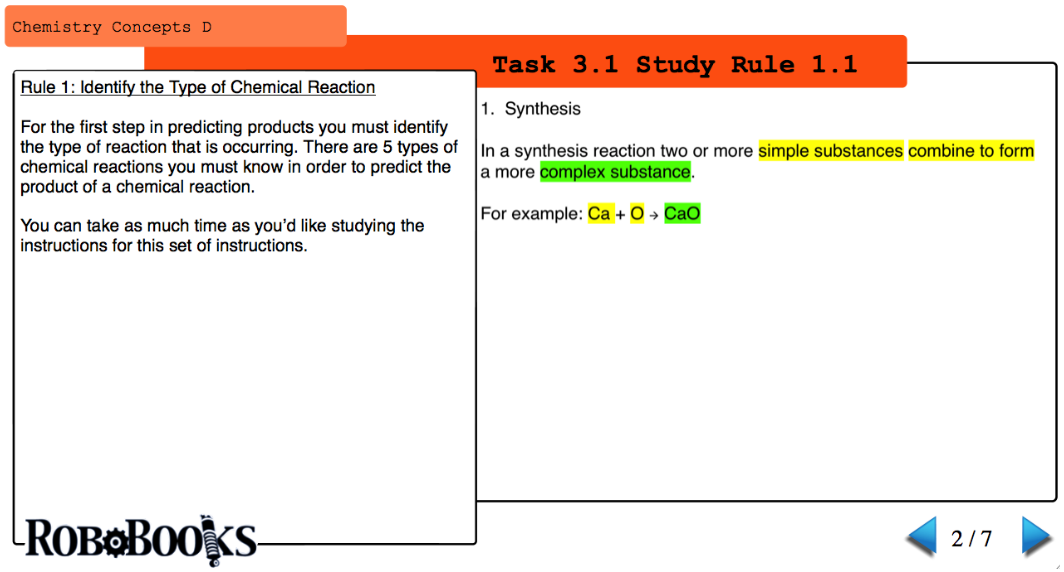


Figure 4.16. *UDL 2 Design: Equation Task 3 (Low I Subtask) example screenshot*

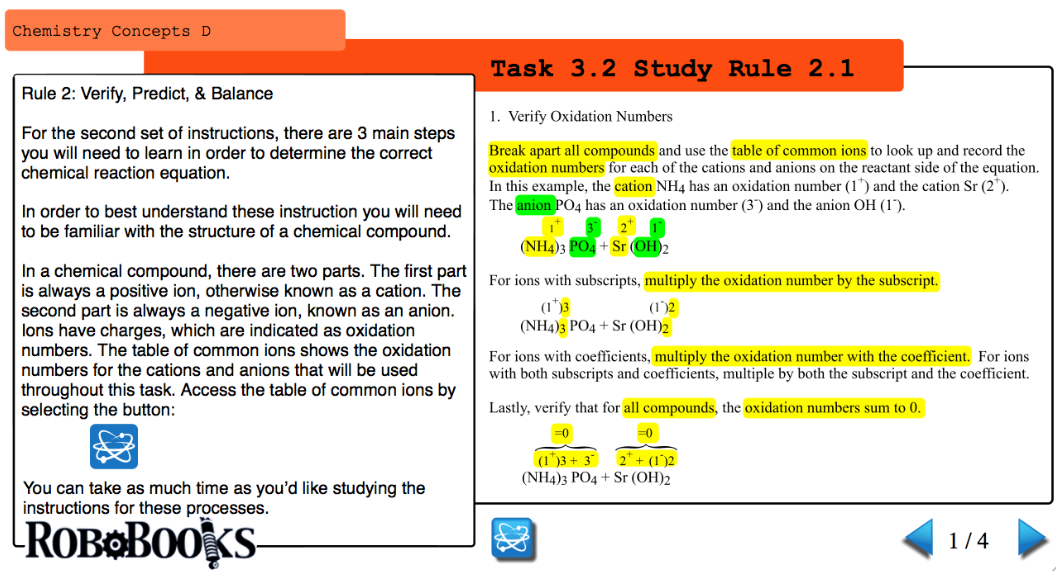


Figure 4.17. *UDL 2 Design: Equation Task 3 (High I) example screenshot*

4.4.3 Subject information.

For this session of the second phase, 20 new student participants from Tufts University were given the UDL 1 and UDL 2 instructional formats. Of the 20 subjects, 9 were male and 11 were female. Participant ages ranged from 18-28, and 7 of the participants were graduate students. The participants had varying amounts chemistry background and were categorized into novice and expert categories, accordingly, though performance means were aggregated. The number of experts and novices in each condition was balanced. 10 of the students were randomly assigned the "UDL 1" Condition and 10 were assigned the "UDL 2" condition.

4.4.4 Task 2: Naming compounds.

4.4.4.1 Results.

The measures of interest in Task 2 were test scores, task time, and SWR, as the two learning subtasks were followed by subsequent content tests. The instructional format in UDL 1 design (text-based concept maps and highlighting) and the instructional format in UDL 2 design (diagram concept maps and highlighting) was compared to instructional format in High E (high extraneous load) and Low E (low extraneous load), as the between subject factors. Subtask 1 (low intrinsic load) data was compared to the Subtask 2 (high intrinsic load) data as the within subject factors.

Table 4.6 shows the average test scores, mean seconds to task completion, and average SWR (M) for each of the 4 conditions, with associated standard deviations (SD). Particular attention should be paid to differences in means between High E and Designs UDL 1 and UDL 2, independently. The direction of the means show an improvement in Test 2 scores for UDL1 in Subtask 1 (Low I) and UDL 1 and UDL 2 in Subtask 2 (High I). There is a trend of reduced time to task for both Designs UDL 1 and UDL 2 for learning suffixes. The mean SWR was lower in Designs UDL 1 and UDL 2 for Subtask 1 and 2. Further comparisons also show a decrease in scores, longer task times, and an increase in SWR between Subtask 1 (low intrinsic load) and Subtask 2 (high intrinsic load) for all designs.

Table 4.6. Means and standard deviations for scores, task time, & SWR in Task 2

		High E		Low E		UDL 1		UDL 2	
Group	Max	M	SD	M	SD	M	SD	M	SD
Subtask 1: Learn Prefix									
Test 1 Score	12	11.9	0.3	10.9	1.2	11	1.15	11.3	0.67
Test 2 Score	6	5.4	0.7	5.3	1.3	5.7	0.48	5.3	0.82
Task Time	217	71.8	56.8	60.0	29.1	87.3	46.3	85.4	34.8
SWR	7	3.5	0.9	3.6	1.0	3.1	1.1	3.1	1.1
Subtask 2: Learn Suffix									
Test 1 Score	12	11.0	1.1	10.7	1.5	11.4	0.7	10.7	0.82
Test 2 Score	6	3.1	1.9	4.5	1.7	4	1.41	3.5	1.35
Task Time	651	221.0	179.7	88.1	59.7	167.4	81.5	140.9	63.8
SWR	7	5.3	0.8	4.6	1.0	4.8	1.0	4.8	1.3

Test score data was subjected to a 4 (instructional format [High E] vs. [Low E] vs. [UDL 1] vs. [UDL 2]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I]) ANOVA, with repeated measures on the subtask. For Test 1, no significance was found in either the within or between subject conditions, consistent with Session 1 findings. Test 2 results showed a significant main effect for the subtask $F(1, 36) = 48.16$, $MSE = 54.45$, showing a score difference between high and low intrinsic tasks.

Task time data was subjected to a 4 (instructional format [High E] vs. [Low E] vs. [UDL 1] vs. [UDL 2]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I]) ANOVA, with repeated measures on the subtask. There was a significant main effect for the subtask, $F(1, 36)$ 25.485, $MSE = 122382.013$, suggesting that there were time differences between the learning prefixes and suffixes. There was an additional significant design x subtask interaction, $F(1, 36) = 4.241$, $MSE = 36663.0$, indicating that between tasks of low extraneous load (Low E, UDL 1, and UDL 2) and high intrinsic load (Subtask 2), the performance time decreased.

Subjective workload data was also analyzed in a 4 (instructional format [High E] vs. [Low E] vs. [UDL 1] vs. [UDL 2]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I]) ANOVA, with repeated measures on the subtask. Results showed a significant main effect for the subtask, $F(1, 36) = 86.93$, $MSE = 48.05$. This finding suggests that overall subjects reported higher mental workload for subtask 2 (high intrinsic load).

4.4.4.2 Analysis.

The data analysis shows that learners performed higher, with less workload in tasks that have lower levels of element interactivity than higher levels. While not statistically significant, the means for Test 2 scores suggest a higher performance in the UDL 1 and UDL 2 conditions for the second subtask, consistent with session 1 findings for these scores. Evidence of overall extraneous load reduction in Low E, UDL 1, and UDL 2 is supported by reduction in overall task time and a lower reported mental workload for this condition.

Figure 4.11 compares the task time measure for each of the 4 designs. As predicted, in low levels of intrinsic load, performance should be similar between the 4 designs. In higher intrinsic load conditions, UDL 1 and UDL 2 conditions were faster than the High E condition, which suggests a possible reduction in extraneous load similar to the Low E condition. As such, the means for UDL 1 and UDL 2 show reduced task times when compared to the High E condition.

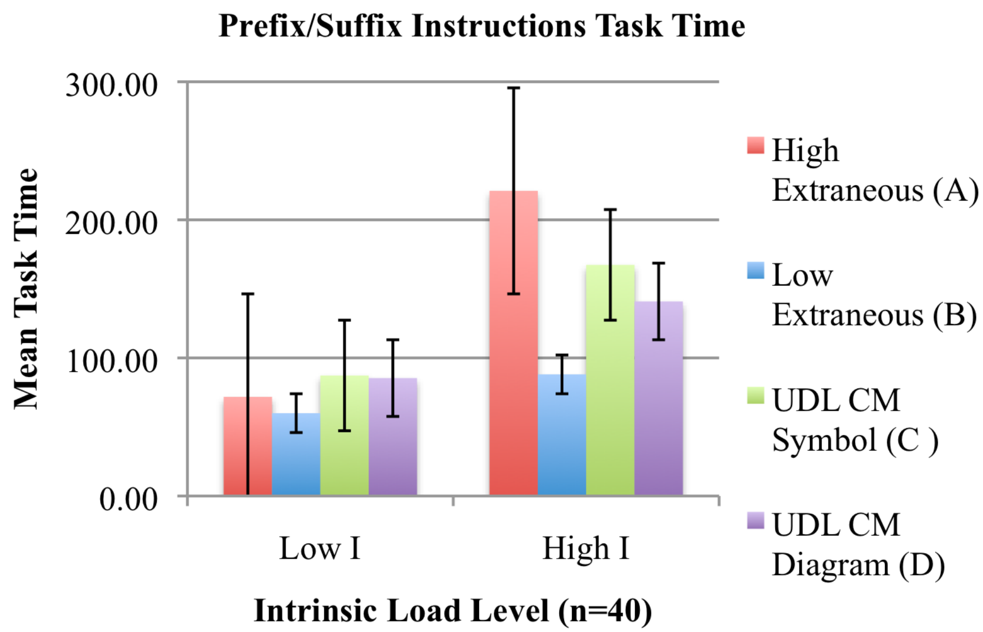


Figure 4.11. *Mean comparison of task times for designs in Task 2*

Figure 4.12 compares the SWR measure for each of the 4 designs. As predicted, in low levels of intrinsic load, performance should be similar between the 4 designs. In higher intrinsic load conditions, UDL 1 and UDL 2 conditions had lower SWR than High E and suggest a possible reduction in extraneous load approaching similarity with Low E.

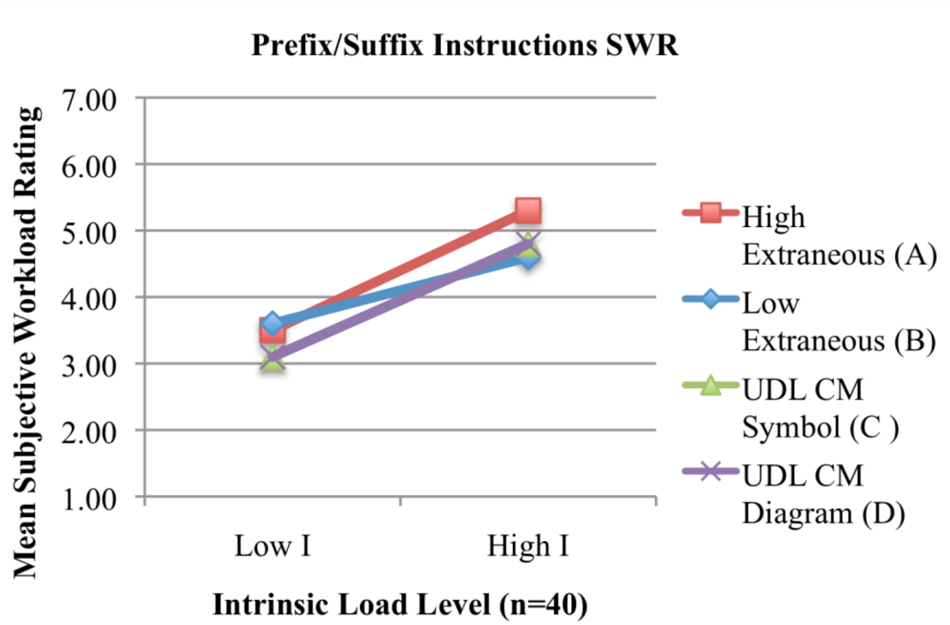


Figure 4.12. *Mean comparison of SWR for designs in Task 2*

4.4.5 Task 3: Chemical equations.

4.4.5.1 Results.

The measures of interest in Task 3 were test scores, task time, and SWR, as the two learning subtasks were followed by subsequent content tests. The instructional format in UDL 1 (text-based concept maps and highlighting) and the instructional format in UDL 2 (diagram concept maps and highlighting) was compared to instructional format in High E (high extraneous load) and Low E (low extraneous load), as the between subject factors. Subtask 1 (low intrinsic load) data was compared to the Subtask 2 (high intrinsic load) data as the within subject factors.

Table 4.7 shows the average test scores, mean seconds to task completion, and average SWR (M) for each of the 4 conditions, with associated standard deviations (SD). Particular attention should be paid to differences in means between High E and the two UDL designs, independently. The direction of the means show an improvement in Test 3 scores for UDL 2 for both subtasks. There is a trend of time reduction to task for both Designs UDL 1 and UDL 2 for learning to ID reactions when compared to the High E condition and a reduction for UDL 2 in the balancing equations subtask. The mean SWR was lower in Designs UDL 1 and UDL 2 for Subtask 1. Further comparisons also show a decrease in scores, longer task times, and an increase in SWR when comparing Subtask 1(low intrinsic load) and Subtask 2 (high intrinsic load) for all designs.

Table 4.7. Means and standard deviations for scores, task time, & SWR in Task 3

		High E		Low E		UDL 1		UDL 2	
Group	Max	M	SD	M	SD	M	SD	M	SD
Subtask 1: ID Reactions									
Test 3 Score	10	9.5	1.3	9.7	1.0	9.1	2.1	10.0	0.8
Test 4 Score	5	4.5	0.5	4.9	0.3	4.6	3.5	4.9	0.3
Task Time	250	93.6	74.5	50.0	21.1	88.1	61.3	69.5	30.5
SWR	7	4.1	1.2	3.2	0.79	3	0.94	2.9	0.99
Subtask 2: Balance Eqts									
Test 3 Score	10	7.3	3.0	7.8	1.6	6.8	2.4	8.3	2.2
Test 4 Score	5	4.1	0.9	3.9	0.7	3.5	1.4	3.9	1.5
Task Time	488	234.1	191.0	185.8	101.5	267.7	106.3	209.9	106.7
SWR	7	4.8	0.9	4.3	1.2	5.5	0.7	5.1	1.2

Test score data was subjected to a 4 (instructional format [High E] vs. [Low E] vs. [UDL 1] vs. [UDL 2]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I]) ANOVA, with repeated measures on the subtask. Test 3 and 4 results showed a significant main effect for the subtask $F(1, 36) = 38.67$, $MSE = 82.01$, and $F(1, 36) = 25.46$, $MSE = 15.31$, respectively, showing lower scores on high intrinsic load tasks.

Task time data was subjected to a 4 (instructional format [High E] vs. [Low E] vs. [UDL 1] vs. [UDL 2]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I])

ANOVA, with repeated measures on the subtask. There was a significant main effect for the subtask, $F(1, 36) = 90.45$, $MSE = 444467.11$ suggesting that there were time differences between the learning prefixes and suffixes subtasks.

Subjective workload data was also analyzed in a 4 (instructional format [High E] vs. [Low E] vs. [UDL 1] vs. [UDL 2]) x 2 (Subtask 1 [Low I] vs. Subtask 2 [High I]) ANOVA, with repeated measures on the subtask. Results showed a significant main effect for the subtask, $F(1, 36) = 100.17$, $MSE = 36.45$, suggesting that overall subjects reported higher mental workload for subtask 2 (high intrinsic load). There was also a design x subtask interaction, $F(1, 36) = 4.076$, $MSE = 1.48$, indicating that extraneous load reductions were considered to have less cognitive load when compared to high extraneous load for subtask of identifying reactions.

4.4.5.2 Analysis.

Again, the data analysis suggests that learners performed higher, with less workload in tasks with lower levels of interactivity than higher levels. While not statistically significant, the means for Test 3 suggest a higher performance in the UDL 2 condition for both subtasks. Evidence of extraneous load reduction in Low E, UDL 1, and UDL 2 is supported by reduction in overall task time and a lower reported mental workload for Subtask 1.

Figure 4.13 compares the task time measure for each of the 4 designs. As predicted, in low levels of intrinsic load, performance should be more similar between the 4 designs. In higher intrinsic load conditions, UDL 1 task time mean was longer than High E, yet UDL 2 was faster than High E. This suggests a possible reduction in extraneous load towards for UDL 2 trending in a similar direction as Low E.

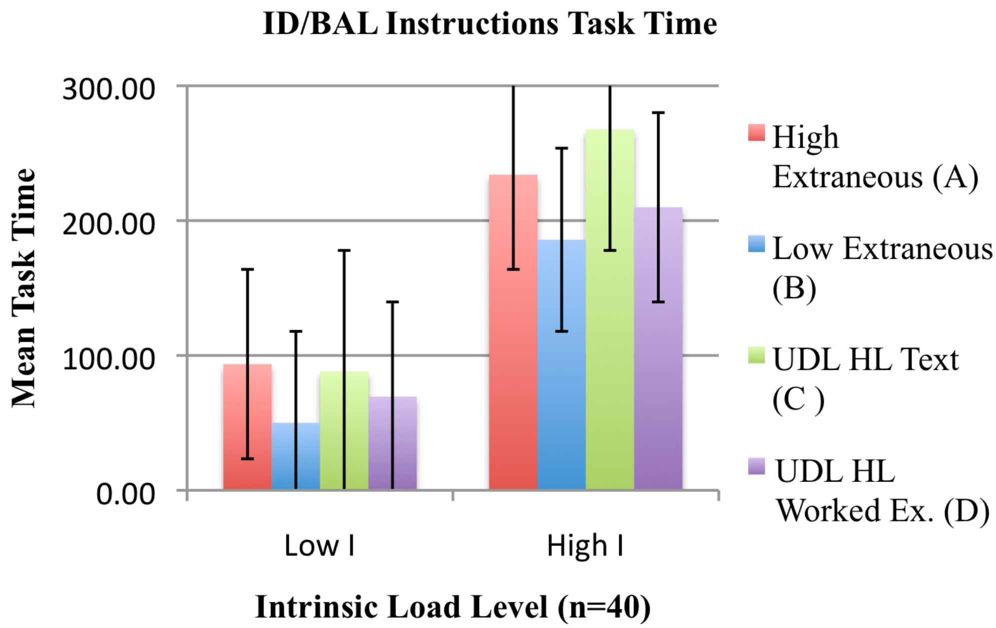


Figure 4.13. *Mean comparison of task times for designs in Task 3*

Figure 4.14 compares the SWR measure for each of the 4 designs. Counter to predictions, reported SWR varied more in lower levels of intrinsic load than in higher. As performance measures in this task were also not indicative of expected results, there may be specific demands of this tasks that are incongruent with the other tasks. Notably, the high standard deviations are likely attributed to the variance of background in the participants, and may represent a difference in behavior between “novice” and “expert” participants, though the number of novice and experts were balanced between conditions.

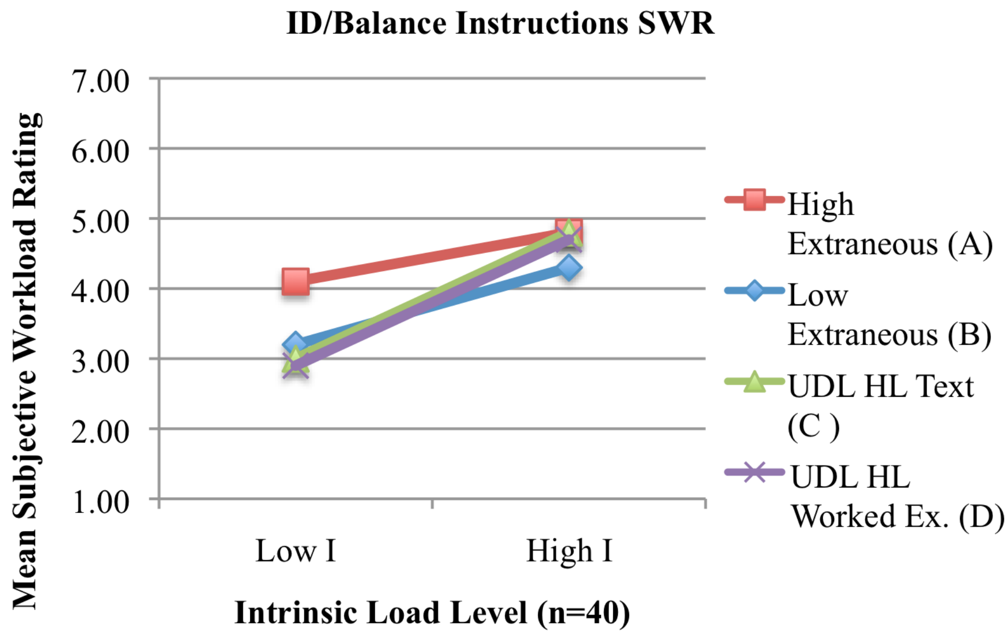


Figure 4.14. *Mean comparison of SWR for designs in Task 3*

4.4.6 General discussion.

Two particular deviations appeared in session 2 findings. Firstly, the employment of UDL tools shows a difference in mean performance for UDL 1 and UDL 2. Moreover, the means for UDL 1 often exceed the High E condition, implying that there could possibly be additional cognitive load imposed by the application of certain UDL tools. Secondly, the performance in Task 3 indicates a divergence from CLT as described by Carlson et al. (2003).

In the first case, it may be that UDL tools have different affordances for the type of instructional format. For example, the highlight UDL may be adding processing time to the task, as shown in mean Task 3 times. Additionally, it appears a diagrammatic concept map may afford gains over a text-based concept map, when comparing mean performance. Further analysis is needed and should focus on differences in performance gains for different UDL types and employment methods.

Furthermore, consider the case of task differences. For example, the expected element interactivity for the identification of chemical reactions (subtask 1) was 4-5 elements. This could actually be considered a medium level of element interactivity, as compared to a low level (2-3 elements). Interestingly, this subtask seemed to behave more like a high intrinsic load condition, which implies that there may be a continuum for intrinsic load, which is not captured in “low” and “high” level definitions. The balance equations instructions have 14-16 interacting elements, which could impose too high a level of cognitive load that performance again converges. The nature of this task also requires sequentially processing of information, for which later elements are dependent on the understanding of previous rule elements. Alternatively, the rules for learning subtasks in Task 2 and the first subtask in Task 3 can be understood, and stored in the memory, independently. Hence, more specific considerations for task definitions may be necessary to determine performance gains.

However, there was evidence to merit the continuation of investigating UDL usage in the CLT context. For example, the repeated significance for the within subject conditions does show strong evidence of successful intrinsic load manipulations. It can be deduced that estimates of element interactivity, as described in Appendices A, B and C, provides a means to measure inherent complexity of instructional material. As such, varying degrees of complexity in subject matter can be assessed in a systematic way. While still unclear if UDL tools provide a similar benefit as established principles in extraneous load reduction, mean performance for both UDL designs in Task 2 and findings from UDL 2 in Task 3 suggest that further exploration with additional subjects may yield promising results. As such, these tools may be best implemented in cases of high intrinsic load for best performance benefits.

The next chapter further explores the use of UDL tools in the classroom. If theory applies to actual context, it can be hypothesized that in higher levels of intrinsic load, UDL tools would aid in cognitive load reduction and consequently be utilized more frequently. As such, UDL tools will be optional for supplementing instructional materials, in order to determine if utility is indeed increased as element interactivity increases.

4.5 Phase II Summary

This lab work explored the concept of cognitive load within varying chemistry tasks. Intrinsic load, or inherent complexity, was studied by looking at learning tasks that have low and higher levels of element interactivity. Extraneous load was studied by varying instructional format, primarily comparing a text only version of instructional material with diagrammatic and worked example formats. This work replicated a study done by Carlson et al. (2003) in the digital platform and extended methodology into a new subject task - identifying and balancing chemical equations. In the RoboBooks environment, results showed that indeed manipulations of intrinsic and extraneous load in the digital environment showed similar results to traditional, fixed media.

In the second session two UDL tools were selected to explore for their possible role within the context of CLT. Here, it was hypothesized that the tools aid in information acquisition by organizing and representing information supporting parallel with the use of concept maps and sequential processing by use of highlight tools that emphasize key components. The tools were applied to high intrinsic load conditions and studied to explore if performance gains were similar to established extraneous load reduction techniques. Results showed that indeed intrinsic load manipulations were maintained and mean trends suggest with increased sample sizes that performance gains can be achieved by application of these UDL tools.

Consequently, this research phase introduced the idea of cognitive load management. The idea being, that use of UDL tools, instructional formatting, and even germane load manipulations are techniques to enable learners to manage cognitive load, increasing access to complex ideas, prevalent particularly in science, technology, engineering, and math context. The following phase looks at this idea in a naturalistic setting, back in the classroom, to explore the relationship between UDL tool utilization in varying levels of cognitive load conditions.

5. UDLs in the Classroom

5.1 Objectives

5.1.1 Phase III description.

Cognitive load theory provides a means to understand task differentials in instructional design with respect to both content and information presentation. As observed in the classroom in Section 3, there may be task dependencies that drive

UDL tool usage and have consequent bearings on cognitive gains. As determined in the lab in Section 4, a learning task is comprised of components of cognitive load, one of which is inherent to the complexity of the task (intrinsic); another, which can be manipulated by means of information presentation (extraneous). This phase details a final classroom session that was conducted in order to explore cognitive load and UDL tools in a naturalistic setting. This phase looks particularly at CLT in curriculum presented in the digital RoboBooks workbook and compares varying levels of cognitive load components to particular navigation and utilization behavior exhibited by students in the classroom.

5.1.2 Research questions.

The research questions explored in the third phase were derived from outcomes determined in the previous phases. As such, the application of UDL tools in the context of CLT in this field study focused on comparing varying levels of intrinsic and extraneous load with UDL tool utilization. In accordance with findings from Phase 2, which indicated that UDL tools may best promote performance gains in conditions of high intrinsic and extraneous load, the following hypotheses were investigated:

> H4: UDL tools will have higher utilization in conditions of high intrinsic load
>
> H5: UDL tools will have higher utilization in conditions of high extraneous load

It should also be noted that utilization, here, indicates that the learner accessed the features within the digital environment via a button click. Additionally, supplementary data was collected pertaining to the students' cognitive gains and self-reported feature utility, congruent with Session 3 of Phase I.

5.2 Methods

5.2.1 Instructional materials.

As in Phase I, the instructional materials for the RoboBooks for this phase were scheduled to correspond with the classroom syllabus. In the Chemistry class, the content for the RoboBooks pertained to compounds, chemical bonds and reactions, predicting products, and balancing equations. The physics content covered circuit diagrams, resistors and conductors, Ohm's law, and parallel and series circuits. The classroom intervention spanned 5 days for Chemistry, in 75-minute periods, and 4 days in Physics in 3, 85-minute blocks and one 120-minute block. The RoboBooks consisted of several sets of instructional content with supplementary images, videos, and simulations, as well complementary hands-on activities, which required instruction from, and input to, the RoboBook.

5.2.2 Cognitive load by page.

5.2.2.1 Intrinsic load.

For each content page in the RoboBook, an estimate of element interactivity was derived following the Carlson et al. (2003) method defined in Phase II. For example, the page shown in Figure 5.1 has the following estimate number of element interactivity:

1) Note current to be flowing
2) Associate flowing current to closed circuit
3) Note diagram to the right
4) Identify elements in the diagram as battery, lamp, and switch
5) Note switch is closed
6) Associate closed switch to electrons flowing

7) Associate electrons flowing to lamp lighting

This value was used to quantify intrinsic cognitive load imposed by the content in each of the pages. Pages were subsequently categorized as having high, medium, or low intrinsic load, in accordance to the number of interacting elements, such that: “low” equates to 3 or fewer interacting elements; “medium” means 4-5 interacting elements; and “high” is more than 5 interacting elements. For the example in Figure 5.1, the page would have high intrinsic load because the estimate number of interacting elements is 7.

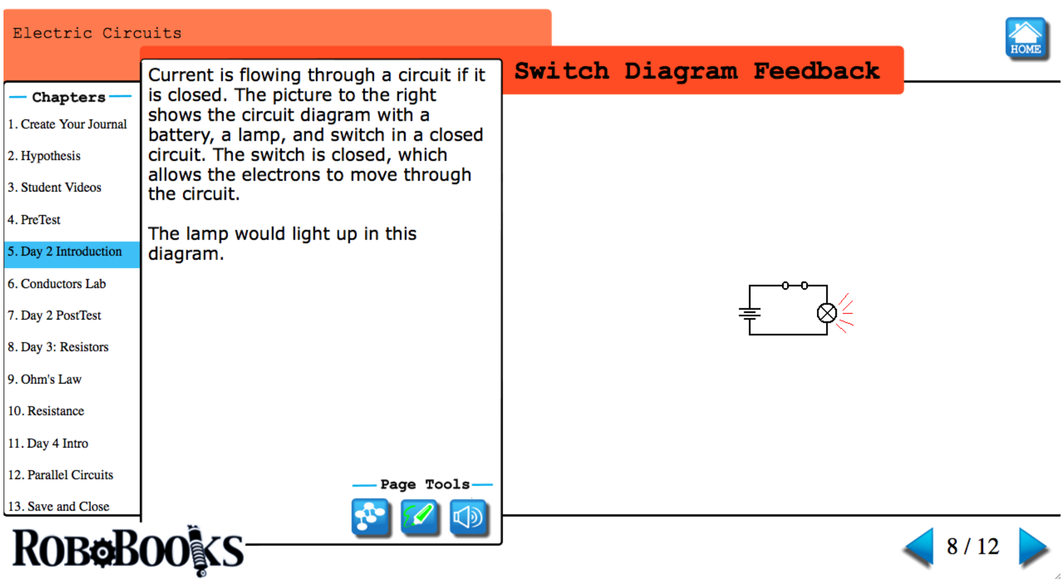


Figure 5.1. *Phase III, Physics screenshot, example page*

5.2.2.2 Extraneous load.

Each page in the RoboBook was also assigned to an extraneous load categorization. A low extraneous load assignment would be made for pages with a complementary image, movie, simulation, or worked example that accompanied the text. For content with no accompanying extraneous load reduction, for example pages with only text, a high extraneous load categorization would be assigned. The example in Figure 5.1, which shows a supplementary diagram depicting a closed switch, a circuit, and lighted lamp, exemplifies extraneous load reduction. In this case, the page would have low extraneous load.

5.2.3 Apparatus.

5.2.3.1 RoboBooks.

The same hardware and software specifications, as detailed in section 3.2 of Phase I, are also applicable in this phase. All students used 13" MacBooks displaying the RoboBooks interface. The built-in mouse pad, web cam, and microphone were used for collecting user input across response modalities.

5.2.3.2 Data collection.

5.2.3.2.1 User Survey.

Upon completion of the content presentation, all students were provided a modified user survey used in Session 3 of Phase I, presented in Appendix C. The questions in this survey were used to determine preferential use of features in the RoboBooks interface as self reported by the students. In this phase, the survey was administered to the students in paper form. The 6-point Likert scale, from which the students selected their responses to the first set of 10 questions, ranged from most negative (1) to most positive (6), with the intervals evenly spaced. Similarly, a 4-point Likert scale, from which the students selected their responses to the second set of 4 questions, ranged from most negative (1) to most positive (4). The evaluation questions look at 10 UDL features designed to gather information on the value of those features in RoboBooks for the purpose of supporting science instruction. The second set of questions looks at the utility of the digital workbook itself for science instruction in the classroom.

In both classes, students were administered a pretest and complementary posttest designed by the teachers in order to evaluate learning gains. The test questions were designed to assess the learning of the main science ideas in the selected topics presented in the RoboBooks. The tests were scored as instructed by the teachers.

5.2.3.2.2 Navigation Tracking.

Full navigation tracking was deployed in this phase to accompany access of RoboBook features and pages. Accordingly, a tracking script captured, to the second, each interaction of the student with the interface. Additionally, the navigation tracking noted UDL tool utilization behavior and all times per page. This action tracking was used to determine page times and frequency of use for the following UDL tools: hint, highlight, concept map, and text to speech.

5.2.4 Subject information.

The classroom implementation for this phase was conducted in one class at Fenway High School and one class at Boston Arts Academy (BAA). Students in the Fenway class were exposed to RoboBooks during the three sessions in Phase I and are each categorized as having a high incidence disability. However, only 6 students were present for the RoboBooks session in this phase. Additionally, 7 students from the BAA class completed the RoboBooks for this phase, of which 29% of the students are categorized in the high incidence disabilities group. Each of the BAA students was paired at a laptop.

5.3 Results

5.3.1 User feedback.

Students were surveyed upon completion of the RoboBooks activities for both high school classes. Table 5.1 shows a summary the 13 students' responses to 10 items pertaining to the usefulness of UDL features on the 1-6 rating scale. The mean values in the user survey tend toward the positive, with no items below the 3.5 threshold for moving into more positive judgments. 100% of the items generated mean responses in excess of 4.0, such that all features were found to be at least fairly helpful for supporting understanding of the science content. Furthermore, 30% of the features (audio record response, picture response, and text-to-speech) generating mean responses in excess of 5.0; these features were found to be very helpful in understanding the science content. The narrow standard deviations for several of the features provide additional indication that the students were strongly united in their assessments of the many features of RoboBooks.

Table 5.1. Average response to UDL feature utility, student self-response

Item	Mean	SD
The option to record audio for your responses?	5.00	1.05
The option to type for your responses?	4.64	1.39
The option to take pictures for your responses?	5.23	0.73
The text read out loud?	5.00	1.04
The movies that went with a text passage?	4.57	1.40
The images that went with a text passage?	4.38	1.04
The concept map?	4.18	1.47
The highlighting?	4.00	1.67
The hints?	4.50	1.38
The games and simulations?	4.08	1.55

Table 5.2 shows a summary the 13 students' responses to 4 items on a 1-4 rating scale. The mean values in the user survey tend toward positive, with no items below the 2.5 threshold for moving into more positive judgments. 100% of the items generated mean responses in excess of 3.0, such that students positively responded to: the RoboBooks tool aiding in understanding the content, would like to use the software again in the future, enjoyed learning about science using the software, and overall enjoyed working on science in a digital environment.

Table 5.2. Average response to RoboBooks utility, student self-response

Item	Mean	SD
Did RoboBooks help you understand the science information?	3.39	0.63
Would you like to use RoboBooks in the future?	3.23	0.60
Overall, did you enjoy learning about science when using RoboBooks in class?	3.07	0.83
Did you enjoy working on science activities in a digital environment?	3.11	0.49

5.3.2 Cognitive gains.

Table 5.3 shows pretest and posttest data gathered from questions designed to explore the learning gains for the RoboBook in this phase. While the average percent improvement for Fenway students was 5.7%, three students showed improvements over 9%. The average percent improvement was 32% for BAA and students ranged from 12% to 59% improvement in performance.

Table 5.3. Pre and post test average scores by school

School	Max Score	Mean: Pre	SD: Pre	Mean: Post	SD: Post
Fenway	33	17.62	4.27	19.33	4.89
BAA	17	6.90	2.56	13.22	2.91

5.3.3 Navigation behavior.

5.3.3.1 Page tracking.

The navigation tracking data for students was used to map behavior patterns of the students throughout this classroom session. Similar to findings in Section 3, aggregated data shows diverging and converging patterns that are task dependent. For example, Figure 5.2 shows the progression of the Fenway students through the Session 4 RoboBook, by page and over time. This visual representation of tracking data shows behavior that trend similarly, as seen in the Fenway class. This session was largely driven by classroom hands-on activities with circuit boards and such tasks drove much of the student behavior. Figure 5.3 likewise shows similar patterns in behavior exhibited by BAA students in their final session, also while completing a task, which consisted of a sequence of very similar subtask components.

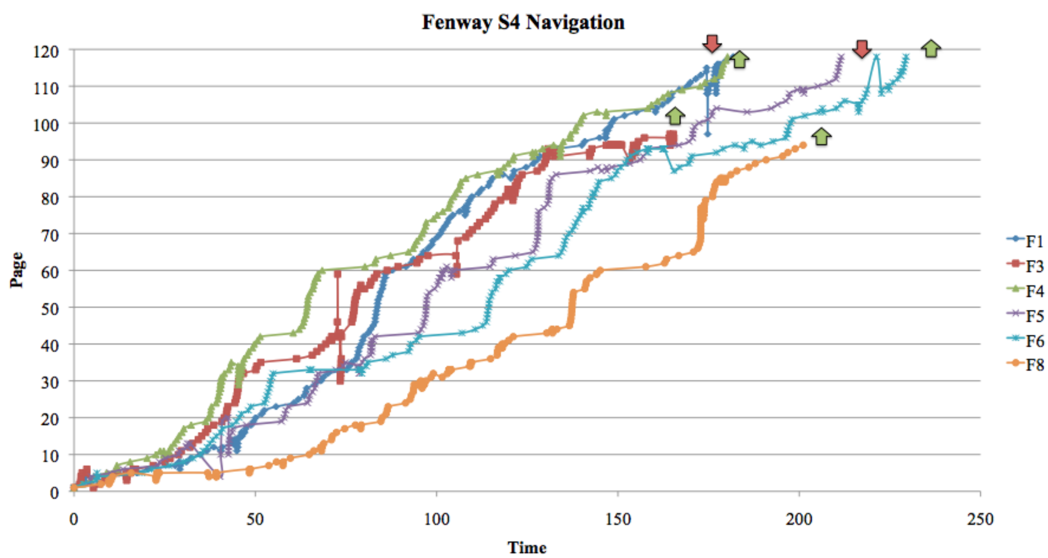


Figure 5.2. *Fenway navigation data for Session 4*

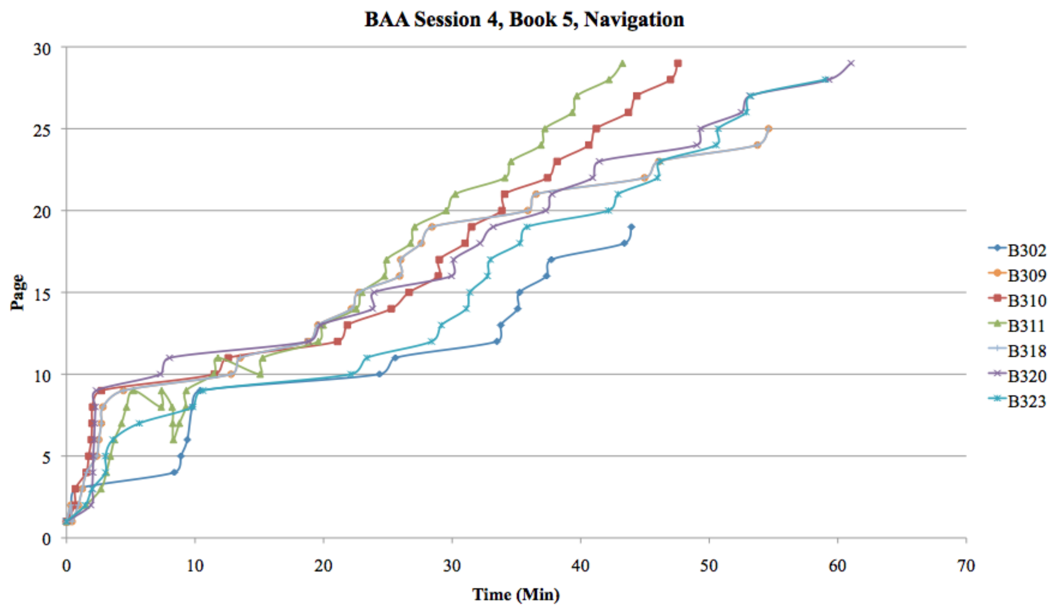


Figure 5.3. *BAA navigation data for Session 4, Day 5*

Again, navigation behavior with inclusion of UDL tool usage was investigated on an individual student basis, as shown in Figures 5.4 and 5.5. In both figures, the students use text and image inputs throughout the book. In the first of these graphs, the tool utilization of a student that who showed 19% pre to post test cognitive gain is presented. This student used each of the additional features (TTS or text to speech, hint, concept map, and highlight) throughout the book. However, the student shown in Figure 5.5 did not show performance gains. This student utilizes each of the additional features initially, but only returns to the hint and text to speech features throughout the book. A possible explanation of use of features through a book is the student is actively managing cognitive load by accessing features as needed, such as when intrinsic or extraneous load is high.

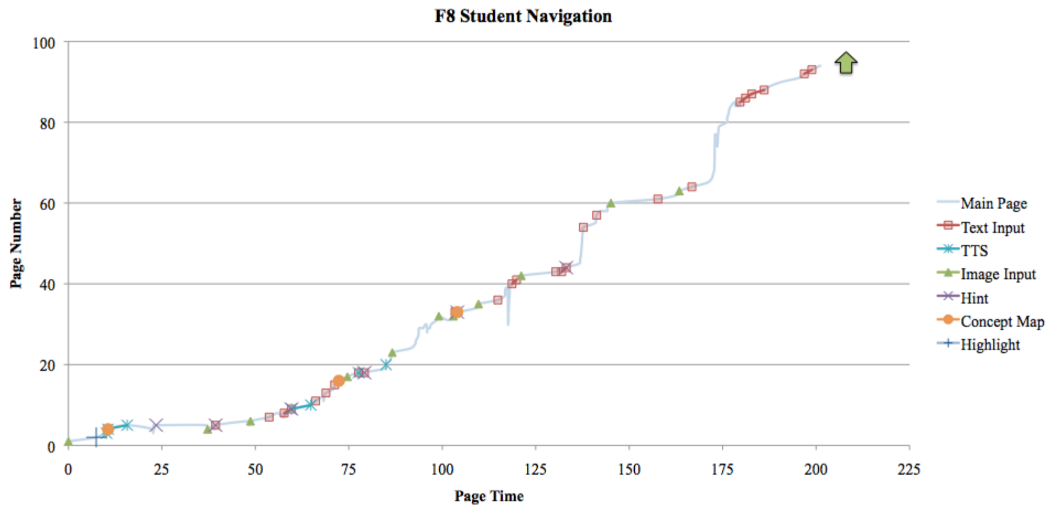


Figure 5.4. *Feature utilization for Fenway student with improved performance*

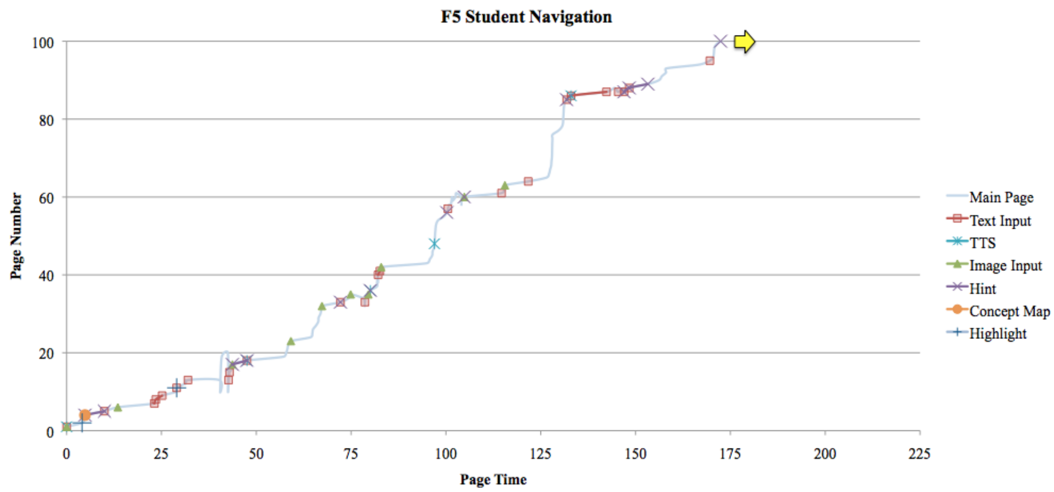


Figure 5.5. *Feature utilization for Fenway student with no performance gains*

Both of the Fenway students represented the high incidence disability group. Figure 5.6 shows a high performing general education student from the BAA classroom for day 5. Though there may be task differentials that drive tool usage, notice that selection of concept map and highlight features throughout the book. This may indicate an advanced strategy for managing cognitive load through utilization of UDL tools.

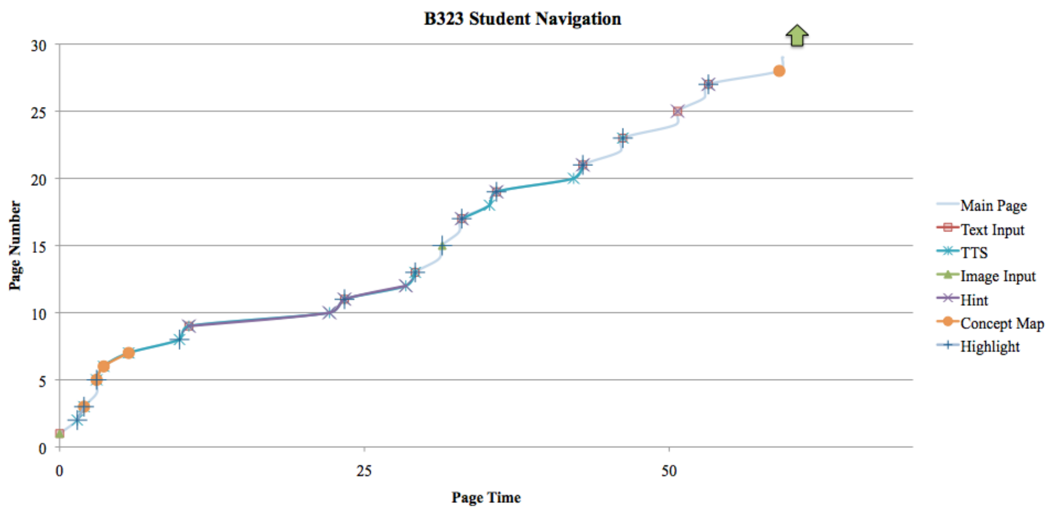


Figure 5.6. *Feature utilization for BAA student with improved performance*

5.3.3.2 Feature utilization by page.

In order to determine if UDL tools were actually being used in a manner that is differential by task, a comparison was made between tool usage and page characteristics, as described in section 5.2.2. These relationships were investigated in order to examine the related hypothesis, that UDL tools, similar to other techniques for extraneous load reduction, will have higher utilization when intrinsic load is high (H4). Furthermore, it is expected that UDL tools will have higher utilization in conditions of high extraneous load (H5).

A crosstab analysis shows a significant correlation (R=0.219, p=0.026) for the element interactivity number per page with the average tools used on the page in the Fenway classroom. Significant correlations in average tool use and element interactivity number were also found in the BAA classroom (R=0.24, p=0.025). A similar analysis further investigated low, medium, and high levels of intrinsic load. As such, for intrinsic load, and corresponding average tool use for these pages, a significant correlation was found for high intrinsic load (R=0.881, p=0.00) and medium intrinsic load (R = 0.549, p=0.00) in the Fenway classroom. Additional analysis was run for high and low levels of extraneous load. For high extraneous load, significance was found between text-only pages and corresponding average tool use (R=0.496, p=0.00) in the Fenway classroom, such that the tools were utilized more in text-only formats. Significance was not found in levels of intrinsic load when correlated to tool usage for BAA.

Germane load, the cognitive load component attributed to level of engagement, was not investigated in the lab nor initially considered for investigation for a relationship with tool utility in this research. However, during informal observations in the classroom, where it appeared tool use may be higher in more active sections of the RoboBook session, a post hoc germane categorization was assigned to each page. While this is not an established categorization practice, this peripheral investigation looked at the utility of the tools when engagement could be considered higher; for example, tasks that require students to actively synthesize information, make decisions, and subsequently convey understanding. This was compared to tasks that required students to acquire information in a more passive form of information presentation. Pages with only information presentation were coded to have low germane load. Alternatively, pages that required a more active information acquisition, processing, and response from the student, such as during an activity requiring student input, were assigned to the high germane load category. Fenway students showed a significant correlation between average tool use and high germane load (R=0.226, p=0.05).

5.4 Discussion

Significant correlations between tool usage and high intrinsic load (as predicted in H4) and tool usage and high extraneous load (as predicted in H5) indicate that in cases where either of the cognitive load factors is higher, UDL tool utilization is greater. Moreover, significance in higher levels of germane load further support CLT, which claims that instructional materials and presentation should be constructed such that for complex material (high intrinsic load), extraneous load should be kept low and germane load should be kept high. While these findings were not found across both classrooms for all levels of each load time, a relationship between tool use and intrinsic load was maintained in both classrooms.

The correlations suggest a relationship between tool utilization and high intrinsic and extraneous load. Consequently, use of UDL tools should in turn reduce cognitive load in a manner similar to extraneous load reductions determined in the lab data. Ultimately, manipulations to cognitive load by application of UDL tools should promote performance gains.

The UDL tools were rated highly amongst the students for aiding in understanding the science content, which reinforces their use. However, "why" the students' assigned merit to these tools can be accordingly explained as the UDL tools are conscientiously being utilized in order to manage cognitive load. Hence, when cognitive load is too great, tools are utilized in order to reduce the overall cognitive load; thus, promoting accessibility to the instructional materials. Consequently, as predicted in H4 and H5, the students find merit in the tools to aid their understanding of the science instructional materials. This is reflected in the self-reported utility of the features and is enforced by the correlation of actual tool use with high levels of intrinsic and extraneous load, and low levels of germane load.

6. General Discussion

Deriving cognitive load components in learning tasks provides a means to systemically predict behavior and consequently devise improved opportunities for individualized learning. Moreover, by understanding UDL within the context of CLT, these opportunities can be extended to increase accessibility and best promote equitable learning opportunities. Ultimately, by approaching the design of digital instructional media from a cognitive load theory framework, active management of cognitive load is achieved through utilization of supplementary accessibility tools.

If we conceptualize each individual short-term memory capacity as the bottleneck for learning, considerations for instructional design must be made for a range of learner capacities. And, with a range of cognitive capacities, designing UDLs in digital instructional media within the context of CLT provides active mechanisms for students who may have limited capacity for information acquisition. For these students, management of cognitive load is increasingly imperative, especially when skills for schemata creation may also be limited. Yet, tools for managing cognitive load are additionally pertinent to aid all learners in managing cognitive load, particularly with the many complex ideas prevalent in science instruction. As shown by the Phase 3 findings, learners do indeed utilize tools more in situations where intrinsic and extraneous load are higher, evidence that learners are actively managing cognitive load in situations where it is needed.

Moreover, a CLT approach has implications for the design of educational software from a human factors perspective for both curriculum and interface development. As shown in this research, learning tasks can be defined in terms of cognitive load components. The research findings suggest that tasks that are high in intrinsic load should be accompanied with extraneous load reduction; this includes utilizing multiple information channels, offloading memory demands, and the provision of cognitive aids. These componentized tasks and associated techniques for extraneous load reduction can be tested with users prior to production, in a true learner-centered design fashion. Furthermore, UDL tools can be assessed such that a blanket, "apply to all" formula is avoided; informed design decisions can determine best use for the development of UDL tools within an interface. Consequently, resources can be selectively allocated to the learning tasks that warrant inclusion of supplementary aids and focus can be directed towards providing the most effective means of extraneous load reduction for different learning tasks.

7. Conclusion

This research encapsulated observed behaviors from an initial phase of study into a theoretical framework and tested resulting design alternatives in a controlled laboratory experiment during a secondary phase. A final phase reapplied cognitive load theory in the naturalistic setting to capture the true behavior of the learners in their normal environment. Consequently, this thesis was a holistic, user-driven, theoretically founded, and system-tested approach to understanding the role of UDL within instructional media design.

During the first phase, learner-centered design principles were utilized in several classroom implementations of evolving interface designs. User needs were determined by leveraging research on accessibility and by soliciting teacher input to each session. The iterative nature of the design process was established through the three-session process, from which previous session feedback resulted in design parameters for subsequent sessions. As such, the UDL features were developed in the RoboBooks platform over several test sessions with inclusion of both the needs of the students and the teachers. The features were well received for their utility in the classroom, with the third design receiving over 80% agreement in utility for understanding the science instructions from students for 70% of the features. Moreover, navigation tracking provided insight into different behavior trends from the students and launched investigation for task dependencies on UDL feature utilization.

As seen in the second phase lab experiments, the robustness of concepts across learning tasks in a controlled environment showed that behavior could be manipulated in a manner predictable to cognitive load theory. In this context, performance gains in conditions of higher intrinsic cognitive load are found by application of extraneous load reduction. Benefits of the UDLs trend in a manner that implies, with further testing, UDL usage too produces performance gains by similarly lowering extraneous load. Results showed indications of extraneous load reduction by use of diagrams and worked examples, and in certain usages of diagrammatic concept maps and highlighting UDL tools. Manipulations of intrinsic load were firmly established throughout each of 4 design implementations, which tested 40 students performing several different learning tasks.

When extending CLT back into the classroom, differential tool use was discovered with varying instructional presentation within the CLT framework. Consequently, UDL tools used in high intrinsic and extraneous load conditions may be best situated to promote performance gains. Furthermore, peripheral indications of differential usage due to varying levels of germane load further support CLT in the design of accessible, educational technology.

When considering the findings in a progression throughout the research, Figure 7.1 shows key findings for each of the phases. As depicted, navigation tracking in Phase I showed that student behavior was converging and diverging in manner dependent upon the task. Phase IIa consequently looked at tasks within the CLT framework, distinguishing high and low intrinsic and extraneous load levels and determining that extraneous load reduction in situations of high intrinsic load resulted in increased performance. Similarly, UDL designs in Phase IIb showed indications of performance gains trending toward extraneous load reduction techniques employed in the original designs. In the last phase, UDL tool utilization was investigated particularly for correlations in use with respect to high intrinsic and extraneous load. Findings suggest that UDL tools are utilized in situations of high intrinsic and extraneous load such that cognitive load can be managed and improvements are made in performance.

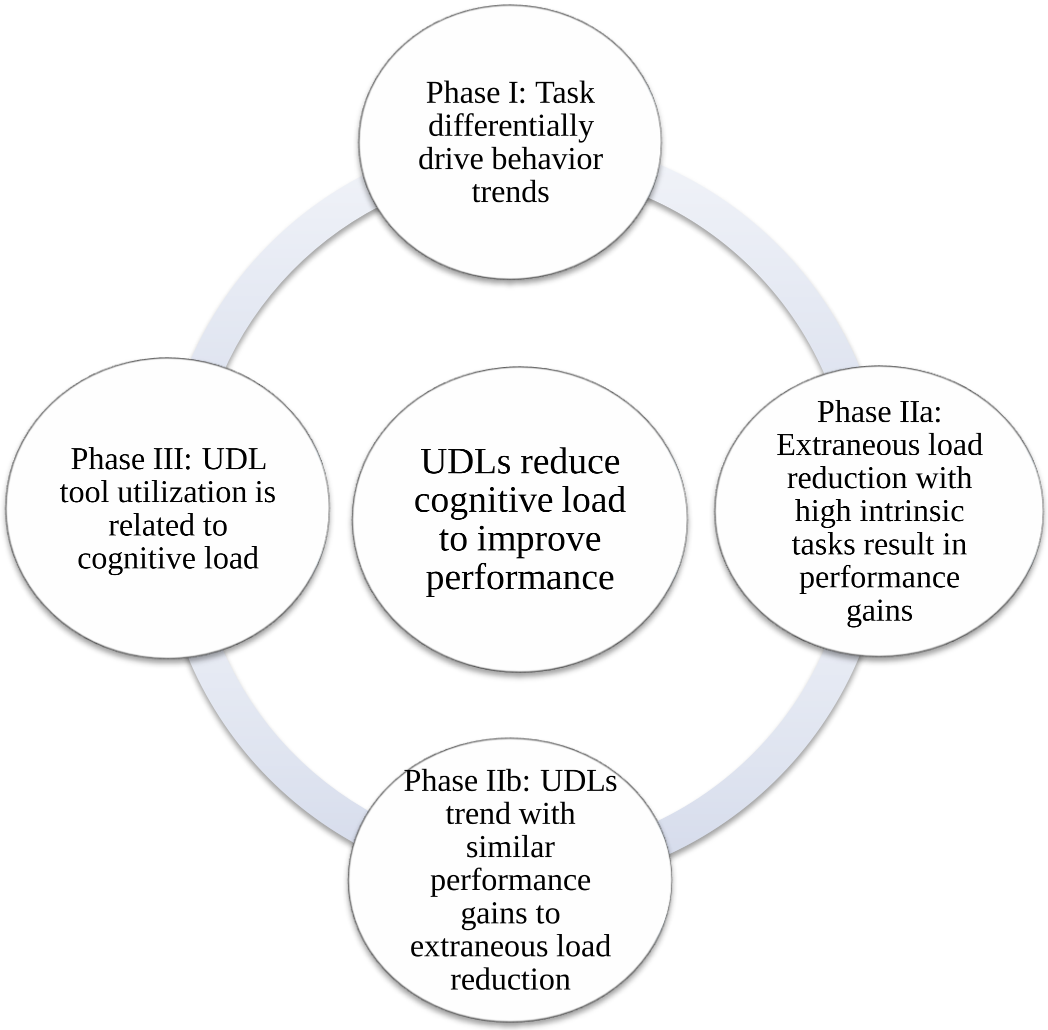


Figure 7.1. *What CLT tells us about the design of educational technology & UDLs*

These research claims can be further supported with an increased sample size for participants in the Phase II laboratory research in order to definitely determine UDL application on performance gains. In the naturalistic setting, following digital "pages" of varying cognitive load with subsequent cognitive evaluations would also promote directly connection of UDL utilization and performance gains. To continue research with a CLT approach to digital instructional design and for developing UDL tools in educational technology, functional relationships between intrinsic, extraneous, and germane load should be explored. Further investigation into the many different types of UDL tools and their individual affordances for cognitive load management would provide additional design guidelines for the implementation of accessibility tools within digital instructional media. Further exploration could also look at schemata creation, long-term memory retention, and consequent learning gains over time.

Approaching digital instructional media with a CLT framework, particularly with respect to accessibility, provides a means of considering the utility of UDL tools for managing cognitive load. This has several design implications for the development of accessibly in educational technology. As such, digital instructional, from a cognitive load theory approach, could be developed with intent to manage cognitive load. These implications are pertinent, but not limited to, the interface design, curriculum development, and the learner interactions.

When designing digital instructional media interfaces, for example, a cognitive load theory approach can aid in determining when and where to expend development resources. As such, UDL tools can be designed in a manner that maximizes performance opportunities. Content for the digital environment can be systematically evaluated for intrinsic cognitive load; subsequently, techniques to reduce cognitive load can be expended in situation where content complexity is high. And the user, a learner interacting with the digital media and embedded content, can actively participate in managing their cognitive load. These design implications, and the potential for associated performance gains, warrant consideration of CLT for learner centered design of digital instructional media.

8. Appendices

Appendix A: Student Survey for Session 1

Please help us design the next RoboBooks for your classroom. For each of the items, please indicate whether you agree (Yes) or disagree (No) with the statement.

Navigation

It was easy to navigate through RoboBooks.
I liked using the chapters and subsection menu to see where I was in the RoboBook.
I understood what would happen when I selected each of the buttons on the page.
I would like it if I could explore different chapters and topics when I decide to (instead of in order).

Information

Reading the lessons helping me understand the unit.
Listening to the audio recordings helped me understand the unit.
The illustrations and videos helped me understand the unit.
The simulations helped me understand the unit.

Response

I liked typing my answers to questions.
I liked audio recording my answers to questions.
I liked adding an image for my answers to questions.
I would like to have more ways to respond to answers on page (not just one way).
I prefer to have just one way to answer on a page (example: text only).

Content

I liked having activities to help better understand the lesson.
I liked being able to choose what I wanted to learn about.
I liked having a Journal where I could see all my work in one place.

Layout

The page layout was clear and consistent.
I liked the color scheme.
There was too much text on the pages.

Appendix B: Student Survey for Session 2

Please help us design the next RoboBooks for your classroom. For each of the items, please indicate whether you agree (Yes) or disagree (No) with the statement.

Navigation

I think that most people would learn to use RoboBooks very quickly.
Learning how to use all the functions in RoboBooks is difficult.
Navigating through RoboBooks is straightforward.
I don't feel comfortable using RoboBooks.

Information

The organization of information the page is confusing.
I like that RoboBooks allows me to access information through multiple options such as text and movies.
The sequence of information and activities is confusing.
It is easy to find the information I need to complete the RoboBook activities.

Tools

I liked being able to choose how I would respond to a question.
Getting feedback when an answer was right or wrong is helpful.
I like that I can chose the functions (like Highlight or Hint) that I want to use on a page.
There are too many functions on the page that I did not want to use.

Content

Taking a break for the class discussion is helpful to understanding the information.
It was too hard to become engaged with RoboBooks after we had a discussion.
Getting feedback when an answer was right or wrong helped me know if I understood the unit.
The content of this RoboBook did not meet my expectations of what I should learn today.

Layout

There is too much inconsistency in this RoboBook.
This RoboBook is visually appealing.
This RoboBook is too complex.
It is easy to move from one page to another in RoboBooks.

Appendix C: Student Survey used in Session 3

For each of the items, please use the scale below to rate how helpful the features were in supporting you understand the unit.

1 = I don't agree at all
2 = I somewhat disagree
3 = I somewhat agree
4 = I completely agree

<u>Response</u>

The option to record audio for your responses?
The option to type for your responses?
The option to take pictures for your responses?
The option to make stop motion movies for your responses?

<u>Information</u>

The text read out loud?
The movies that went with a text passage?
The images that went with a text passage?
The science content that was presented in the RoboBook?

<u>Tools</u>

The concept map?
The highlighting?
The hints?
The calculator?
The activity timer?
The option to answer different ways?
The questions with right or wrong feedback?
The games and simulations?

<u>Content</u>

Did RoboBooks help you understand the science information?
Would you like to use RoboBooks in the future?
Overall, did you enjoy learning about science when using RoboBooks in class?
Did you enjoy working on science activities in a digital environment?

Appendix D: Estimates of Number of Interacting Elements for Instructional Material Used in Phase 2, Task 1

Examples from Low Element Interactivity Material of Subtask 1

The following is an estimate of the number of different interacting elements that need to be considered when constructing the molecular models:

3 element questions [Q1, Q2, Q3, Q4, Q5, Q7, Q8]:

Example: Sodium chloride

1) Select sodium ball.
2) Select chlorine ball.
3) Select single link and attach each end of link to both balls.

5 element questions [Q6, Q9, Q10]

Example: Dihydrogen monoxide

1) Select oxygen ball
2) Select hydrogen ball
3) Select single link and attach each end of link to both balls.
4) Note location of remaining hole(s) on oxygen ball.
5) Repeat Steps 2-3.

Examples from High Element Interactivity Material of Subtask 2

The following is an estimate of the number of different interacting elements that need to be considered when constructing the molecular models:

16 element question [Q1]: Sodium hydrogen sulfate

1) Select sulphur ball.
2) Select oxygen ball.
3) Select double flexi-links and attach end of links to both sulphur and oxygen balls.
4) Note location of where double link is attached to hydrogen ball and select opposite side of ball.
5) Select new oxygen ball.
6) Repeat Step 3.
7) Note location of remaining holes on sulphur ball.
8) Select new oxygen ball.
9) Select single link and attach end of links to new oxygen ball and sulphur ball in Step 7.
10) Repeat Steps 7-9.

11) Select and oxygen ball that has a single link to sulphur ball.
12) Select hydrogen ball.
13) Select single link and attach each end to both the oxygen ball in Step 11 and the hydrogen ball.
14) Select the remaining oxygen ball that has a single link to the sulfur ball.
15) Select sodium ball.
16) Select single link and attach each end to the oxygen ball in Step 14 and the sodium ball.

19 element question [Q2]: Acetic Acid

1) Select carbon ball.
2) Select second carbon ball.
3) Select single wooden link and attach end of link to both carbon balls.
4) Note location of where single link is attached to first carbon ball and select opposite side of ball.
5) Select hydrogen ball.
6) Select single wooden link and attach end of link to the carbon ball and hydrogen ball.
7) Note location of remaining holes on carbon ball.
8) Select new hydrogen ball.
9) Repeat Step 6.
10) Repeat Steps 7-9.
11) Note location of remaining carbon ball.
12) Select an oxygen ball.
13) Select double flexi links and attach end of links to both carbon and oxygen ball.
14) Note location of remaining holes on second carbon ball.
15) Select new oxygen ball.
16) Select single link and attach each end to the remaining hole in the carbon and the new oxygen ball.
17) Note location of remaining holes on oxygen ball with single link to carbon ball.
18) Select new hydrogen ball.
19) Select single link and attach ends to the remaining oxygen ball and new hydrogen ball.

Adapted from Carlson et al. (2003)

Appendix E: Estimates of Number of Interacting Elements for Instructional Material Used in Phase 2, Task 2

Example from Low Element Interactivity Material of Subtask 1

The following is an estimate of the number of different interacting elements that need to be considered when learning the structure of the carbon compound prefixes:

2 Element Rules: [R1.1, 1.2, 1.3, 1.4, 1.5, 1.6]

Example: Rule 1.1 (propan)

1) Note the number of carbon atoms
2) Record the name of the corresponding prefix

Example from High Element Interactivity Material of Subtask 2

The following is an estimate of the number of different interacting elements that need to be considered when learning the structure of the carbon compound suffixes:

5 Element Rule [2.2, 2.5,]

Example: Rule 2.2 (yne)

1) Note the carbon atom in the molecular model.
2) Note a second carbon atom in the molecular model.
3) Ensure the two carbon atoms are joined by a triple bond.
4) Note the hydrogen atom in the molecular model.
5) Ensure the hydrogen atom is joined to a carbon atom by a single bond.

7 Element Rules [2.1, 2.2, 2.3, 2.4, 2.5, 2.6]

Example: Rule 2.1 (amide)

1) Note the carbon atom in the molecular model.
2) Note the oxygen atom in the molecular model.
3) Ensure that the oxygen and carbon atoms are joined by a double bond.
4) Note the nitrogen atom in the molecular model.
5) Ensure that the carbon atom in Step 1 is joined to the nitrogen atom by a single bond.
6) Note that there are two hydrogen atoms in model.
7) Ensure that the nitrogen atom in Step 4 is joined to both hydrogen atoms by single bonds.

8 Element Rules [2.3, 2.4, 2.6]

Example: Rule 2.3 (ol)

1) Note the carbon atom in the molecular model.
2) Note the hydrogen atom in the molecular model.
3) Note a second hydrogen atom in the molecular model.
4) Ensure the carbon and hydrogen atoms are joined by single bonds.
5) Note the oxygen atom in the molecular model.
6) Ensure one of the hydrogen atoms is joined to the oxygen atom by a single bond.
7) Note an additional hydrogen atom in the model.
8) Ensure oxygen atom in Step 6 is attached to hydrogen atom by a single bond.

Adapted from Carlson et al. (2003)

Appendix F: Estimates of Number of Interacting Elements for Instructional Material Used in Phase 2, Task 3

Example from Low Element Interactivity Material of Subtask 1

The following is an estimate of the number of different interacting elements that need to be considered when learning types of chemical reactions:

4 Element Rules: [1.1, 1.2, 1.3, 1.4]

Example: Rule 1.1 (synthesis)

1) Note the first reactant substance
2) Note the second reactant substance
3) Combine the two substances
4) Record the corresponding chemical reaction name

5 Element Rule [1.5]

Example: Rule 1.5 (combustion)

1) Note the first reactant compound
2) Identify second reactant compound as oxygen
3) Note first product as water
4) Note second product as carbon dioxide
5) Record corresponding chemical reaction name

Example from High Element Interactivity Material of Subtask 2

The following is an estimate of the number of different interacting elements that need to be considered when learning how to balance equations:

13 Element Rule [2.1]: Verify Oxidation Numbers

1) Select table of common ions
2) Look up oxidation number for the cation of the first reactant substance
3) Record corresponding oxidation number on first reactant substance
4) Note remaining reactant substances
5) Look up oxidation number for the anion of the second reactant substance
6) Note remaining reactant substances
7) Repeat Steps 3-7 for all remaining reactant substances
8) Note first reactant compound
9) Note if substances in compound have subscript
10) Multiply oxidation number for substances with subscript
11) Note if substances in compound have a coefficient
12) Multiply oxidation number for substances with a coefficient

13) Verify oxidations numbers sum to 0 for each compound

16 Element Rule [2.2]: Predict Product

1) Indicate a yield arrow after reactants
2) Identify the positive ion from first reactant
3) Record the positive ion on product side of equation
4) Record the oxidation number of positive ion on product side of equation
5) Identify the negative ion from second reactant
6) Record the negative ion on product side of equation
7) Record the oxidation number of negative ion on product side of equation
8) Create a compound with negative ion and positive ion on product side of equation
9) Identify remaining positive ions from reactant side of equation
10) Record positive ion and oxidation number on product side of equation
11) Identify remaining negative ions from reactant side of equation
12) Record negative ion and oxidation number on product side of equation
13) Create a compound with remaining positive ion and negative ion on product side of equation
14) Verify all positive ions precede all negative ions
15) Add a plus sign between products
16) Verify all oxidation numbers sum to zero [Apply Rule 2.1]

10 Element Rule [2.3]: Balance Equation

1) Note subscripts are fixed
2) Choose compound with greatest number of atoms
3) Locate molecule on opposite side of equation
4) Place necessary coefficient in front of molecule
5) Apply new coefficient to entire compound
6) Identify other molecule in compound from Step 5
7) Determine number of atoms for other molecule in compound from Step 5
8) Locate molecule from Step 7 on opposite side of equation
9) Apply Rules 4-8 for all molecules in equation
10) Verify the number of atoms for each molecule on left side of the equation equals the number of atoms for each molecule on the right side of the equation

9. Bibliography

Bohman, P. & Anderson, S. (2005). A conceptual framework for accessibility tools to benefit users with cognitive disabilities. *W4A '05 Proceedings of the 2005 International Cross-Disciplinary Workshop on Web Accessibility*. 85-89.

Center for Applied Special Technology (2011). *About UDL.* Retrieved from http://www.cast.org/udl/index.html

Carolson, R., Changler, P., & Sweller, J. (2003). Learning and Understanding Science Instructional Material. *Journal of Educational Psychology.* 95(3), 629-640.

Danahy, E., Goswamy, A., & Rogers, C. (2008). Future of robotics education: The design and creation of interactive notebooks for teaching robotic concepts. I*EEE International Conference on Technologies for Practical Robotic Applications. Woburn, MA.*

Duckworth, E. (1973) Piaget takes a teacher's look. *Learning*, *2*(2), pp. 22-27.

DeLeeuw, K. E., & Mayer, R. E. (2008). A Comparison of Three Measures of Cognitive Load: Evidence for Separable Measures of Intrinsic, Extraneous, and Germane Load. *Journal of Educational Psychology*. 100(1), 223-234.

Gopher, D., & Braune, R. (1984) On the psychophysics of workload: Why bother with subjective measures. *Human Factors*. 26, 519-532.

Graham, S., Harris, K., MacArthur, C., & Schwartz, S. (1991). Writing and Writing Instruction for Students with Learning Disabilities. *Learning Disability Quarterly.* 14(2), 89-114.

Gribbons, W. (2008). Universal Accessibility and Functionally Illiterate Populations: Implications for HCI, Design, and Testing. In A. Sears & J. Jacko (Eds.), *The Human Computer Interaction Handbook* (872 – 881). Mahwah, NJ: Lawrence Erlbaum Associates.

Hollender, N., Hofmann, C., Deneke, M., Schmitz, B. (2010). Integrating cognitive load theory and concepts of human-computer interaction. *Computers in Human Behavior.* 26(6), 1278-1288.

Johnson, L., Smith, R., Levine, A., and Haywood, K., (2010). *2010 Horizon Report: K-12 Edition.* Austin, Texas: The New Media Consortium.

Jonassen (2003). Using Cognitive Tools to Represent Problems. *Journal of Research on Technology in Education.* 35(3), 362-382.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the designof multimedia instruction. *Journal of Educational Psychology*, *92*, 126–136

Kalyuga, S. & Sweller, J. (2004). Measuring Knowledge to Optimize Cognitive Load Factors During Instruction. *Journal of Educational Psychology.* 96(3), 558-568.

Larsen, S. (1995). What is "Quality" in the Use of Technology for Children with Learning Disabilities? *Learning Disability Quarterly*. 18(2), 118-130.

Leahy, W., & Sweller, J. (2005). Interactions Among the Imagination, Expertise

Reversal, and Element Interactivity Effects. *Journal of Experimental Psychology: Applied*. 11(4), 266-276.
Marcus, Cooper, and Sweller (1996). Understanding Instructions. *Journal of Educational Psychology*. 88(1), 49-63.
Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist. 3,* 43–53.
Meister, D. (1999). *The history of human factors and ergonomics.* Mahwah, NJ: Lawrence Erlbaum Associates.
National Assessment of Educational Progress, 2005. Available at http://nces.ed.gov/nationsreportcard/.
National Science Education Standards, 1996. Available at http://www.nap.edu/html/nses/.
Paas, F., Renkel, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science*. 32, 1-8.
Paas, F. & Van Merriënboer, J.J.G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach, *Journal of Educational Psychology.* 86, 122–133.
Phillips, R. (2005). Pedagogical Institutional and Human Factors Influencing the Widespread Adoption of Educational technology in Higher Education. *Proceedings of the 22nd annual conference of the Australasian Society for Computers in Learning in Tertiary Education.* Brisbane: Ascilite.
Rose, D. H., & Meyer, A. (2002). *Teaching every student in the digital age: Universal Design for Learning.* Alexandria, VA: ASCD.
Rose, D. H., Meyer, A., & Hitchcock, C. (2005). *The universally designed classroom: Accessible curriculum and digital technologies.* Cambridge, MA: Harvard Education Press.
Smith, T. J. (2007). The ergonomics of learning: educational design and learning performance. *Ergonomics*. 50(10), 530-1546.
Soloway, E., Guzdial, M., & Hay, K. E. (1994). Learner-Centered Design: The Challenge for HCI in the 21st Century. *Interactions* 1(2), 36-48.
Stary, C. (2001). Handling User Diversity in Task-Oriented Design. Proceedings from CHI 2001: *The 9th International Conference on Human-Computer Interaction*. New Orleans: Lawrence Erlbaum.
Stephanidis, C. (2001). User Interfaces for All: Concepts, methods, and tools. Lawrence Erlbaum Associates, Inc. Mahwah, NJ: Lawrence Erlbaum Associates.
Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4*, 295–312.
Sweller, J. (1999). *Instructional design.* Melbourne, Australia: Australian Council for Educational Research.
Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 215–266). San Diego, CA: Academic Press.
Sweller, J., Van Merrienboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review,10,* 251–296.

van Dam, A., Becker, S., Simpson, R. (2005). Next-Generation Educational Software: Why We Need It and a Research Agenda for Getting It. *EDUCAUSE Review*. 40 (2) 26–43.

van Merrienboer, J., & Ayres, P. (2005). Research on Cognitive Load Theory and Its Design Implications for E-Learning. *Educational Technology Research & Development 53*(3), 5-13.

Wallace, R., Soloway, E., Krajcik, J. S., Bos, N., Hoffman, J., Hunter, H. E., Kiskis, D., Klann, E., Peters, G., Richardson, D., & Ronen, O. (1998). Artemis: Learner-centered design of an information seeking environment for K-12 education. *Human Factors in Computing Systems: CHI '98 Conference proceedings*. Los Angeles, CA.