The effects of the variable spatial contiguity of digital vocabulary annotations on reading comprehension by English language learners at American universities

---

A Dissertation

Presented to

The Faculty of the Curry School of Education

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Brendan O. Downey, B.A., M.Ed.

June 2014

Abstract

Since the 1990s, electronic glossary systems have been a popular topic of inquiry for language educators and instructional designers. However, no published research describes a fully replicable toolset for performing empirical glossary user research in applied educational settings, leaving scant evidence about the effects of different styles of glossaries on reading comprehension and other outcomes (Welker, 2010). This study asked if learners would access vocabulary annotation features when offered in an electronic reading system, and if different modes of presentation would affect reading comprehension, according to the spatial contiguity principle (Mayer, 2005). Using off-the-shelf materials adapted for use with the system, students in an English for Academic Purposes (EAP) program (n = 20) at an American university read a short passage with an integrated glossary mechanism, followed by a reading comprehension assessment. Almost half of the participants (45%) opted to utilize the glossary features. No significant differences were observed between the contiguous and non-contiguous versions of the gloss in terms of reading comprehension, gloss clicks, gloss time, or reading time, and no correlations were observed between self-efficacy ratings of glossary and information technology and the reading comprehension or gloss activity measures. Data quality limitations precluded the use of robust inferential statistics. A platform for replications and modifications of the study in English-language academic settings was made available for future research through a freely available Web application and open-source PHP code.

Dedication

This project is dedicated to all people mystified by language.

Acknowledgements

I thank the department of Anthropology, especially Chuck and Nan Perdue and Eve Danziger; everyone at CAELC, especially Dudley Doane and Liz Wittner; everyone at the CTTE, especially Glen Bull; Vonna Hemmler, Sarah Moseley, Linh Li, Missy Maceyko, and Michael Navarrete, my teaching assistants for LING 108 and 111; Patricia Gordon, Yitna Firdyiwek, and everyone at UVaCollab; Rafael Alvarado and everyone at SHANTI; my friends from class, especially Tory Jackson, Michael Kelly, Tom Hammond, Jing Ma, Kathy Neesen, Roger Geyer, Sarah Johnston, Diallo Sessoms, Eunhee Jung, Mark McDonald, Curby Alexander, Jenny Eppard, Rob Oliver, Dan Pryor, Justin Humphries, Johanna Craig, Jie Chao, Yongju Lee, Eric Yoder, Dennis Pipes, Jason Bryant, Henrietta Siemens, and Ilknur Akyildiz; my friends at ASCIT, Steve West, Eric Stauffer, Alejandra Montaño, and Veronica Yeh; my friends at the Curry School ETO, Marc Perdue, Bill Booz, Tom Rose, and Heather Burns; Jo Tyler, Kevin Brady, Carolyn Kaltenbach, Holbrook Mahn, Paul Edmunds, Patricia Reynolds, Nick Ekkizogloy, Carmen Figueroa, Rob Childress, Terry Elvers, Delaney Blom, and Warren Roby, who provided support and encouragement from other campuses; everyone in Fauquier; and, Kay Buchanan and Carole Lohman from the library. I thank my friend Spring Brennan and Gumption Technologies for helping me code the Web site used in this study, and Christina Dendy for proofreading. Finally, I would like to extend special thanks to Ruth Ferree, Valerie Larsen, John Bunch, Mable Kinzie, and Bill Ferster for their mentorship.

Table of Contents

List of Tables

List of Figures

Chapter 1: Problem

In an exploratory pilot study, English language learners (ELLs), or learners whose second language (L2) was English, who were enrolled in a graduate program and receiving services through the Center for American English Language and Culture at the University of Virginia, responded to a survey about their history of dictionary use by describing electronic aids such as CD-ROM-based dictionaries ($N = 60$). In follow-up interviews, participants ($n = 4$) demonstrated a broad array of electronic dictionaries on stand-alone devices (i.e., "e-dictionaries") as well as networked devices with Internet access. Students described using these devices to assist with academic tasks such as reading and writing, and used the devices at home more than at school. Participants expressed a preference for speedy electronic dictionaries over paper dictionaries. Participants who utilized a Web browser for vocabulary support described specific combinations of resources used to understand academic English, including digital English dictionaries, bilingual dictionaries, machine translators, and Google search queries that included the word "dictionary" or "translation". Although participants did not refer to these supports as "glosses" or "glossaries" during interviews, the function of the tools that were demonstrated was, in part, to generate on-demand annotations for words and phrases. It was determined that further exploration of the autonomous use of electronic

vocabulary annotations by learners of English may shed light on the effectiveness of such supports in academic environments (Downey & Eppard, 2010).

Annotation is the term used to describe a note added to a text, whether that text be verbal, video, or audio (Bird & Leiberman, 2001). Electronic annotations may contain sound, images, or animations. Chen (2006) asserts that *annotation* is synonymous with *gloss*. *Gloss* is defined as "any explanatory annotation [in the students' native language] or translation … of a word or phrase" (Roby, 1991, p. 72). A *glossary,* or a collection of glosses, is similar to a dictionary; however, glossaries are generally specific to a certain text. In an historical discussion about dictionaries, Roby (2005) explains that our present conception of dictionaries originates with glosses, which were "localized annotations" of "so-called 'hard words'" (p. 54). Extractions and compilations of glosses gained an independent identity and started to circulate (Roby, 2005). A gloss can contain definitions, explanations, or translations of a term, and glosses can exist in several forms: verbal, visual, and audio (Roby, 1999, as cited in Ben Salem, 2007). In order to acknowledge that a gloss may contain media in forms other than verbal ones, it can be helpful to use the term *annotation*; however, when denoting that an annotation is intended as a lexical support, per se, *gloss* may be most accurately termed *vocabulary annotation* (Nation, 1990, as cited in Chen, 2006), because a gloss can be accessed through an integrated glossary, or by integrating a dictionary or other support into the task of reading.

Research about ELLs and vocabulary supports is an important area of inquiry because vocabulary supports for reading comprehension may impact ELLs' access to

academic English (Cummins, Brown, & Sayers, 2007), and because language proficiency may correlate with grade point average (GPA) and other measures of academic achievement (Thomas & Collier, 1997; Wongtrirat, 2010). Extant research on the topic of glosses from the United States has focused mostly on populations of language learners other than ELLs. As discussed in Chapter 2, only five electronic glossary studies with adult ELLs were performed at U.S. academic institutions between 1991 and 2011 (Ariew & Ercetin, 2004; Al Seghayer, 2001; Chen, 2006; Al Ghafli, 2011; Yoshii & Flaitz, 2002). As more information technologies continue to become available in future years, instructional technologists will need research-based guidelines that assist in effectively matching technologies with individual learners in academic contexts (Kopriva, Emick, Hipolito-Delgado, & Cameron, 2007).

Similarly, instructional designers need research-based guidelines to support the development of media that efficiently communicates with learners, especially as digital media offer more options for learner control of interactive features. Research has shown that learner control is an important requirement for effective computer-based instruction (Kinzie, 1990; Kinzie, Sullivan, & Berdel, 1988, as cited in Roby, 1991). Learners often prefer optional supports to ones that appear in an obligatory manner, even if they have the option to remove the automatically presented support (Schnotz & Heib, 2008). Research in learner control has focused on not only learner utilization of optional features, but also the effects of more specific aspects of learner behavior, such as the length of time a participant spends interacting with optional supports, what information participants choose to interact with, and so on (Ross & Morrison, 1989).

This study designed, developed, and implemented an interface for displaying electronic glossaries, for the purpose of determining whether two treatments, alternate styles of gloss presentation (as illustrated in Figure 1), affect reading comprehension or user behavior, according to the spatial contiguity principle (Mayer, 2005). No published studies from the United States have explored the spatial contiguity principle as it pertains to the use of electronic glossaries by L2 learners of English.

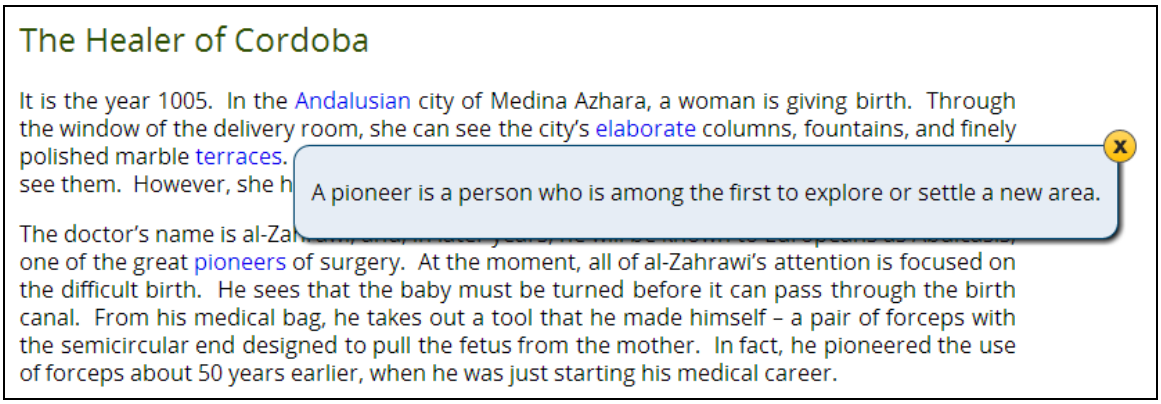


Figure 1. Reading activity with contiguous gloss displayed for the term *pioneer*.

This instructional design principle states that when annotations to on-screen material are positioned close to the associated information, learners' cognitive load can be minimized (Mayer, 2005).

Within the limited amount of research on the topic of electronic glossaries and English language learning, one gap is the lack of a common instrumentation platform. Thus, one aim of this study was to develop materials that were forward-compatible with Web technology standards, and so that the code base was made portable. A learning

management system (LMS) with mechanisms for delivering randomly assigned treatments (i.e., variations on the positioning, or spatial contiguity, of a pop-up glossary) to online participants, along with integrated data collection mechanisms, was developed and implemented by the researcher. Figure 2 shows the home page of the experimental Web application.

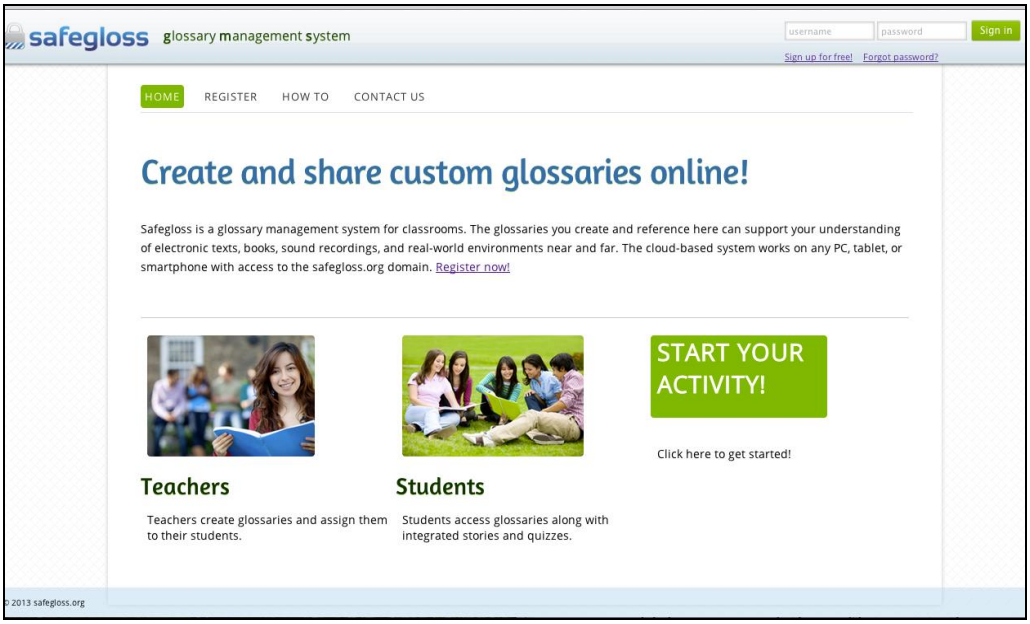


Figure 2. Safegloss.org home page.

Research Questions

Using this system, which is described in Chapter 3, the study addressed the following research questions:

1. Will a group of English language learners utilize hyperlinked vocabulary annotations when they are made available during a reading comprehension activity?
2. Will variations in the presentation of a pop-up glossary window affect reading comprehension, time spent glossing, reading time, or number of gloss clicks?
3. Will ratings of self-efficacy in the use of glossaries or information technology correlate with time spent glossing, reading time, or number of gloss clicks?

The first treatment condition was glossary-near-text (spatially contiguous). In this condition, participants who clicked a word saw a glossary pop-up bubble appear just above the word that was clicked. The second was glossary-in-margin (spatially non-contiguous). In this condition, participants saw the bubble appear in the right margin. Each activity that participants performed on the system was followed by a multiple-choice comprehension test. The score on each of these tests was recorded as a percentage value of correct responses. The first user behavior measure was total time spent reading. For each activity, a timestamp was recorded into the log indicating when the participant began and finished the reading.

The system calculated the differences in these timestamps to yield a measurement of time spent reading. These were recorded into a database, rounded to the microsecond. The second user behavior measure was total time spent using glossary. Each time a participant activated or closed a gloss, the system recorded a timestamp into the log. The time difference between each opening and closing of the gloss was calculated, and all

such durations were added to yield a measurement of total time spent using the glossary. These were recorded into the database in the same fashion.

To provide data on the background characteristics of participants, participants' reports of dictionary and information technology self-efficacy were recorded on 10-point Likert scale survey items. This value was recorded because participants' extant skill with the use of lexical supports could confound the measurement of the effect of the independent variable. The second value recorded for this purpose was perceived technological self-efficacy. As with dictionary self-efficacy, a general level of skill with computer technology could also confound measurement of the effect of the independent variable.

Hypothesis

Based on the rationale that English language learners find academic texts difficult to interpret, partly because of unknown vocabulary items, it was expected that participants would avail themselves of the glossary features when reading. Based on research findings from non-language learning applications, it was expected that some spatial contiguity effect would be observed in user behavior or reading comprehension scores, that is to say, the positioning of the gloss may affect or correlate with measures of the dependent variables.

Contribution to Knowledge

The study makes a contribution to the field in several ways. 1) It provides a longitudinal review of research into dictionary interfaces over a 20-year period, marking

the shift from a largely paper-based paradigm to a multimedia-capable, then Web-based one. 2) It describes a novel instrumentation that capitalizes on the features described in previous dictionary research systems, as well as current affordances in Web development technologies. 3) A platform for future replications of this quasi-experiment was made live as a freely available Web application, and the source code was donated to the research community using software repository GitHub so that it can be validated or modified (forked). 4) It presents original findings on learner control, specifically the relationship between the spatial contiguity of glosses and reading comprehension.

The next chapter will review the literature relating to the four topics outlined above: 1) vocabulary and reading comprehension; 2) the spatial contiguity principle; 3) learner control; and, 4) research about electronic glossaries and reading comprehension conducted from 1991 to 2011.

Chapter 2: Review of the Literature

Chapter 1 briefly discussed how limited English language proficiency can constitute a barrier to academic achievement. In order to understand how glossaries relate to literacy, this chapter will begin with a discussion about vocabulary and reading comprehension. Because glosses, like all annotations, are subject to questions of relative positioning, the spatial contiguity principle (Mayer, 2005) will be explored second. The topic of learner control, which relates to students' utilization of supports like glossaries, will be explored third. Finally, the chapter will review research studies from applied settings that tested electronic glossary systems with language learners. The review includes studies published from 1991 to 2011, a period during which cognitive theories of multimedia developed, and when the proliferation of information technologies made electronic glosses, or vocabulary annotations, a popular topic of inquiry.

Vocabulary and Reading Comprehension

As one of the subordinate skills of reading comprehension, a reader's breadth and depth of vocabulary knowledge can contribute to how well he or she understands a text (August, Carlo, Dressler, & Snow, 2005; Brown, 2010; Golkar & Yamini, 2007; Rashidi & Khosravi, 2010). Reading comprehension is a complex skill that depends on vocabulary knowledge. Only when a reader can decode a word accurately can he or she

comprehend clauses and paragraphs (Lesaux, Kieffer, Faller, & Kelley, 2010). Vocabulary is of particular importance to ELLs for several reasons. 1) Learners consider vocabulary errors to be the most serious kind of errors. 2) Lexical errors are the most common among language learners. 3) Native speakers find lexical errors more disruptive than grammatical ones. 4) The lexicon is an important aspect of comprehension. 5) Learning words is a recursive process that does not happen instantaneously (Gass and Selinker, 2001, p. 372). The importance of vocabulary knowledge to reading comprehension would seem to give the topic high priority in curriculum and instruction, but opportunities for vocabulary learning can be scarce (August et al., 2005).

Providing a dictionary is one strategy to support the cognitive processes involved in vocabulary acquisition: attention to form-meaning connections, rehearsal of words for storage in long-term memory, and elaboration of associations with other knowledge (Fraser, 1999, p. 73). However, students who are not skilled with dictionaries or glossaries may not benefit from them, and may even find them detrimental (Chun & Plass, 1996; Jones & Plass, 2002; Plass, Chun, Mayer, & Leutner, 2003). One problem with dictionaries and glossaries is that they require the reader to look away from their material, which may cause him or her to become distracted (Ali Farhan, 2011). This consideration applies to all annotations, whether they are paper or electronic, and whether the annotation is integrated or separate. This generalization is based on the cognitive theory of multimedia learning.

Cognitive Theory of Multimedia Learning

This theory describes how human senses interact with memory (Mayer & Moreno, 2002). The tenets of the theory provide the basis for the spatial contiguity principle (Mayer, 2005), the theoretical basis for the research questions of the present study. In this theoretical model, which is drawn from research in psychology, memory is described as short-term and long-term. During instruction, stimuli activate sensory registers in the brain. Because short-term memory is limited, the amount and type of stimuli experienced by a learner may overload its capacity. This load is called "cognitive load" (Sweller & Chandler, 1991, p. 294). Thus, according to this theory, for new knowledge, skill, and ability to be transferred to long-term memory (i.e., the goal of instruction), the instructional experience must not require a learner's short-term, or working memory, to process more than its capacity allows.

The integration of multiple modalities during instruction (i.e., the engagement of more than one of the senses) risks causing this cognitive overload if the multimedia elements are not appropriately coordinated (likewise with a plurality of elements of a single modality, e.g., text on text). This phenomenon has been studied as a construct called the "split-attention effect" (Chandler & Sweller, 1992, p. 233). "Split attention" can be broken down into types (i.e., spatial contiguity, temporal contiguity, modality, etc.) (Mayer & Moreno, 2003). Corresponding effects, spatial contiguity, coherence, and multimedia effects, have been observed during individuals' interactions with media of all modalities (Mayer, 2003). The observance of these effects across a range of studies supports the cognitive theory of multimedia learning (Mayer & Moreno, 2002), and has

led to the naming of a set of principles that can be used to describe the best practices (i.e., those that minimize cognitive load, that is, which avoid these generally undesirable effects). For instructional designers, these principles can serve as guidelines for the positioning of annotations (i.e., the focus of the present study), but also for decisions on timing, the integration of sound, images, and animation, the pacing of a presentation, and so on.

Spatial Contiguity Principle

The spatial contiguity principle states that placing related elements together on a display, rather than apart, can reduce the cognitive load that is introduced by diverting attention across space to the associated element, or what is a type of split-attention, or spatial contiguity, effect (Mayer, 2005). Mayer's 1997 review of the effectiveness of multimedia instruction found consistent evidence of a spatial contiguity effect, an outcome also found by others (Chandler & Sweller, 1991; Mayer & Anderson, 1992; Moreno & Mayer, 1999; Paas & Van Merriënboer, 1994; Sweller, Chandler, Tierney, & Cooper, 1990). While electronic glossaries are a popular topic of inquiry around the world, no study from the United States has explored the spatial contiguity principle (Mayer, 2005) in the context of the use of electronic glossaries by L2 learners of English while reading (Welker, 2010).

Learner Control

Scheiter and Gerjets (2007) distinguish between system-controlled multimedia learning environments and hypermedia systems, in that the latter is characterized by a

high level of interactivity, or *learner control* (p. 285). Whereas multimedia environments are often characterized as linear, hypermedia environments imply non-linear paths of information access (Scheiter & Gerjets, 2007, p. 286), meaning that learners may control their path through information. This preference can be exercised in several ways: by determining the sequencing of information, the selection of contents, or the representation of information (i.e., whether it is verbal or pictoral) (Scheiter & Gerjets, 2007, p. 287).

Learner control, when it impacts learning, can either improve or degrade instructional effectiveness, depending partly on how students exercise that control. On one hand, providing additional help only when a student requests it may prevent unnecessary distractions, but on the other, students may not be skilled in the use of embedded help systems such as glossaries, in which case, the on-demand nature of the assistance may prove to hide the instructional support that the student needs. This has been termed an "assistance dilemma" (Koedinger & Aleven, 2007, p. 239). In a study of glossaries conducted with students in a geometry class, fewer than half of the participants utilized a glossary, to the surprise of the researchers (Aleven & Keodinger, 2000).

Five main potential advantages of hypermedia environments have made exploring their educational effectiveness a point of continued interest in the research community: 1) that linking structures of hypermedia environments resemble the associative nature of the mind; 2) that learner control can increase interest and motivation; 3) that hypermedia environments allow for adaptation to preferences and cognitive needs; 4) that hypermedia provides affordances for active and constructive information processing; and that 5)

hypermedia may promote the acquisition of self-regulatory skills (Scheiter & Gerjets, 2007, p. 288).

In one study that tested the effect of learners' control of the pacing of a presentation, it was found that the group who had control of the pacing outperformed a control group (Hasler, Bernd, & Sweller, 2007). In another study, learner control resulted in longer instructional times, which was thought to render the learning experience inefficient, although outcomes for the group that exercised more learner control were significantly better than for the group that exercised less (Gerjets, Sheiter, Opferman, Hesse, & Eysink, 2008).

Although research has explored the extent to which learner control is a factor in the effectiveness of computer-based study, the effectiveness of self-controlled learning with hypermedia is difficult to demonstrate due to 1) usability problems such as disorientation, distraction, and cognitive overload, 2) moderating learner characteristics such as prior knowledge, self-regulatory skills, cognitive styles, and attitudes towards learning, 3) lack of conceptual foundations, and 4) methodological shortcomings of hypermedia studies (Scheiter & Gerjets, 2007, p. 285).

Electronic Glossary Research: 1991–2011

Since the 1990s, when the first electronic dictionary user studies were conducted, personal computers have become fixtures in many language education settings. Instructional contexts related to the teaching of language have been an ideal place for the application of computers, because language instruction, when it addresses all four of the skills of listening, speaking, reading, and writing, is an inherently multi-modal subject;

that is, language-learning exercises adapted for computer-based study lend themselves to the application of multimedia, or media that include text, sound, images, and animations; moreover, language-learning is both receptive (listening and reading) and productive (speaking and writing), making it an ideal venue for the implementation of technological capabilities related to the authoring of interactive presentations, or those that allow teachers and learners to provide their own input to software systems (in contrast with the static, paper-based paradigm which had been the norm).

Yet even as computers' ability to produce rich combinations of text, sound, and images was making an impact on language education in the late 1980s and early 1990s, the advent of the Internet in the mid-1990s brought new ways to interconnect these ever-more-capable machines. The field of dictionary and gloss research, like the broader field of education research (and indeed, research at large) has undergone a transformation that has been made possible by the proliferation of networked information technologies. Whether to deliver treatments, measure behavior, or calculate outcomes, new technologies have afforded research studies the potential for increased rigor. For example, Larsen (1997) parsed and interpreted the Web server logs for the first online interactive frog dissection simulator, which was accessed by tens of thousands of individual client devices via the Web (Kinzie, Foss, & Powers, 1993).

Still, as the need for more effective educational solutions for ELLs in the United States has persisted, only five electronic dictionary studies with adult ELLs were performed at U.S. institutions between 1991 and 2011 (i.e., Ariew & Ercetin, 2004; Chen, 2006; Yoshii & Flaitz, 2002; Al Seghayer, 2001; Al Ghafli, 2011), and of those, one

sampled from a population outside of the United States (Al Ghafli, 2011) and two did not measure reading comprehension as a dependent variable (Al Seghayer, 2001; Yoshii & Flaitz, 2002) as the present study did. Thus, there is a scarcity of comparable studies. To get a broader, albeit less-focused perspective, seven electronic glossary studies conducted with non-ELL populations are also included in the literature reviewed in this chapter. All twelve studies are methodologically and ecologically comparable in that the annotations displayed on the experimental user interfaces were activated by a hyperlinked word, and that the populations were made up of language learners. The studies differ in terms of what kinds of media were contained in the annotations. In addition to verbal annotations, which were activated through various forms of user interaction, some programs contained a short animation, audio (Akbulut, 2007; Ben Salem, 2007) or translations in their glosses (Roby, 1991; Aust, Kelley, & Roby, 1993; Ben Salem, 2007). And while the glossary information presented to participants in these studies was placed in various locations on the screen, spatial contiguity, per se, was not a discretely measured variable.

However, the studies are also comparable in terms of their theoretical frameworks. Whereas the first electronic dictionary study (Roby's foundational 1991 work, described in this chapter) did not reference theories of multimedia, as those addressing digital multimedia were not yet well developed, Ben Salem's dissertation study, published 16 years later (2007), cites Mayer's generative theory of multimedia learning as the relevant theoretical framework. Paivio's dual-coding theory and cognitive load theory (CLT), also popular theoretical frameworks for annotation studies (e.g.,

Chen, 2006), were both effectively re-worked as the generative theory of multimedia learning (Lin & Liu, 2010).

Perhaps the most important similarity between the studies involves replicability; that is, while electronic glossary studies have seemed to successively build upon one another over time, they have not been able to form a common methodological platform. That is to say, almost every new study has produced an original system for housing and delivering the materials used for reading activities and assessments, and with the exception of Aust, Kelley, & Roby (1993), no study has recycled the instruments developed from any previous study (the system used in Roby's 1991 study were used again in the Aust, Kelley, & Roby, 1993 study). This lack of a common platform has made it impossible for researchers to replicate each other's work. Studies such as Chen (2006) demonstrate the affordances offered by the Web as an option for delivering variable treatments to participants. Yet no study in this review has made the full complement of its materials available to general user registration as a "live" Web site supporting both the development and consumption of instructional design experiments involving glossaries and dictionaries, nor as a shared code base. This is likely, in part, because of the longtime lack of open standards for the authoring of Web media, and because of the unavailability of easy-to-use content versioning and shared storage mechanisms.

The popular technologies used to author multimedia for the Web historically have been based on closed-source platforms like Adobe Flash, meaning that any repository suited for the materials of many previous studies, had one existed, likely would have

received material unable to be readily examined or adjusted. Likewise, even as studies like Chen (2006) utilized standards like HTML and JavaScript to develop their experimental interfaces, no widely adopted service allowed archiving and sharing software code with others. The most popular "social coding" platform that serves as a public repository of this type, GitHub, was not established until 2008, and its forerunners were not widely used.

The absence of code-sharing mechanisms and open multimedia authoring standards did not prevent researchers from using multimedia platforms to integrate data collection functionality into interactive experimental glossaries. The first electronic glossary study published in the United States was a 1991 dissertation by Warren B. Roby, then a doctoral student at the University of Kansas, who used an early model of the Apple Macintosh to develop a stand-alone experimental multimedia glossary system. This study pre-dates the Web by two years, as Mosaic, the first Web browser, was not introduced until 1993.

Roby (1991). The study by Roby (1991) addressed whether combinations of semantic support (dictionary, gloss) and presentation mode (paper, desktop computer) affect reading comprehension or speed of reading. The participants were 95 L2 learners of Spanish enrolled in Spanish 206 at the University of Kansas. Treatment conditions were gloss and dictionary (paper); gloss and dictionary (electronic); gloss only (paper); and gloss only (electronic).

The materials were an interactive activity built by the researcher using HyperCard, a programming platform organized around the concept of slides. The system

was programmed to activate a glossary entry when target words were clicked, to count how many times each word was clicked, and to measure for how long the participant used the gloss. The system also calculated the overall time that the participant spent reading the passage. Participants came to one of two classrooms, either a computer lab or a normal classroom (one for each of the presentation mode treatments). All participants completed a background questionnaire that gathered descriptive elements such as major area of study, gender, and age. Participants clicked through an on-screen text presentation comprising a biographical sketch in Spanish. While reading, participants had access to one of two kinds of semantic support: either a dictionary and a glossary or a glossary only. The dictionary items were extracted from off-the-shelf dictionaries. When finished with the activity, participants were presented with a reading comprehension test. Some participants were interviewed for their thoughts about the user interface and the semantic supports.

The dependent variables were number of queries, comprehension, and reading time. Measures were recall protocols, which are writing prompts that ask the participant to write down each thing they remember from the story that they read, as well as reading duration, as logged by the software (see Figure 3).

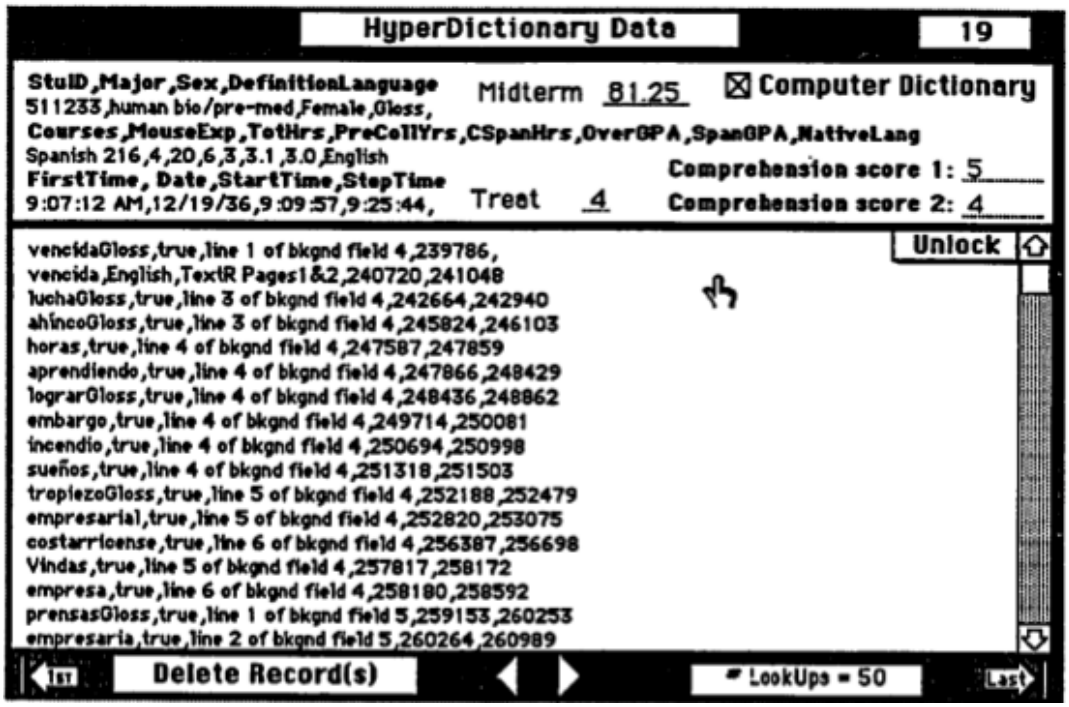


Figure 3. Data collection card providing log function used in Roby (1991).

Roby included a covariate in the analysis, the students' grades in Spanish from the previous mid-term examination, to control for any major differences in the Spanish proficiency of the participants in the two rooms. It was found that the group with access to both the dictionary and the gloss read the passage in significantly less time (M = 18:40) than those with the dictionary alone (M = 21:24) (F = 4.62, p = .034). Also, participants in the desktop computer presentation conditions looked up significantly more words (M = 49.98) than participants in the paper conditions (M = 17.15) (F = 63.69, p = .000). These findings have a high degree of face validity, in that it would seem likely that electronic glosses, being easier and faster to use than paper versions, would promote more consultation and would correlate with shorter reading times. Comprehension,

however, was not significantly different across the two media. Qualitative data indicated higher satisfaction with the digital forms of semantic support than with the analog counterparts, which is also to be expected given that the electronic dictionaries were likely easier to use. This also is sensible because less difficulty would seem to lead to more satisfaction.

This study is important because it combined elements from the corpus of dictionary user studies back to 1955, inclusive. Some of the important improvements that Roby made to previous studies were 1) reporting of a design that was replicable, 2) a quasi-experimental methodology, 3) electronic authoring tools and delivery platforms, and 4) a comprehensive, worldwide historical literature survey. Several of the methodological components of this study were included in later studies (notably, Ben Salem, 2007), as discussed in the following sections.

Aust, Kelley, & Roby (1993). The Aust, Kelley, and Roby article published in 1993 was the report of a study conducted in 1990, and in Roby (1991) is referred to as a 1990 forthcoming paper. The study addressed the question of whether there would be any effect of two kinds of reference media and definitions in two languages upon the dependent variables of frequency of consultation, reading time, or comprehension. The participants were 80 L2 learners of Spanish. Treatment conditions were two kinds of reference media (electronic, paper) and either monolingual or bilingual definitions. The materials consisted of a custom-made HyperCard presentation that automatically logged usage data, including counts of which words were clicked and how often, and measures for how long words were consulted. Gloss contents were provided by the researchers,

whereas dictionary contents had been extracted from off-the-shelf dictionaries. Corresponding materials were also provided in paper form, depending on which condition was assigned to the participant. Participants arrived at a computer lab during their regularly scheduled class, where they were divided into four groups.

Participants read a passage on the computer screen that was accompanied by an available gloss and/or dictionary (depending on the participant's treatment group). Participants actively consulted the gloss and dictionary, and those in the paper treatments noted which words caused them to consult the gloss or dictionary. Time spent consulting the gloss and dictionary was recorded for all participants, either manually for the paper group or automatically for the electronic group. The dependent variables were the following: frequency of consultation, reading time, number of consultations, and comprehension. The measures consisted of a written recall protocol (from Bernhardt, 1983) that asked participants to write down everything they could remember about the text, and a short survey connected to the software. Reading time was measured by a function within the software that logged data about user behavior. Grade point average (GPA) in Spanish was used as a covariate.

It was found that participants consulted the dictionary more frequently when it was electronic (M = 28.3) versus the paper version (M = 13.1). In addition, users accessed the bilingual version of the dictionaries 25% more often (M = 23.6) than the monolingual ones (M = 17.7), $F(1, 76) = 4.09$, $MS_e = 170.28$, $p < .05$). In terms of comprehension, as measured by mean number of recalled ideas from the text, the hyper-reference dictionary users (M = 10.95) did not demonstrate a significant difference compared to the paper

version (M = 12.65), $F(1, 76) = 1.21$, $MS_e = 47.74$, $p = .28$. No significant difference was found in comprehension when users of bilingual dictionaries (M = 12.45) were compared to users of monolingual ones (M = 11.15), $F(1, 76) = .07$, $MS_e = 47.74$, $p = .80$. Post-hoc analysis showed a positive correlation between participants' comprehension and overall GPA, as well as their GPA in Spanish courses, $r = .35$, $p < .01$.

In its suggestions for further research, the report mentions incidental vocabulary learning, which was included in Knight (1994) and several subsequent electronic glossary studies (Akbulut, 2007; Chen, 2006; Yoshii & Flaitz, 2002). Furthermore, the authors suggested using a longer reading passage so that a more sensitive measurement of the total time spent reading could be acquired. Table 1 illustrates the factorial design of the 1993 (i.e., 1990) study.

Table 1

Factorial Design of Treatment Conditions Described in Aust, Kelley, and Roby (1993)

	L1	L2
Paper	Paper/L1	Paper/L2
Electronic	Electronic/L1	Electronic/L2

Despite bolstering the measurement by using multiple raters and calculating inter-rater reliability, a multiple-choice test may more reliably measure differences in performance between subjects. In addition, the proposition recall measure elicits a writing performance by the participant, and if skill in the domain of writing was not measured specifically in terms of participants' background variables, then variation in writing skill

could confound measurement of reading comprehension. The participants in Al Ghafli's study (2011) were given a multiple-choice reading comprehension test because the researcher had reason to believe that the writing proficiency of the participants was substantially lower than their reading proficiency. Similarly, the treatment conditions involving the paper modes of support may have been subject to measurement error. It seems likely that it was more difficult for a person to accurately log these details when compared to software calculating the same things automatically.

Despite the potential of measurement error in the proposition recall measure versus the multiple-choice option, a proposition recall may be superior to a multiple-choice measure in terms of construct validity. The proposition recall measure allows the evaluators to consider the entire range of possible responses made by participants (within the constraints of the finite number of idea units established by the raters), whereas the multiple-choice measure only measures the responses prescribed in the test items. Thus, to the extent that the test items on the multiple-choice may not measure every element of the participants' comprehension of the test, the proposition recall measure could have superior construct validity. Finally, a no-treatment group could have been employed to make the study more closely resemble a true experiment; however, this may have been impractical in the specific academic context. The next study in this stream of research parcels out vocabulary acquisition as a constituent element of comprehension and measures it discretely, which may be sensible in that each of these studies tacitly asks whether knowledge of individual word meanings impacts comprehension.

Knight (1994). The study by Knight (1994) addressed the question of whether there would be a significant difference in vocabulary and reading test scores whether or not participants see new words in a list or in the context of a text and when they have access or do not have access to an on-screen electronic dictionary. The participants were 105 L2 learners of Spanish at Central Michigan University. Treatment conditions were access or no access to a dictionary and access or no access to a text that shows the word in context. The materials were all contained in a multimedia presentation that was saved to floppy disks, with different versions of the presentation corresponding to different treatment conditions. Participants picked up a disk at the front of the computer lab and loaded it at an individual workstation. Each disk contained text and an electronic dictionary (if applicable). The materials were custom-made by the author, but entries were extracted from an off-the-shelf dictionary.

The dependent variables were incidental vocabulary acquisition and reading comprehension. Measures were a recall protocol and an immediate and delayed vocabulary test. The recall protocol is a system whereby participants write down all ideas remembered from the text that they read, and then, raters assign a score according to the number of idea units expressed by the participant in terms of the finite set determined a priori by the raters. The vocabulary tests were multiple-choice. Subjects with dictionary access learned the most; all participants learned more when they were exposed to the words in context; all participants were able to correctly supply meaning of words without first being able to see the targeted words in context. This study utilized a similar method to Roby (1991) for counting the amount of information remembered, which was a

weighted proposition analysis. The proposition analysis refers to the scoring of the written recall measure. These measures generally require raters to define discrete bits of information within a text, and then compare participants' written recall to those bits of information, scoring the writing accordingly. The study was unique in that it included high or low reading ability on the part of the individual participants as a variable of interest. Likewise, Knight employed a computerized mechanism for logging data about the gloss consultations. Knight found a significant difference in reading comprehension between the dictionary group (M = 74.01) and the no-dictionary group (M = 56.65, p = .739). While this study adds a measure of vocabulary learning to Roby's template, it, like Roby (1991), utilizes a written recall measure to evaluate reading comprehension. As discussed earlier, this measure may produce measurement error because of the presence of subjectivity in rating the test, as well as the potential for the writing aspect of the test to produce measurement error because of the potential for variations in writing skill to confound the accuracy of the measurement. A strength of Knight's study is the large number of participants, which can mitigate selection bias and increase power.

Al Seghayer (2001). The study by Al-Seghayer (2001) compared the effect of different multimedia annotations on vocabulary acquisition. Three treatment conditions– text alone, and text with one of two annotations, video or still image—were implemented. The participants were 30 participants enrolled at the English Language Institute at the University of Pittsburgh. Participants were native speakers of several languages with TOEFL scores ranging from 450 to 500. Materials consisted of an HTML reading activity that was published to CD-ROM. The reading activity allowed for the clicking of

hyperlinks to activate gloss features, including pronunciation. Measures consisted of two kinds of vocabulary tests, administered after the reading activity by surprise, as well as post-experiment surveys and interviews. The receptive vocabulary test displayed an image and asked participants to choose the correct response from a list of options. The productive test asked participants to respond to items by typing into a text-input area. The results from the test were automatically transmitted to the researcher by email. The target vocabulary words, 21 in total, were a mixture of verbs, nouns, adjectives, and adverbs. Five words had text-only annotations, and five had text and an accompanying image. Four participants completed a Cloze exercise to determine whether the words would be unknown. With a mean score of 65% on the Cloze activity, the reading was deemed appropriate for the intermediate level. The reading, addressing a topic about Native American culture, and thereby thought to be equally novel to all participants, was 1,300 words. The materials contained an integrated "tracking device" that ensured that no support other than the annotations could be accessed by participants.

This study leveraged the capabilities of the Web browser not in terms of networking over a wide area, but in terms of its multimedia capabilities, specifically, to leverage the fact that Web standards for multimedia provided a way to affordably reach learners on either Mac or PC platforms with no problems of compatibility. The procedure for assigning participants into groups for each treatment were not made clear in the report: it is difficult to know whether three versions of the CD-ROM were distributed throughout the group of 30, or if all three treatments were applied to each participant simultaneously. The report also is not clear about the manner in which multimedia

annotations were used in the multiple-choice test versus in the reading activity. As limitations, the report lists the small sample size, the lack of log data indexing user behavior, and the use of only multiple-choice tests. To analyze the data, an alternative to repeated-measures one-way ANOVA, the Friedman test, was used. The study found that a video clip was more effective than a still image for learning new vocabulary.

Yoshii & Flaitz (2002). The study by Yoshii & Flaitz (2002) addressed whether three types of multimedia annotations would affect vocabulary retention. While this study departs from the others reviewed here in that it did not measure reading comprehension, it is worth reviewing because it is one of only four electronic gloss studies from the post-1991 period that were conducted with ELLs in the United States. The participants were 151 adult L2 learners of English at five Florida universities. The participants consisted of beginning and intermediate-level ELLs enrolled in an intensive English program. The independent variables were three kinds of multimedia glosses: text-only, picture-only, and text and picture. The dependent variable was incidental vocabulary learning. Measurements used for the dependent variables were three kinds of vocabulary tests: definition supply, picture recognition, and word recognition. Although there was no measure for reading comprehension in the study, participants were told that they would be tested on comprehension. In order to test for incidental vocabulary learning, some studies present a vocabulary test as a surprise. This study did so, but opted not to test for comprehension as expressed. The authors mention, however, that a brief comprehension check was administered immediately after participants completed the reading task. Participants took a pre-experiment vocabulary test to check for prior knowledge of the

target words used in the glosses. Two participants knew several of the words and were excluded from the remainder of the study. All the target words were verbs, and those words were compared with words from native languages of the participants to ensure that there were no cognates, which could make guessing word meanings easier for some L1s than others. The materials were custom-made by the authors, with the reading passage being authored previously by one of the researchers and the pictorial annotations being developed by a graphic artist. The contents of the gloss were reviewed by non-participating participants to ensure that they effectively conveyed the meaning of the target word.

Participants were stratified by L1 and randomly assigned into treatment groups for the experiment using pre-determined code numbers that corresponded to individual computer workstations also bearing code numbers. Participants were given an introduction to the exercise, and completed a consent form and a questionnaire asking for gender, age, and extent of English study before moving to the United States. Another section asked about attitudes toward computers and familiarity with them. During a subsequent session, participants accessed the hypertext reading with access to the glosses. Vocabulary tests were given immediately after the reading and also two weeks later. Data were analyzed with a 3 x 2 ANOVA, with annotation type and language proficiency level as the variables. With respect to immediate vocabulary retention, the group accessing a combination of picture and text annotations outperformed the other groups on all of the vocabulary test types. There was a significant difference found among the groups for the Picture Recognition Test $F(2, 145) = 4.04$, $p < .05$. The Combination group

outperformed the Text-only group significantly ($p < .0167$). The differences in proficiency levels did not yield an interaction effect as measured on any of the test types. In other words, the Combination group outperformed the other two groups irrespective of language proficiency level.

The report gives no information about the procedure employed to determine the suitability of the text used for the reading exercise. Therefore, it is left to speculation whether the lack of any effect of language proficiency on the dependent variables could have been due to the fact that the text was too challenging for the beginner students, for example. No information is given on the instrument used to measure language proficiency. This would make it impossible to replicate this study precisely. Descriptive statistics about the responses on the questionnaire were not given, but were reportedly recorded.

Ariew & Ercetin (2004). The study by Ariew & Ercetin (2004) addressed the question of whether two kinds of annotations (contextual and textual) affected reading comprehension. It also asked if annotation use would be a predictor of reading comprehension for advanced level learners, and if prior knowledge about a topic was an important predictor of reading comprehension for either intermediate and/or advanced level learners.

The participants were 84 English for Academic Purposes (EAP) students, with a variety of cultural and linguistic backgrounds, at the University of Arizona. Participants were in nine English classes, with 34 participants designated as intermediate and 50 designated as advanced, based on an institution-specific language proficiency instrument

called the Comprehensive English Language Test (CELT). Treatment conditions were two versions of annotations integrated into a computer-based reading activity. The dependent variables were the following: total amount of time spent on textual annotations (seconds), time spent on contextual annotations (seconds), and reading comprehension assessment scores. These were tracked using the hypermedia reading software, and recorded into log files saved to each participant's local machine (i.e., hard drive).

In a pilot study, the reading material was evaluated for its difficulty by asking readers to underline the difficult words in the text and then respond to interview questions about the difficulty of the text. The annotated words for the study were determined partially by the results of this procedure, with the words most frequently cited as difficult by the readers making up one part of the glossary. Other annotations were created according to the researcher's judgment about which terms were most crucial in understanding the text. The textual annotations were pop-up windows that appeared in the left margin, giving a brief explanation of the word. The contextual annotations were activated by buttons located at the bottom of the page and labeled with topics from the reading.

The reading comprehension measure was a 21-item test comprising both multiple-choice and short-answer items. The questions were differentially weighted, with the multiple-choice questions given a value of two points, and the short-answer items given a value of one point. To minimize a testing effect, prior knowledge was measured before the reading exercise by asking questions about the topic that were not addressed in the reading. A background questionnaire was given to participants after the reading exercise;

it asked participants to provide details about their level of skill with computers, their thoughts about the experimental software, and demographic items. Participants spent two class sessions in a computer lab completing a questionnaire and learning about the software and how to use it, and then completing the activity. Follow-up interviews were conducted within three days.

The study used multiple regression analysis, with reading comprehension as the criterion and amount of time spent on annotations and prior knowledge as the predictors, to determine whether a relationship existed. Using an independent-samples *t*-test, it was found that the advanced group performed significantly better on the reading comprehension test than the intermediate group, $t(82) = -2.5$, $p = .01$.It was also found that prior knowledge (accounting for 28 percent of variability, $F(1, 33) = 12.50$, $p < .001$), time spent on contextual video (accounting for 10%, $F(2,33) = 4.32$, $p < .001$), correlated with comprehension. No relationship was found between advanced learners and reading comprehension. The researchers concluded that the availability of resources may have proven distracting, and compare their findings to those of Aust, et al (1993).

The researcher admits that the lack of random assignment and the absence of a control group were limitations to the study. This study adds to the research base by providing an example of how to integrate an investigation of the relationship between ability level and reading comprehension into the study of hypermedia annotations. It is not comparable to the present study in that it did not isolate the annotation type and explore the effect of its presentation on reading comprehension. Finally, the study did not

examine the differential effect of annotation types, but rather their combined effect across two ability levels in terms of other background characteristics.

Chen (2006). The study by Chen (2006) addressed whether two different types of multimedia annotations would affect immediate vocabulary recall and reading comprehension. The participants were 78 L2 learners of English from three U.S. universities. Treatment conditions were text/picture annotations and audio/picture annotations in incidental and intentional learning conditions. These two learning conditions are differentiated according to the learning task, learner attention, and the pedagogical context of the learning (Read, 2004, as cited in Chen, 2006). Incidental learning can be tested by withholding the ostensible purpose of an exercise, for example, using a surprise test to measure the construct of interest (Chen, 2006). By contrast, intentional learning is measured by providing a posttest that matches the expressed purpose of a learning activity (i.e., it is not a surprise to the test-taker) (Chen, 2006). The materials were a hypermedia reading on the Internet with integrated annotations. Participants interacted with an HTML and JavaScript slideshow with access to a gloss. The dependent variables were reading comprehension and vocabulary learning. Measures were a modified version of the Vocabulary Knowledge Scale, a word-recognition test (a multiple-choice test asking to identify the correct meaning of a word), a written recall protocol, and a multiple-choice test for reading comprehension. This study is relevant because it used a slideshow interface and glosses to test research questions about reading comprehension.

Results of the reading comprehension measure did not show a significant effect for annotation type (text and picture vs. text and audio) for either comprehension measure (i.e., the reading comprehension test, $F = .44$, $p = .5089$, or the written recall, $F = 3.38$, $p = .0700$). Participants in the intentional learning condition had higher scores in vocabulary recall than those in the incidental learning condition $F(1, 74) = 19.16$, $p < .0001 < .05$). No effect for learning condition was found on reading comprehension ($F = .66$, $p = .4209 > .05$). The intentional learning condition did produce a significant effect on the written recall measure ($F = 8.97$, $p = .0037 < .05$).

Chen acknowledges a weakness of the study that regards internal validity: "One possible threat to internal validity was selection of participants. [Those] in different experimental groups might not be functionally equivalent in respect to their knowledge of the target words" (2006, p. 12). The use of several experimental groups who are different cohorts portends several other problems in addition to possible variations in prior vocabulary knowledge, including varying degrees of English or L1 proficiency, varying levels of experience with online environments, and varying experience with multimedia annotations (i.e., glosses). Moreover, the variety of L1s present in U.S. English contexts may also variably impact the dependent measures. Several demographic items were used in a questionnaire given to participants before the experiment, and were reported, but these variables were not a part of the analysis. Chen implemented random assignment to the four treatment groups, and posited that potentially moderating variables such as age, gender, and computer familiarity were effectively mitigated by the randomization (Chen, 2006).

One strength found in the design is that the written recall measure asked participants to write down their understanding of the text in their native languages. This may have minimized any concern about variations in English writing skill; however, it likely introduced some increased difficulty on the part of raters, who were tasked with interpreting and standardizing scores for the measure across several languages. This may transfer the error from the participants to the raters, as their varying degrees of language proficiency across the number of participants' L1s may also produce error. The written recall protocol was strong in that it granulated the reading passage into 52 idea units, which is a high degree of resolution for such a short reading passage (i.e., 425 words). Another strength is found in the data analysis procedures. The researcher calculated Chronbach's alpha on the dependent measures with a brief pilot study involving participants from the cohort but not involved in the study. Likewise the suitability of the reading was confirmed using a Cloze procedure to determine what portion of the text was likely comprehensible by the participants according to their (intermediate) English language proficiency level. A Cloze passage is one in which key words are omitted and replaced with a blank. This aspect of the study is addressed similarly in Ben Salem (2007).

Ben Salem (2007). The study by Ben Salem (2007) addressed the influence of electronic glosses on reading comprehension and word retention. The participants were 93 L2 learners of Spanish at the University of Kansas. Treatment conditions were five electronic glosses: 1) bilingual text and bilingual audio; 2) audio and picture; 3) translation, audio, and picture; 4) translation, audio, and picture; and 5) handwriting the

gloss, which refers to a process by which the participant copies the gloss entry in his or her own hand. The materials comprised an interactive hypermedia slideshow with variations of the glosses depending on treatment group. The materials were authored in Macromedia Dreamweaver. The reading passage was custom-written by the author, and the glossed words were chosen using a two-step process. First, a list of 75 pictures was taken from a previous study (i.e., Szekely et al., 2003) and presented to the Spanish faculty to be rated for their difficulty. After averaging the scores, only items with a score of seven or higher were retained. Then, those words were tested with participants using a Cloze reading. Participants completed a questionnaire asking about language learning anxiety as well as perceived skill with computers and perceived proficiency in the target language. The questionnaire also asked participants to self-report their native languages and whether they had a parent that was a native speaker of the target language. Participants read an electronic text with access to glosses, and took vocabulary and reading comprehension tests. The dependent variables were word retention, reading comprehension, and time spent reading. Measures were multiple-choice immediate and delayed vocabulary tests (two types, receptive and productive) and a multiple-choice reading comprehension test (which had been tested for internal consistency using Chronbach's alpha, equaling .79). Participants were randomly assigned to treatment groups using a random number generator. Only 63 of the 93 participants completed the post vocabulary test because one of the partnering instructors lost their materials. Analysis was a 5 x 2 ANOVA, with treatments as a between-group factor and pretest vs. posttest (i.e., "time") as a within groups factor. Bonferroni was used in post-hoc analysis.

It was found that participants who used glossed words had higher reading comprehension and vocabulary scores (M = 17.2) than non-glossers (M = 9.2), (F = 82.2, p > .001). Ben Salem (2007) also found that the number of times a gloss was used was positively correlated to comprehension (r = .40, p > .002) and vocabulary acquisition (r = .42, p > .000). The study materials had been piloted with a group of six students to check for usability. Likewise, two instructional designers checked the materials.

Akbulut (2007). The study by Akbulut (2007) addressed whether different hypermedia glosses had an effect on incidental vocabulary learning and reading comprehension. The participants were 69 L2 learners of English at a Turkish university. Treatment conditions were text-only annotations, picture and text, and video and text. The study used a 3 x 3 design, with annotation and time of test being the independent variables. The materials were an interactive slideshow that logged data points on user behavior. Participants took a pre-treatment vocabulary test. Next, they took a background questionnaire that asked about prior topic knowledge, topic interest, L2 reading ability, PC aptitude, Internet use, learning style, language proficiency, gender, and frequency of access to annotations. Participants then went to a lab and worked with a slideshow. Participants were randomly assigned to the three treatments (23 participants in each). Then, participants took three vocabulary tests and a reading comprehension test. The dependent variables were scores on the four tests. Measures for the dependent variables were a three-part vocabulary test and a combination true/false and multiple-choice test. It was found that combining definitions of words with associated visuals regardless of the type of visual used was more effective in facilitating vocabulary learning than providing

only definitions of words, but there was no significant difference between the groups in reading comprehension ($F_{2, 66} = 2.054$, $p = .136$). The power of the test was only .41, which is "less than the desirable power of .80" (p. 512), and so the author cautions that the sample size was not large enough to detect a significant difference. In this experiment, participants were notified that they would take a comprehension test, but the vocabulary tests were a surprise.

In analyzing the data, the researchers used ANOVA, and for post hoc analysis, Scheffe. The researchers checked to see if any of the participants' background variables created any significant differences between the groups, and found that no moderator was significant at the .05 level.

Liu & Lin (2010). The study by Liu & Lin (2010) addressed whether certain cognitive processes are associated with different types of aids; whether differences in the kinds of dictionaries lead to differences in reading comprehension and incidental vocabulary learning; and what the relationships are between cognitive processes and learning performance. The participants were 80 university students in Tao-Yuan, Taiwan. Treatment conditions were pop-up dictionary, type-in dictionary, paper dictionary, and no dictionary. The four variables related to participants' cognitive processes were the following: 1) degree of willingness to use a dictionary (i.e., consultation frequency); 2) effort exerted to find a word (i.e., search time); 3) effort exerted to read a word in the dictionary (i.e., average time spent with the definition); 4) effort exerted to read text (i.e., reading time overall). The five variables for learning performance were the following: 1) overall vocabulary performance (i.e., score on test); 2) performance on checked items; 3)

accuracy of vocabulary learning (i.e., performance on checked items divided by the total number of checked items); 4) learning efficiency for vocabulary (i.e., accuracy of vocabulary learning divided by the average dictionary using time); and 5) reading comprehension (i.e., score on test). The electronic dictionaries generated usage logs containing information about times spent with gloss, and so on.

Participants completed a background questionnaire, and then read a passage for comprehension. Next, they took a comprehension test and a vocabulary test. The dependent variables were scores on the vocabulary and reading comprehension test. Measures were a surprise vocabulary matching test, a reading comprehension test, and user behavior measurements from the software. Correlation analysis was used to investigate the relationship between cognitive processes and performance. It was found that participants with a pop-up glossary had the shortest overall reading time when compared with the other two conditions, $F(2,57) = 12.06$, $MS_e = 193.95$, $p < .001$. Users of the pop-up dictionary consulted it twice as many times as participants using the two other types of aids, $F(2,57) = 13.46$, $MS_e = 3827$, $p < .001$. The type-in dictionary took five times longer to perform a search. The pop-up dictionary was the most efficient dictionary type, in terms of vocabulary learning. No significant main effect on comprehension was found across the four conditions, $F(3,76) = 1.53$, $MS_e = 646.09$, $p = .21$.

This study placed dependent measures in the context of constructs named according to contemporary research on cognitive load (i.e., as total effort exerted). The study also counted time spent searching, interpreted as effort exerted to find a word.

Al Ghafli (2011). The study by Al Ghafli (2011) addressed the effectiveness of glosses for understanding technical terms. The participants were 222 L1 speakers of Arabic and L2 speakers of English enrolled in petroleum engineering courses. Cohorts came from two main sources: matriculated students at a university as well as others working at a nearby company. Participants had a variety of English proficiency levels. All participants were male. Treatment conditions were types of glosses: 1) audio and text–term was written in L1 and L2, pronounced in L2, and defined in L2; 2) audio, text, and picture (i.e., condition 1 features plus a picture); and 3) audio, text, and animation (i.e., condition 1 features plus an animation). The materials were an online text with glosses. Participants read the text with access to the glosses and then took a comprehension and vocabulary test. They also took a demographic and attitude questionnaire. The dependent variables were reading comprehension and vocabulary learning. It was found that, when controlling for language ability, those who received audio, text, and picture glosses ($M = 9.72$) had higher comprehension scores than the text-based group ($M = 8.35$), $F(2, 218) = 3.07$, $p < .05$. Audio, text, and picture was received more positively ($M = 3.88$) than the text-based group ($M = 3.41$); $F(2, 216) = 3.10$, $p < .05$. There was a positive relationship between participants' language-learning anxiety and reading comprehension ($r = .203$, $p < .05$).

One distinct advantage of this study is the large number of participants, the highest number in this stream of studies. Al Ghafli (2011) did not record total reading times, but instead asked participants to make a note of their start and end times during the orientation phase of the study. This study was not implemented within a normal academic

setting. In fact, the researcher mentions that participants were notified that their performance on the exercise would not be shared with their classroom teacher.

Al Ghafli (2011) wrote the reading passage independently, and then, consulted with subject matter experts at the institution in order to evaluate the relative difficulty of the words (190 in total). Al Ghafli (2011) contracted a graphic artist to make the images and animations for the gloss entries. Additional procedures were employed to ensure that the glossary was suitable for the students. In one, an unrelated cohort of undergraduate students were asked to match the animations with definitions to ensure that it was possible to associate the two. Another procedure was called the Translation Accuracy Survey, in which five native Arabic speakers were asked to rate the accuracy of the Arabic translations intended for inclusion in the gloss.

The reading passage in this experiment contained a total of 902 words. Fifty words in the passage were contained in the glossary, or approximately 7% of the words. Rather than administer a post-experiment interview to get participants' impressions of the text and gloss interfaces, Al Ghafli (2011) performed a design phase analysis in which three people from the cohort each interacted with one of the treatment conditions, and then, provided feedback on the usability of the interface. Each of the measures used to evaluate comprehension and vocabulary retention were modeled on interfaces used by Al Seghayer (2001) and Ben Salem (2007).

Participants were randomly assigned to treatment conditions using a random number generator. After completing the consent form, the researcher described the reading task and demonstrated how to access the glosses. The large number of

participants required that Al Ghafli administer the treatment in twelve separate sessions. Data were analyzed using ANOVA to analyze the effects of media presentation on the three tests. Four univariate analyses of covariance were used to test the hypotheses. The covariates used were language-learning ability, English level, English reading ability, and number of years of speaking English.

Ali Farhan (2011). The study by Ali Farhan (2011) investigated the effect of the location of a gloss (i.e., annotation) and its contents on reading comprehension and vocabulary acquisition. Participants were 78 EFL learners in Jordan. Independent variables were gloss type (i.e., word synonym or full definition) and gloss position (i.e., in the margin, at the bottom of the screen, in a pop-up window, or after a glossed word, no gloss). Dependent variables were reading comprehension (i.e., "text memory," p. 182) and vocabulary acquisition. Prior to the experiment, participants responded to a survey that requested the following data points: age, years using computer, years studying English, average on secondary English exam, GPA, and courses attended.

The reading material and quiz were both analyzed prior to the experiment, and 32 raters from a variety of perspectives (i.e., similar students, faculty, etc.) evaluated the reading material during a pilot study. The words in the story were rated for difficulty in order to determine which words to gloss. A total of 171 words across the seven stories were glossed. The multiple-choice test was evaluated with Chronbach's alpha (.84), and responses on the test had an inter-rater reliability of .95.

Participants came to a large computer lab for one hour per week over seven weeks. In this environment, neither their instructor nor the researcher, but rather lab

assistants, supervised and provided technical assistance. Participants received one of five versions of instructional software that contained the same reading material and glossary. The software was written in Visual Basic. Participants read seven texts and took seven vocabulary and reading assessments. Each test was worth 21 points, and the overall score across seven tests was aggregated. The comprehension test was a written proposition-recall test that asked participants to recall idea units from the story. The vocabulary assessment was a multiple-choice test that contained items asking participants to correctly identify a target vocabulary word.

It was found that participants who had access to a gloss performed significantly better than those who did not, $p > .05$. Participants in the spatially contiguous condition (i.e., gloss after word) performed significantly better on both vocabulary and reading comprehension assessments. Groups with short glosses (3-5 words) performed better than groups with longer glosses (6-7 words), $p > .05$.

One weakness of the study involved selection: with 79 participants split into five groups, each treatment had only between 14–16 individuals. While some conditions reportedly yielded statistically significant results, no measure of power was provided in the report. While Sheffe was reportedly used for post-hoc comparisons, it was not clear from the report whether a repeated-measures design was used for hypothesis testing.

There is only brief discussion of the authoring tools used to create the experimental interface, and no discussion of how the software was distributed to the participants, although it is clear that participants were all at the same site and in the same laboratory. Also, it is not clear if selection may have been stratified by cohort.

Conclusion

This review of the literature discussed reading comprehension, vocabulary, the spatial contiguity principle (Mayer, 2005), and learner control, and analyzed studies published between 1991 and 2011 that explored instructional design questions relating to electronic glossaries and reading comprehension. Over this period, the proliferation of information technologies transformed the field of glossary research. Tools enabled researchers to more precisely measure user behavior, thus reducing measurement error. Evolving authoring environments enabled researchers to create sophisticated experimental interfaces. However, there is no record of an experimental interface for electronic glossaries that can be used by other researchers to replicate investigations using a common methodological platform.

The next chapter will describe the design and implementation of the system used to answer questions pertaining to learner control (i.e., whether English language learners will utilize vocabulary annotations) and the spatial contiguity principle (i.e., whether variations in the presentation of a glossary affect reading comprehension) (Mayer, 2005).

Chapter 3: Methodology

This study implemented an experimental electronic interface for a glossary support mechanism that tested the effect of spatial contiguity (i.e., the relative positioning of a glossary pop-up) on reading comprehension. This chapter describes the methods used to collect and analyze the data. Based on the cognitive theory of multimedia learning and research about learner control and the spatial contiguity principle (Mayer, 2005), the study expected to find an effect for variable spatial contiguity that would be reflected in reading comprehension measures.

Site and Participants

The participants in the study were volunteers recruited from a cohort of 100 students in skill-level four (of four levels) in the English for Academic Purposes (EAP) program at an American university. The university was located in an area outside of the United States where English is not the native language; however, the language of instruction at the university is English, and many of its instructors are native speakers of English. EAP programs focus on academic English, which is to say, the English skills that pertain to the classroom. The partnering university explains that the goal of its EAP program is:

> … to prepare non-native English-speaking high school graduates to enter the undergraduate program by teaching them academic English and critical thinking skills and study habits. Our goal is to insure that upon completion of the Program, students have the necessary proficiency in English reading, speaking, and writing and awareness of academic cultural norms and expectations to succeed in their undergraduate studies. (Mission Statement, 2014, URL withheld for privacy)

The organization of EAP programs is often done by levels of language proficiency on one dimension, and the four skills of listening, speaking, reading, and writing on the other. Hence, EAP programs often offer classes that are focused on one or more of these language domains, and these offerings are varied by the language proficiency level of the students grouped within them. The university that participated in the present study offers four levels of reading and writing classes, with the addition of an academic skills class for levels one and two. The following statement from the school's website summarizes the offerings:

> APP offers four levels of instruction in reading and writing, four levels of instruction in grammar, and two levels of Fundamentals for Academic Success (FAS). Grammar instruction involves drills, exercises, frequent quizzes, and interactive practice to ensure mastery. Grammar instruction is reinforced in writing classes, which involve almost daily in-class writing practice as well as creation of formal essays, reports, and short writing projects that emphasize, for example, paragraph development or patterns of organization. Reading classes involve periods of sustained silent reading followed by discussion and quizzes

> testing comprehension, as well as reading aloud for practice in speaking and pronunciation. In listening and speaking classes, students engage in conversations in pairs, present oral reports to the class, or lead discussions about assigned topics. Students practice listening skills by watching informative videos (speeches, debates) and attending lectures that mimic the pattern of classes in the Academic Program. (Structure of APP, 2014, URL withheld for privacy)

Participant characteristics are detailed further in Chapter 4.

Ethical Considerations

Participation in the study was anonymous so that the privacy of participants could be assured. Registration on the Web site did not require the entry of personally identifying information such as age, name, address, etc. A code number for each student was provided by participants' instructors, who maintained a secure key that associated names with code numbers. This key was not shared with the researcher. Information about participants such as their ages, genders, and English proficiency levels, was provided by their instructor without divulging personally identifiable information. Participants and their instructors read and signed a brief informed consent agreement that detailed the nature of the study, how the data would be used, and contact information for inquiries about the study. The University of Virginia (UVA) Institutional Review Board for the Social and Behavioral Sciences (IRB-SBS) reviewed the proposed protocol for the study and exempted it as normal educational practice on 1/7/13. It was assigned project no. 2012-0443-00. Modification requests were approved on 2/7/13 and 1/22/14. Permission to utilize the copyrighted reading passage was requested from Cengage

Publishing on 1/16/14 and approved on 5/1/14. (Facsimiles of the copyright materials and those related to the protection of human subjects who participated in this study are included in the appendix.)

Materials

A custom software solution, taking advantage of existing code libraries where possible, was designed and built to provide the following features:

- User management system that allows login/logout and maintenance of individual user records;
- Randomization function that assigns participants to treatment groups and provides treatment accordingly;
- Reading interface by which hyperlinked words are associated with pop-up glossary entries (i.e., tooltip.js);
- Mechanism for recording the details of experiments, namely control over independent variables, accurate time-stamping, and reporting;
- Content management system allowing teacher and researcher to add multiple stories and glossaries;
- "Sites and rosters" mechanism allowing teachers to assign activities to participants;
- Lesson management function that aggregates individual stories, glosses, and quizzes into a sequenced instructional activity;
- Testing engine that allows for authoring and presentation of quizzes with automatic randomization of answer options;

- Security features to prevent unauthorized account access, database access, or tampering with script functions by URL manipulation;
- Import/export function allowing content to be transferred between users (i.e., to reduce the amount of content development required to use the site);
- Durability that would allow for system load of 50-60 concurrent users;
- Extensibility allowing for additional features to be added or removed by other developers;
- Interface that looks professional, coherent, modern, and easy to navigate; and,
- Conformance to W3C accessibility standards.

Several options exist for creating this kind of software, and the considerations involved in the related design choices are detailed in the following sections.

Database-driven Web site. Dynamic Web sites, or database-driven Web sites, are so named because their content must be systematically stored and recalled whenever pages are loaded or when user input on a page is saved. One popular operationalization of database-driven program design combines Personal Home Page (PHP) scripts with the database framework MySQL.

PHP/MySQL. The interface was created using PHP/MySQL as a part of a commercially hosted Linux, Apache, MySQL, PHP (LAMP) "stack." PHP is a free and open-source server plugin that is supported by most Web hosting providers. Commercial alternatives, such as Microsoft's ASP, require payment for software licensing, whereas PHP, as well as the other components of the LAMP stack, carry no software licensing fees.

Web site vs. native app. Two leading options for content delivery are available to Web developers: 1) a Web page (i.e., a standard Web site that is viewed with a browser), or 2) a native application that wraps code into a form that can be loaded directly onto a device (i.e., in the way that certain computer applications are installed vs. accessed as an online service). Native applications allow for the leveraging of hardware features peculiar to mobile devices; however, the software used in the present study is the first kind—a Web application, or a dynamic Web site. Several toolkits have emerged that give developers the option to convert a set of Web pages into a native application, and consequently many developers develop their code for the Web rather than for a native application. The following sections describe the features of the software.

Registration and form validation. Users create their own accounts on this system using a site code that is given to them by their instructor. Figure 4 shows the registration page.

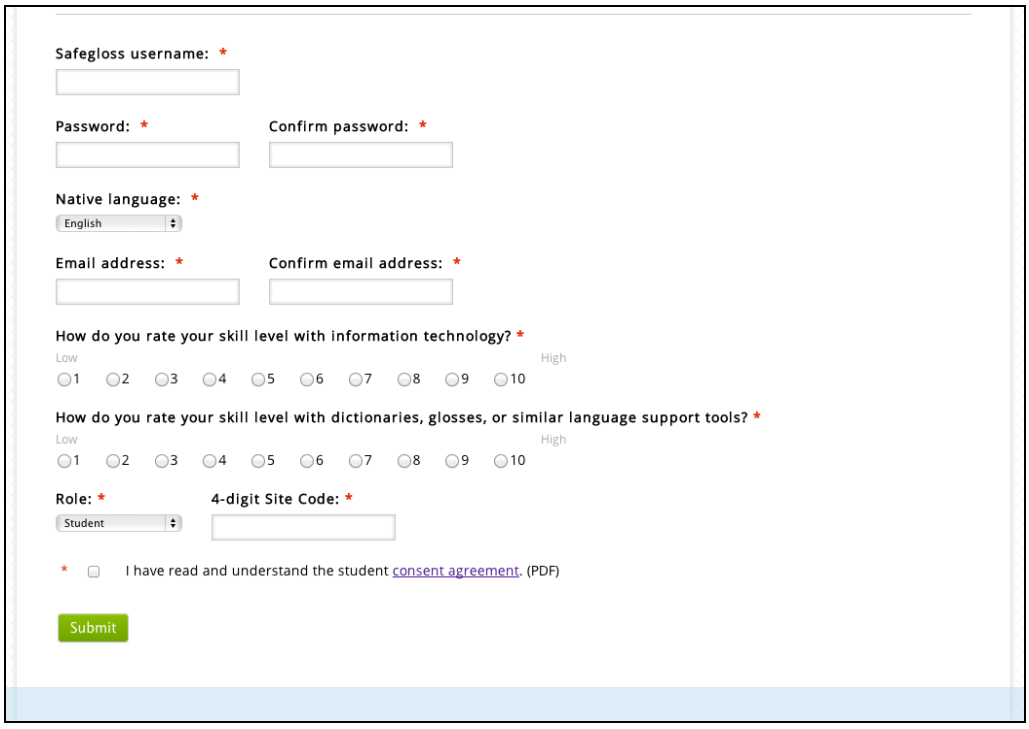


Figure 4. Self-registration page with integrated survey items.

A menu option allows registrants to choose to create either a teacher or a student account, and the informed consent mechanism automatically adjusts to provide the appropriate version of the informed consent paperwork depending on the selection (i.e., there are separate forms for each participant role). Form-validation code included on the page notifies users of any errors in the form when it is submitted. This ensures that all form elements are completed by registrants, and allows participants to make necessary adjustments and submit the form again when errors are detected (see Figure 5).

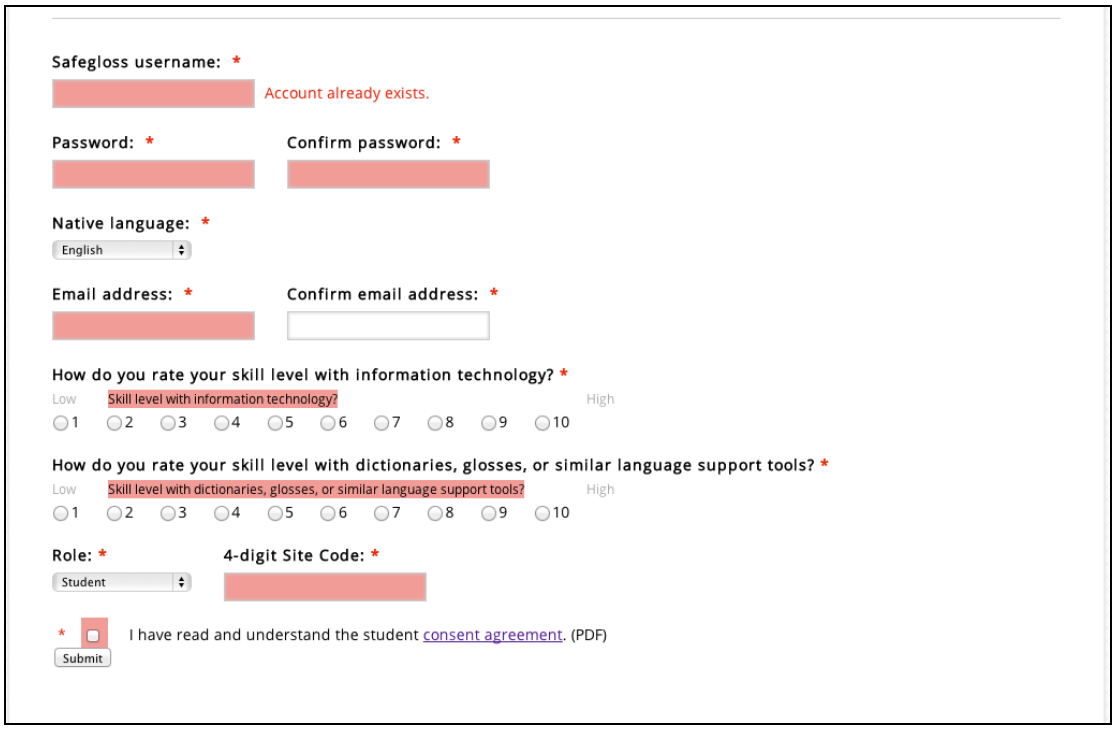


Figure 5. Form validation features.

The form validation functions are important because two data points (the Likert-scale responses) are collected by this form. Figure 5 shows the areas of the form that respond to input errors (in red).

Treatment assignment. The registration page was scripted with PHP to assign each participant to one of the treatment groups in an alternating fashion so that, within a given site (i.e., class), the number of participants in each of the groups would be equal. Participants were selected into a group based upon the assignment of the previous registrant, with the first registrant to the site being assigned to Group A or B at random, the next registrant to the alternate group, and so on. The probability of assignment to one treatment group or the other was equal, although the decision about the individual

assignment of each participant to a group was not made at random, per se. In light of all the intervening variables that could bias one group or the other in terms of the outcomes, the randomization routine serves to ensure that any of these potentially confounding variables are equally likely for either group. That is to say, randomization does not correct for these unknowns, as such, but it is a reasonable countermeasure to prevent them from potentially confounding the validity of between-group comparisons. English-education environments are typified by a very broad array of participant background characteristics, probably more than what would be found in a "typical" academic cohort. Additional considerations regarding the randomization feature will be discussed in Chapter 5.

Participants assigned to Group A see the pop-up window appear in the right margin when clicked, and those assigned to Group B see the window appear next to the hyperlinked word. The variation in treatment was achieved by a line of PHP code that dynamically chose the gloss style sheet according to the variable in the participant's user record in the database. Other than this variation, the reading passage was identically styled, with the same content in the passage and the individual glossary entries. The responses on the form, as well as the assigned treatment group, were saved into a user details table in the database by the PHP script upon submission of the registration form.

Dashboards. Independent views of the Web site were created for three user types: student, instructor, and researcher. The PHP script that creates the home page accordingly contains parameters that alternatively provide the appropriate tabs for each user type, allowing a single home page script to dynamically serve each type of user. As

shown in Figure 6, the instructor sees a “dashboard” upon logging in, which displays the total number of lessons, rosters, students, glossaries, and quizzes associated with his or her account. From this screen, the instructor can navigate to the administration views for stories, glosses, quizzes, lessons, site, and gradebook.

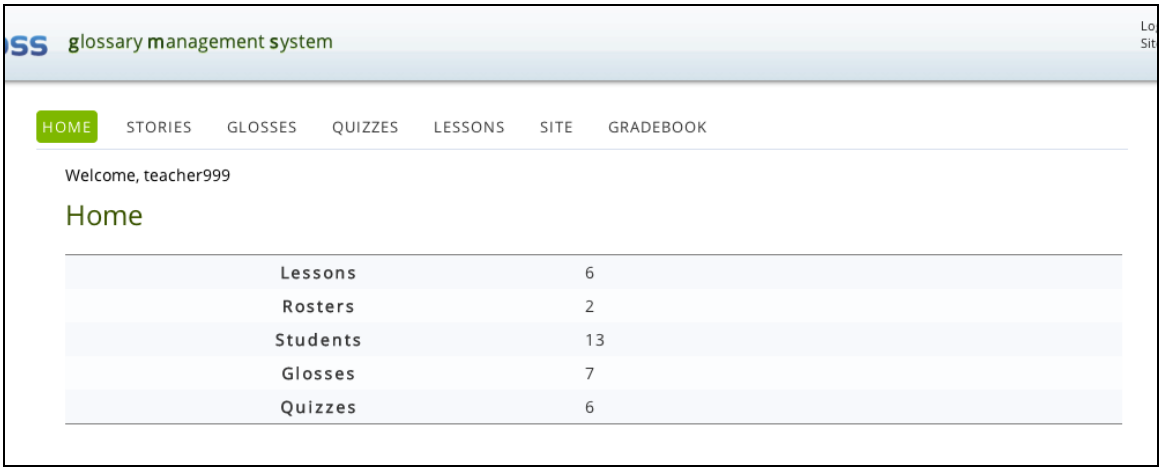


Figure 6. Index.php, instructor view.

Using the same PHP file, the system displays a different view to student users. As shown in Figure 7, student users see fewer options. The dashboard displayed to users in this role indicates the total number of lessons, glossaries, and quizzes that have been assigned to them. For student users, navigation options are limited to My Lessons and My Scores.

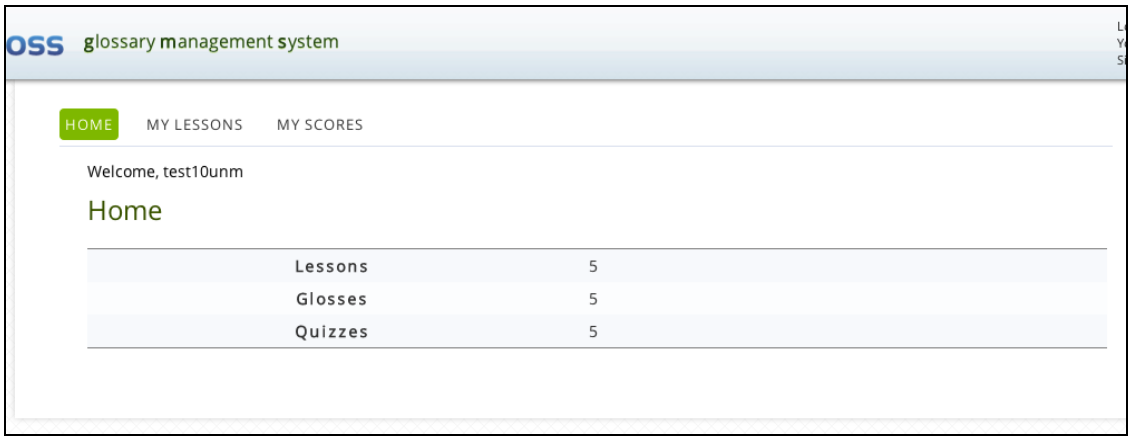


Figure 7. Index.php, student view.

As shown in Figure 8, researchers can access the total number of lessons, rosters, students, glosses, quizzes, and sites that are present in the database. In addition to this information, the researcher can navigate to System Log and Score Log tabs that contain detailed information about system usage.

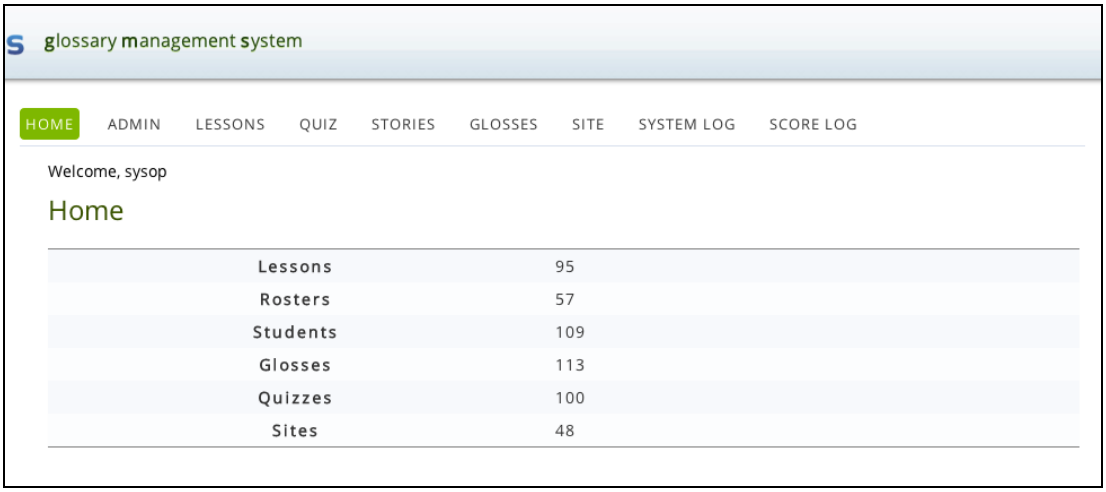


Figure 8. Index.php, researcher view.

Content and Gloss Management

The following sections detail the features of the software that support the generation of texts with integrated glosses (which can be aggregated with other material to form lessons).

Roster manager. When participants register, they are added to a list of participants who have input the same site code. This is called the site roster. When a lesson is created, the instructor has the option to assign the lesson to either the site roster or to a custom roster that may contain only a subset of the participants associated with that site (see Figure 9). This feature allows different lessons to be assigned to subgroups of participants within the same cohort.

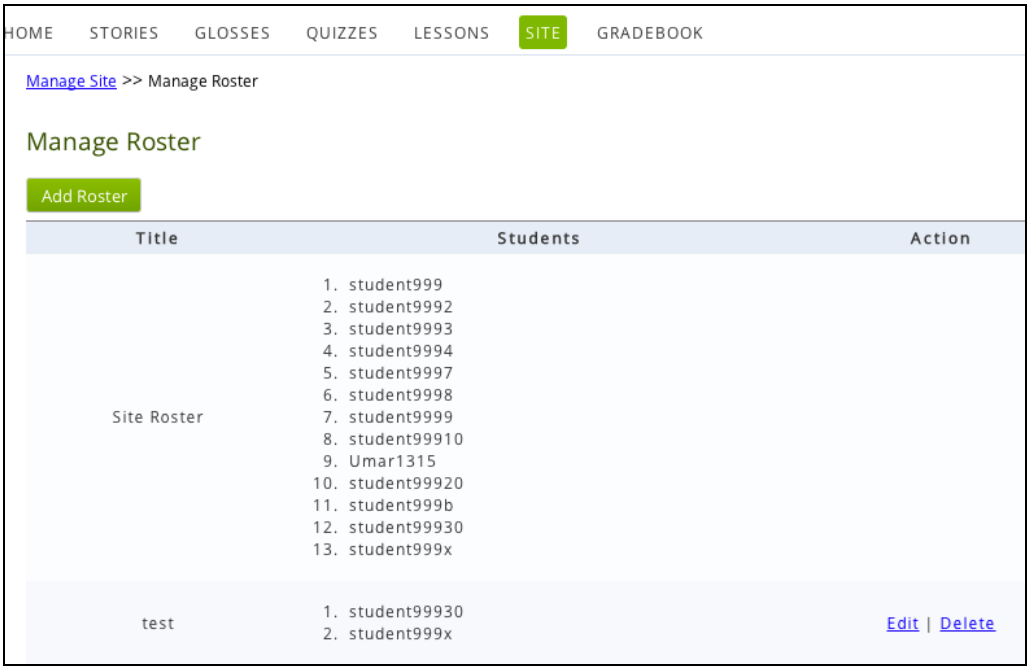


Figure 9. Roster manager.

The roster manager is connected to the lesson manager so that, upon creating a new lesson, an instructor can assign that lesson to any roster in the roster manager. This feature is discussed in more detail in the Lesson Manager section.

Story manager. The story manager allows the instructor to paste or type the text of the story (i.e., reading passage) into a Web page. As shown in Figure 10, the story manager dashboard displays a list of stories, their first lines, the glosses associated with each story, and links to edit or delete each story.

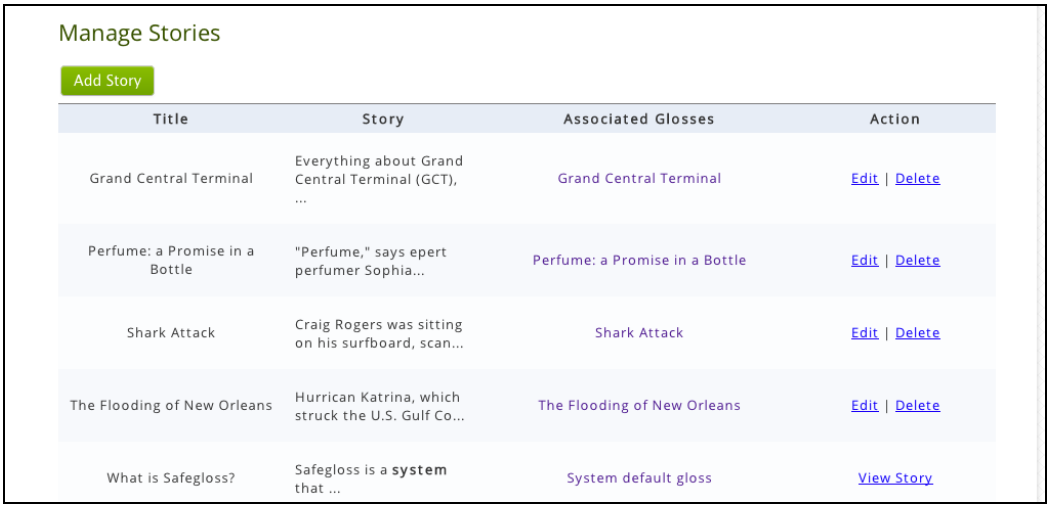


Figure 10. Story manager.

The text input area provided in the story manager, as with the gloss and quiz managers, is provided by CKEditor, a popular, open-source solution (as seen in Figure 11). The interface for the CKEditor installation in the story manager was customized to remove all unnecessary menu options, and to add one button supporting Adobe Flash animations and another supporting the addition of special (e.g., accented) characters.

When a word or phrase in the editor is bolded, it is added to the glossary for that story automatically.

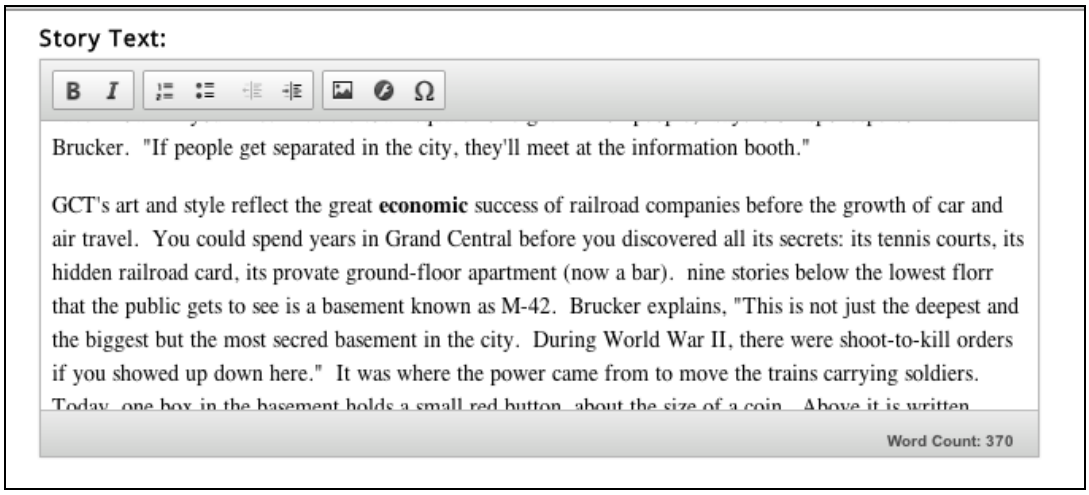


Figure 11. Story editor input area with word count.

When viewed by a student, any word that was bolded in the story manager will receive a blue font color and a tooltip pop-up. The CKEditor window in the story manager provides a word count beneath the text input area.

Gloss manager. Instructors can navigate to the gloss manager in order to add contents to each individual gloss. The text input box is the same CKEditor tool as used in the main story; hence, it supports multimedia annotations. The contents of each pop-up gloss are entered in this area. Glossaries in XML format can be imported directly into the gloss manager. The XML schema for importing and exporting glosses is the same as the one implemented into the glossary features in Adobe Captivate, a popular commercial e-learning authoring system, whose recent version (6.0) introduced an integrated glossary

"widget" that includes an XML import/export feature. Using this XML schema allows glossaries created in the experimental system to be used in Adobe Captivate glossaries, and vice versa. As seen in Figure 12, the word manager within each gloss displays the glossed word, the first words of the definition, and links allowing the user to edit, delete, and add equivalents (i.e., translations).

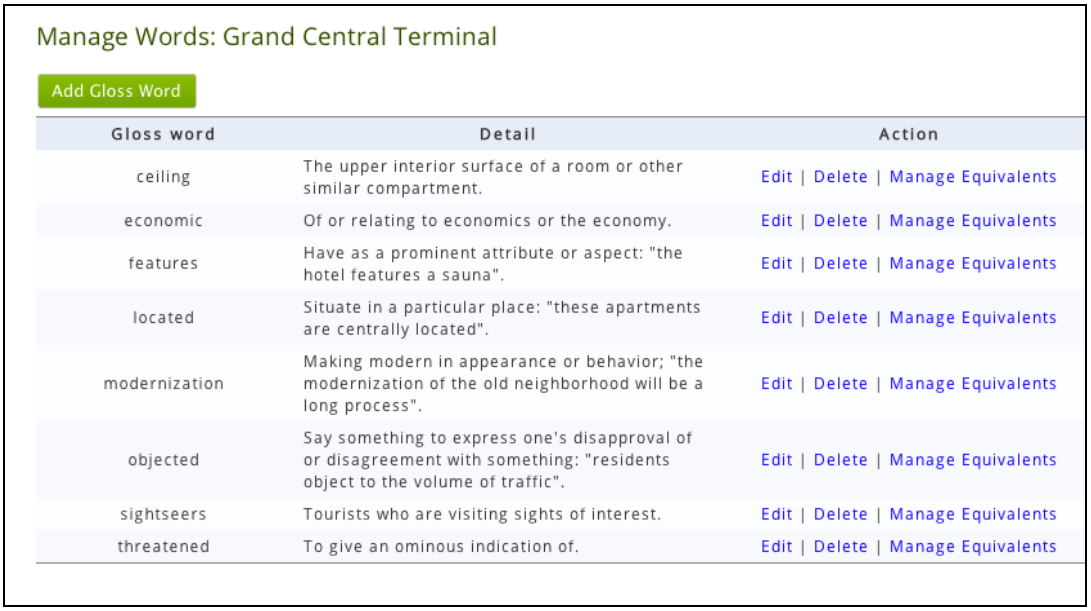


Figure 12. Word manager with add, edit, and delete options.

Quiz manager. A basic quiz engine is provided to make it possible for instructors to include multiple-choice quizzes as a part of their lessons. This area of the software allows the instructor to add question prompts (that can include images and other multimedia) and a list of possible answers, with a simple radio button used to indicate the correct answer. When quizzes are presented to students, the answer options are presented

in random order. As shown in Figure 13, the quiz manager displays all quizzes in an instructor's account, and offers links to manage the questions within each quiz, to edit quiz properties, and to delete quizzes.

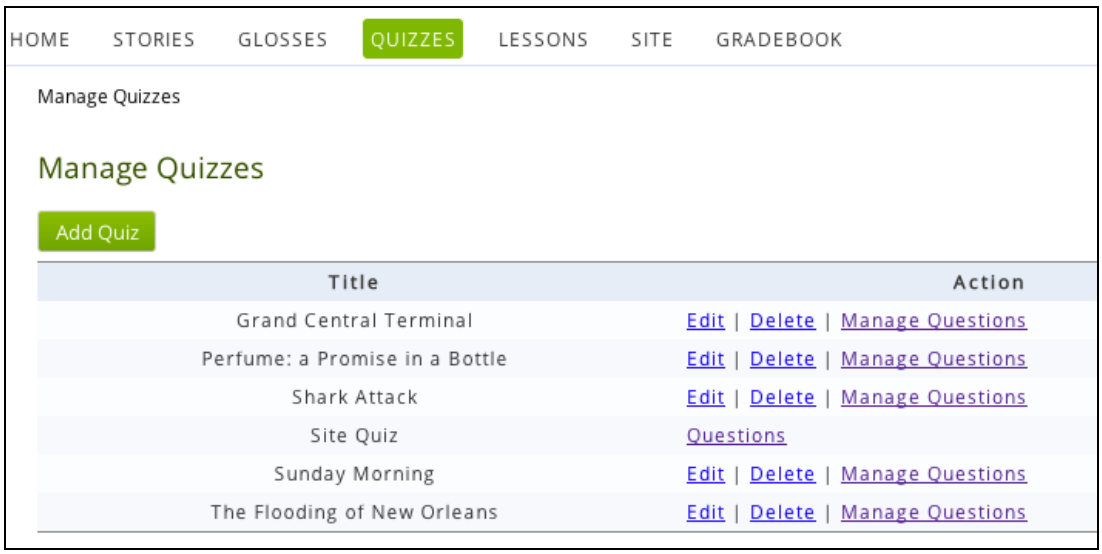


Figure 13. Quiz manager with add, edit, and delete options.

Lesson manager. The lesson manager view offers a list of lessons associated with an instructor's account (see Figure 14). Two large buttons link to the Add Lesson and Import Lesson functions. Beside each lesson's title are options for managing that lesson (i.e., edit, delete, and export). The Export option downloads an XML-separated version of the lesson to the user's local machine to allow for sharing with other instructors.

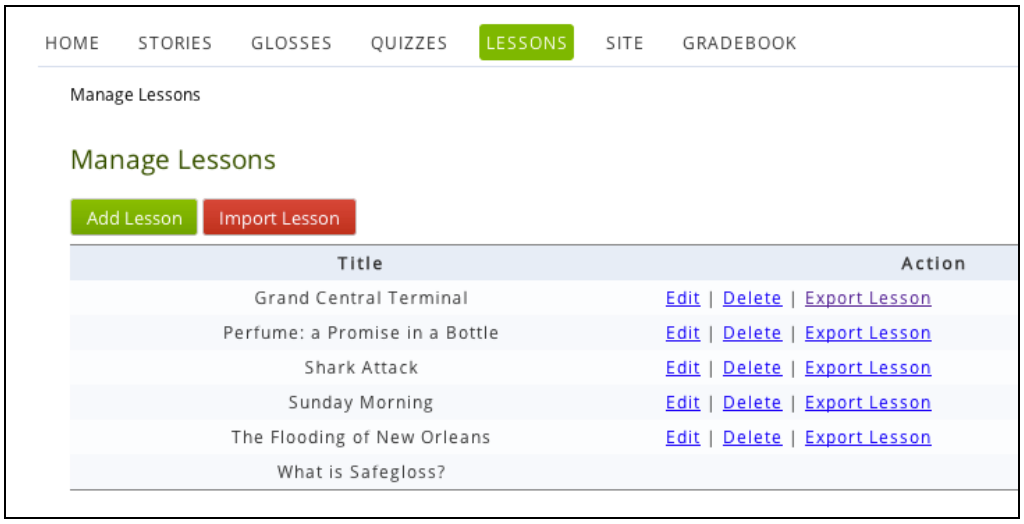


Figure 14. Lesson manager with XML import and export functions.

Upon creating, importing, or editing a lesson, the instructor is presented with a form that allows them to designate which story, glossary, and quiz are components of the lesson. These components are divided into tabs on an accordion menu. As shown in Figure 15, the first tab contains fields for the title of the story and the lesson introduction.

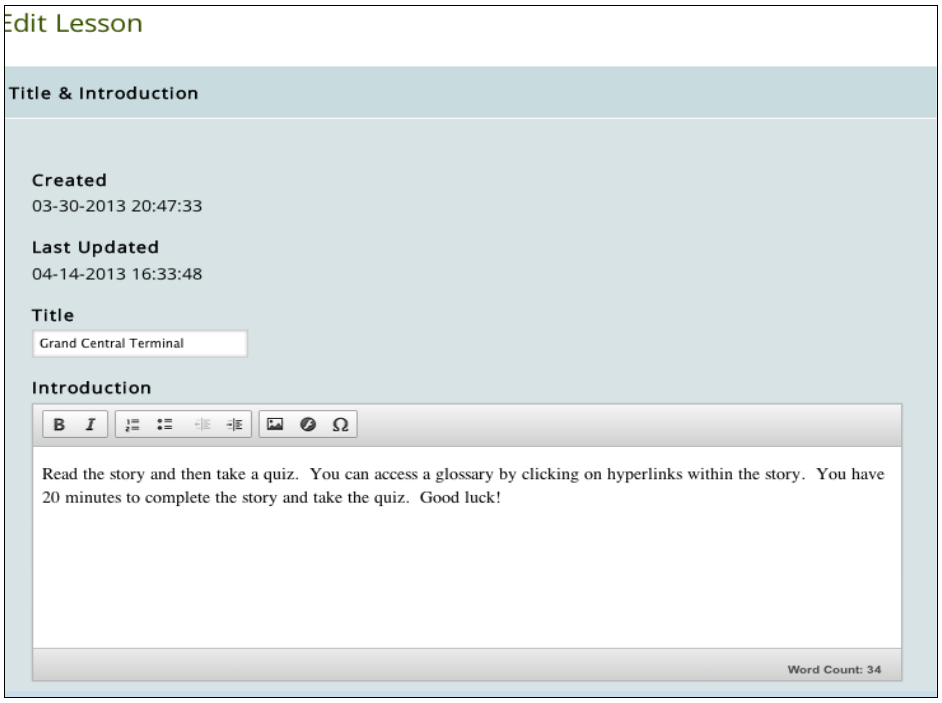


Figure 15. Lesson editor Title & Introduction tab.

The second tab contains selectors for story and quiz. These selectors list each story and quiz that have been added to that instructor's account. The instructor selects the desired one from each list, as shown in Figure 16.

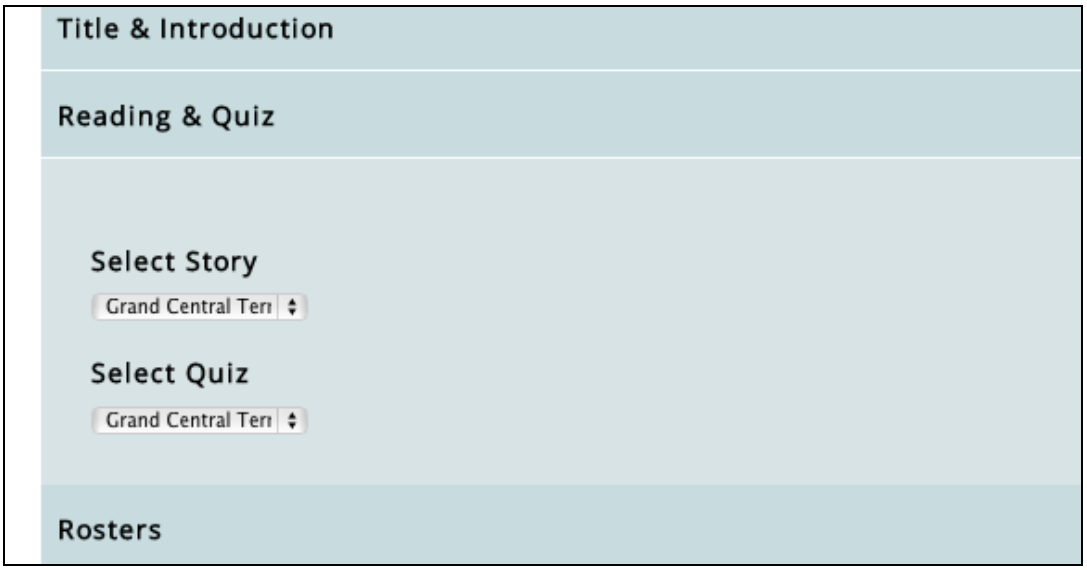


Figure 16. Lesson editor Reading & Quiz tab.

The third tab contains options for assigning the lesson to rosters. The lesson can be saved without assigning to a roster; however, a pop-up message alerts the instructor whenever a saved lesson is not assigned to a roster. Figure 17 shows the area of the lesson manager that the instructor uses to assign the lesson to students

Figure 17. Lesson editor, Rosters tab.

After building the components of a lesson and assigning one to a roster, students can log in and view the lessons.

Lesson sequence. The sequence of displays presented to the student participants may be called a *lesson sequence*. As shown in Figure 18, after logging in, students see a list of lessons assigned to them in their My Lessons tab. Each is a hyperlink that leads to the start of the corresponding lesson.

Figure 18. My Lessons, student view.

Lesson introduction. Upon clicking on a lesson's title, students see a lesson introduction page (as shown in Figure 19). This page displays the title of the lesson, and prompts students to the purpose of the activity, time limits, and so on (as provided by the instructor in the lesson management page). Clicking the Continue button at bottom right loads the main part of the lesson, the reading activity.

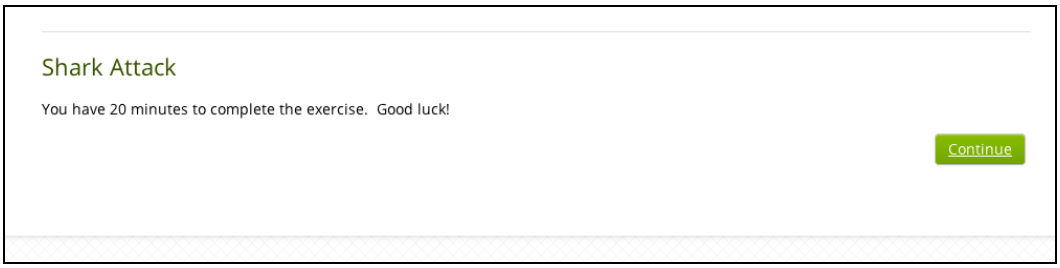


Figure 19. Lesson introduction, student view.

Reading Passage. When the reading is loaded, a timestamp for "reading start" is entered into the database. As shown in Figure 20, the reading portion of the lesson is displayed as plain HTML text with hyperlinks styled in blue, per the convention for displaying hypertext, but without an underline.

Characteristics. The reading used in the present study contained 682 words. The reading passage was from Pathways 3 (Cengage Learning) and was titled *The Healer of Cordoba*. This resource was provided by instructors of the level 4 reading course offered at the institution, from the library of instructional materials used for the course. Post-hoc analysis of the difficulty of the reading passage yielded a Flesch-Kincaid Grade Level of 11.6 and a Flesch Reading Ease score of 45.4. Thus, the reading passage was found to be generally suited to the English language levels of the participants.

Tooltips. The pop-up glossary functionality in the experimental system was created by implementing the popular, free JavaScript code library *tooltip.js*. As mentioned, tooltips are dialogue bubbles that appear when a word or phrase is clicked. Tooltips can be customized with a number of looks and animation behavior. The behavior of tooltips is such that the page does not need to refresh in order to display the dynamic

content when it is called. Tooltips were implemented into the glossary mechanism by installing tooltip code libraries on the server and then calling those libraries from a PHP script in the student's reading page. Depending on the student's treatment group, one of two stylesheets associated with *tooltip.js* is called. One stylesheet designates the contiguous position of the tooltip, where the other designates a position in the right margin. The stylesheets were otherwise identical. *Tooltip.js* allows the tooltips to function reliably on all browsers with JavaScript enabled. In Figure 20, the tooltip is shown in the right margin as it appears when a hyperlink is clicked. A small "x" on the corner of the tooltip allows the user to close the tooltip.

Annotations. The glossed words used in the study were submitted by the participants' instructors, based on pre-experimental guidance that asked instructors to either base the glossary on one of the supports included with the textbook, if possible, or, as an alternative, to select words that, in the instructors' professional judgment, students would click, based on students' limited understanding of those terms. No guidance was given on the number of items that should appear in the glossary. The glossary therefore comprised terms deemed crucial to the understanding of the text and likely to be outside of students' existing vocabulary knowledge. 18 of the words (about 3%) were encoded with hyperlinks that activated a pop-up glossary window. The content of each annotation consisted of a single sentence. Post-hoc analysis of the annotations yielded an average length of 14.7 words, a Flesch-Kincaid Grade Level of 7.8, and a Flesch Reading Ease score of 62.3. Thus, the annotations were generally more comprehensible to the participants than the reading passage, according to this measure.

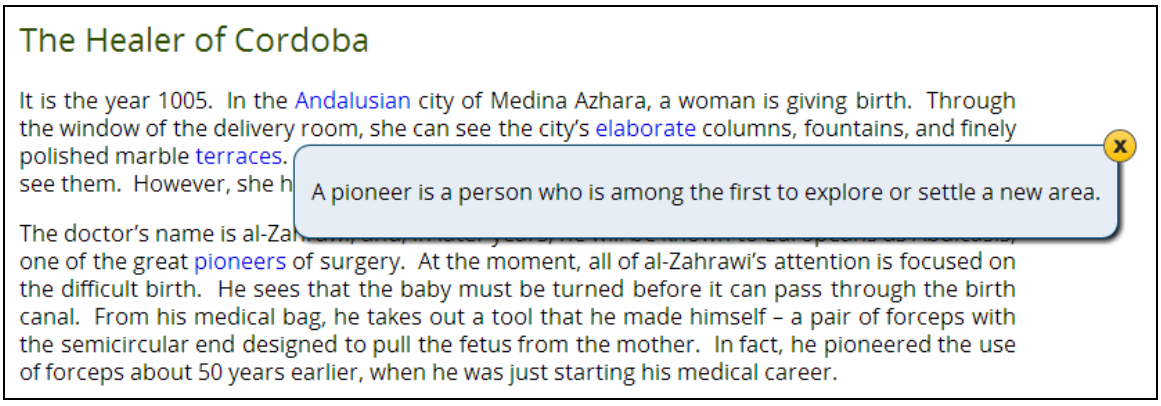


Figure 20. Reading passage with tooltip, student view.

The attribution text provided by the publisher (copyright permission is included in the appendix) was programmed to appear in a space below the story text when viewed by participants, as seen in Figure 21. This displayed field corresponds to an input area in the instructor's story manager (described earlier) .

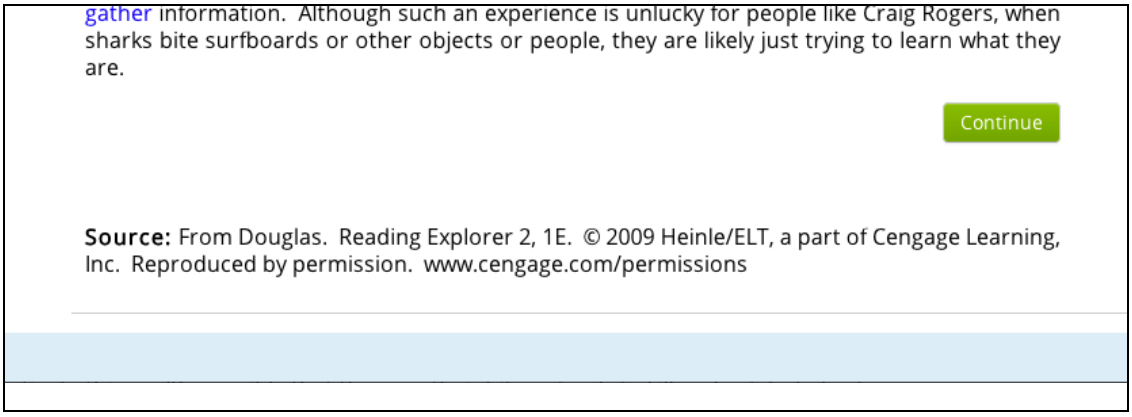


Figure 21. Student view of lesson displaying attribution text at footer.

Comprehension assessment. Upon clicking the Continue button, the student triggers a function to enter another timestamp into the database (i.e., "reading stop"), and then loads the post-reading comprehension multiple-choice quiz, as shown in Figure 22.

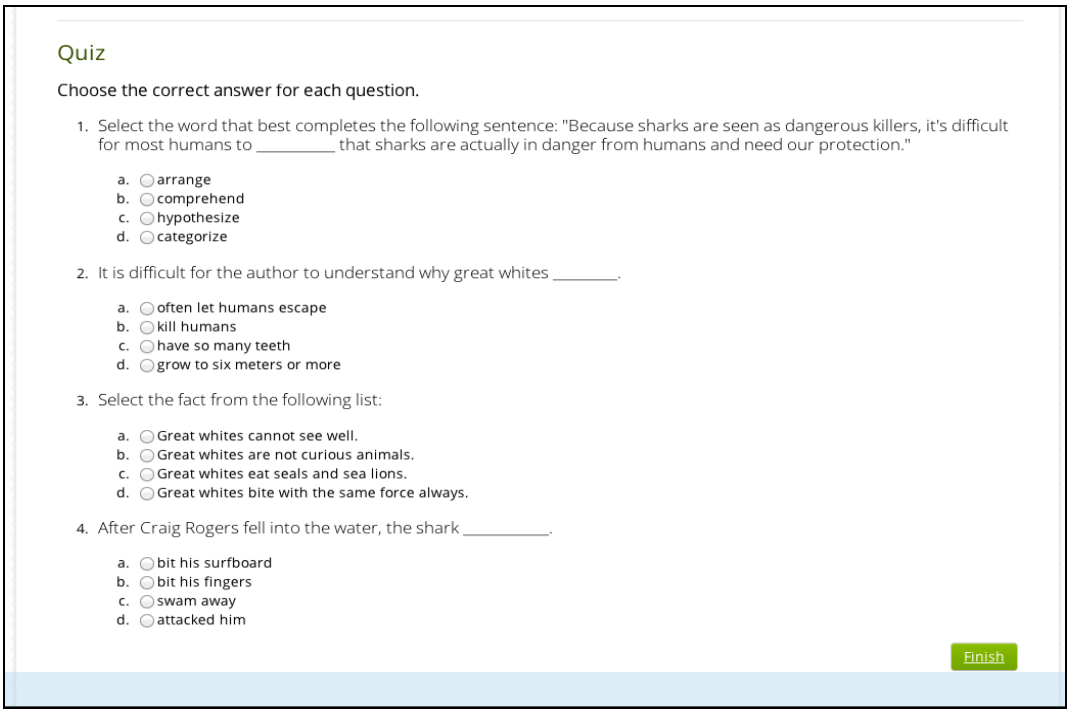


Figure 22. Comprehension assessment, student view.

The assessment was developed by instructors who teach students selected for the study, and was thought to effectively test the participants' overall comprehension of the story. Items on the assessment were limited to those testing comprehension, and any items related to vocabulary acquisition were excluded. The Finish button at lower right activates the quiz scoring function. After displaying the results of the quiz to the student, the system returns the student to the Home screen.

Characteristics. The comprehension assessment was submitted by the participants' instructors, based on pre-experimental guidance that asked them to extract comprehension questions from the textbook, if possible, or as an alternative, to compose test items with responses that would effectively reflect whether the student understood what they read. Because of resource limitations, the quiz engine did not include functionality to evaluate the inter-item consistency or other properties of the comprehension assessment that was used. This is discussed further in Chapter 5. The assessment properties relating to the content validity between the glossed words, their relative importance, and the assessment items were unknown, as also discussed in Chapter 5.

My Scores. When a score is registered and recorded, students can reference the contents of their My Scores page to review their score (for each lesson) and their correct and incorrect responses. The system does not reveal the correct answers. As shown in Figure 23, the My Scores tab displays a record of a completed lesson, listing the individual components of that lesson, the date and time it was completed, and a Detail link.

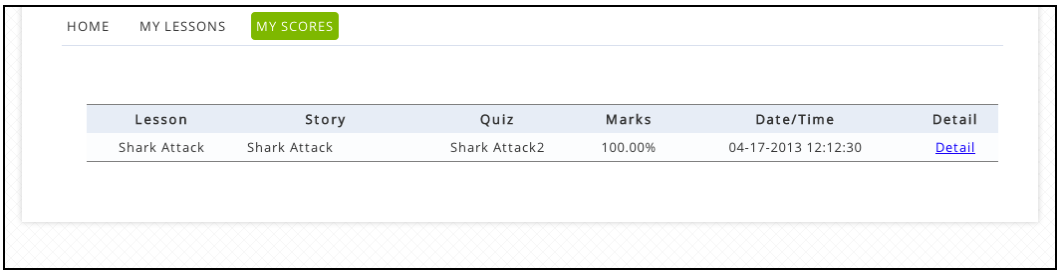


Figure 23. My Scores, student view.

The Detail link activates a lightbox containing a table that shows which items were correct and incorrect (see Figure 24).

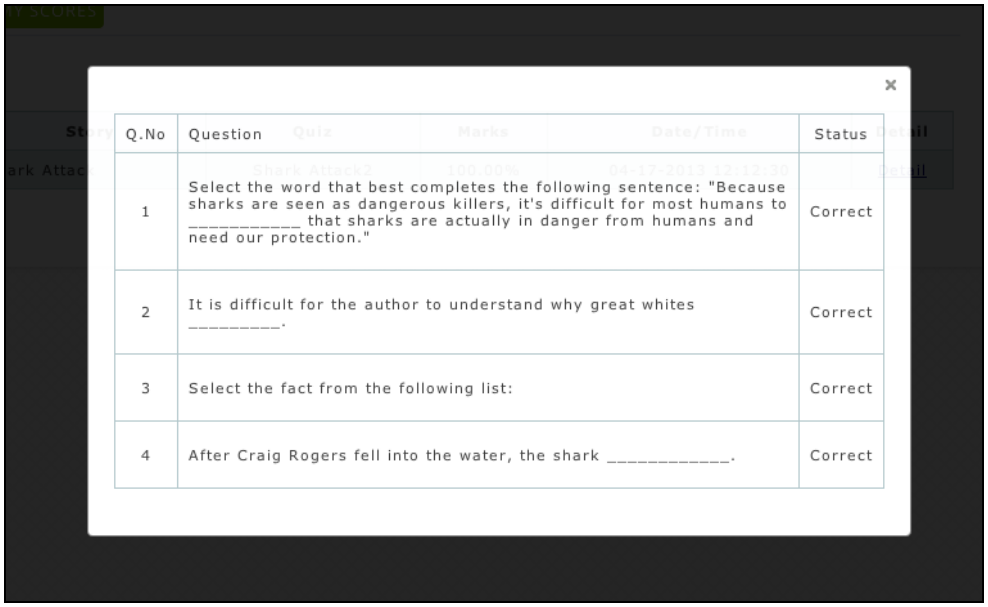


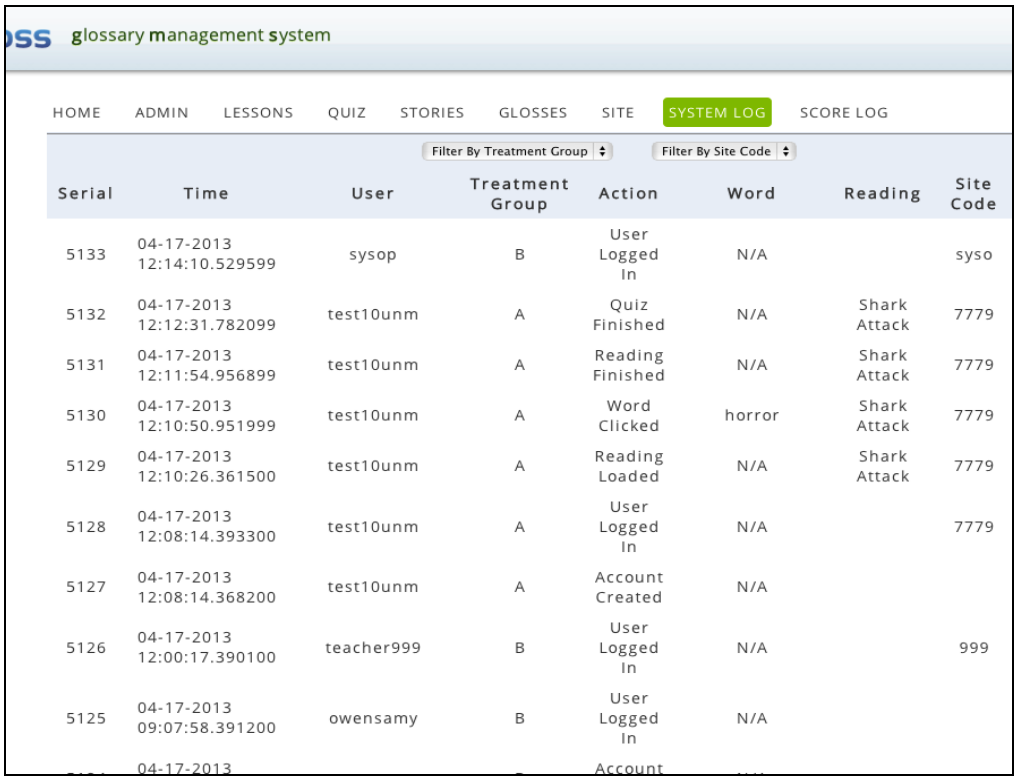
Figure 24. Score detail lightbox, student view.

As mentioned, the data collection mechanism is combined with the materials. The data collection procedures are detailed in the next section.

Data Collection Method

The automated data collection method used in the present study was based, in part, on concepts demonstrated in the dictionary study by Roby (1991), and subsequent replications and adaptations, as well as by Alexander (2009), who used server logs to examine user behavior in PrimaryAccess Storyboard, an experimental software that helps K–12 students plan short video presentations on academic topics (Ferster, 2013).

A randomization function distributed site registrants into one of the two treatment groups. Logging mechanisms coded into the site PHP scripts recorded key actions during user sessions as timestamps, such as start of reading, tooltip open, tooltip close, end of reading, and end of quiz. These timestamps recorded the time at which the action occurred, a description of the action, the name of the user and details about them such as their treatment group ID. The researcher role has the option of auditing the information in the score report by navigating to the System Log tab on the home page. Here, this user role can access a comprehensive log of all system activity related to lessons (i.e., when created, when accessed, glossary usage, and scores). As shown in Figure 25, the System Log contains filters that allow researchers to display log data restricted to one site or treatment group.



oss glossary management system

HOME ADMIN LESSONS QUIZ STORIES GLOSSES SITE SYSTEM LOG SCORE LOG

Filter By Treatment Group | Filter By Site Code

Serial	Time	User	Treatment Group	Action	Word	Reading	Site Code
5133	04-17-2013 12:14:10.529599	sysop	B	User Logged In	N/A		syso
5132	04-17-2013 12:12:31.782099	test10unm	A	Quiz Finished	N/A	Shark Attack	7779
5131	04-17-2013 12:11:54.956899	test10unm	A	Reading Finished	N/A	Shark Attack	7779
5130	04-17-2013 12:10:50.951999	test10unm	A	Word Clicked	horror	Shark Attack	7779
5129	04-17-2013 12:10:26.361500	test10unm	A	Reading Loaded	N/A	Shark Attack	7779
5128	04-17-2013 12:08:14.393300	test10unm	A	User Logged In	N/A		7779
5127	04-17-2013 12:08:14.368200	test10unm	A	Account Created	N/A		
5126	04-17-2013 12:00:17.390100	teacher999	B	User Logged In	N/A		999
5125	04-17-2013 09:07:58.391200	owensamy	B	User Logged In	N/A		
	04-17-2013			Account			

Figure 25. System log showing record of tooltip use.

Data from the activity, including total time spent reading, total gloss clicks, time spent glossing, and score from the reading comprehension assessment were added to the database after the completion of each quiz. The values for time spent reading were generated by calculating the difference of the reading end timestamp and the reading start timestamp, as recorded in a log also stored in the database. To calculate the total time spent glossing, the time durations for which each gloss was displayed (i.e., the time between opening the gloss and closing it) were summed. All time values were converted to seconds prior to analysis.

The system utilized an integrated SMTP mail server to automatically compose and transmit a report to the researcher upon each quiz submission, drawing from the information in the database. Each score report displayed the relevant information for one session (as shown in Figure 27).

From: webmaster@safegloss.org To: researcher@safegloss.org Re: Session Report (24)					
		Participant	24		
		Site Code	----		
		Tech Proficiency	10		
		Gloss Proficiency	10		
		Treatment Group	2		
		Comprehension Score	**33.33%**		
Lesson Title		The Healer of Cordoba		**Lesson ID**	234
Reading					
Reading Start	05:05.5	Reading End	20:35.2	Duration	0:15:29 : 0.65
Gloss Clicks (5)					
Word		**Open**	**Close**		**Duration**
elaborate		05:39.6	05:58.7		0:0:19 : 0.0353
terraces		06:40.9	06:52.3		0:0:11 : 0.4047
contractions		08:23.7	11:10.1		0:2:46 : 0.3402
amputating		11:10.1	14:36.5		0:3:26 : 0.3759
bladder		15:15.8	16:28.2		0:1:12 : 0.4046
				Total Time Spent Glossing	0:7:55 : 0.5607

Figure 26. Script-generated SMTP session report.

Several data sources were used to observe utilization of the experimental interface. These included the following: 1) scores on the reading comprehension test; 2) user input on the gloss and technology scale that measured participants' self-perceptions

of skill in each of those areas; 3) time durations from activity on the reading and integrated glossary, as reflected in logs and reported by integrated messaging features; and 4) glossary click counts (frequencies) from each activity.

Procedure

After watching an orientation session during class that addressed instructors' expectations for the completion of the activity, and receiving a brief, printed pictorial guide on how to use the instructional Web application (see appendix for the guide), participants (N = 20) logged into a custom learning management system through their Web browsers during their free time and completed a registration form that prompted each person to input a username, password, and email address (to facilitate resetting passwords if necessary), and respond to a short survey about their backgrounds with language and technology (as shown in Figure 26), i.e., to specify their native language and their levels of skill in information technology and glossaries (both on 10-point Likert scales).

As shown in Table 2, participants who registered for an account and completed the reading activity and assessment numbered 20, ranging in age from 19 to 27, with an average age of 20, four more males (n = 12) than females (n = 8).

Table 2

Descriptive Statistics of Population Sample per Group

	Group A	Group B
n	10	10
Males	7	5
Females	3	5

Mean Age (years)	20.6	19.8

After submitting the registration form, participants clicked a menu to open an English reading passage. After clicking a button to end the reading, the participants were presented with a multiple-choice comprehension quiz that was composed by their instructors, adapted from material in the textbook.

Chapter 4: Results

This study examined the influence of the spatial contiguity of on-screen glosses on reading comprehension with adult L2 learners of English. Based on relevant literature, the study produced a system for testing this relationship, then implemented it with a cohort of students.

The system was used to conduct an experimental instructional reading comprehension activity with a group of 20 participants. Based on theories of learner control, spatial contiguity and cognitive load, and informed by experimental efforts described in the literature, research question 1 asked if English language learners utilize integrate vocabulary annotations when they are made available during a reading comprehension activity. Research question 2 asked if a variation in the style of a glossary pop-up window affects reading comprehension, time spent glossing, gloss clicks, or reading time. Research question 3 asked if self-reported ratings of technology proficiency or gloss proficiency correlate with reading comprehension, time spent glossing, gloss clicks, or reading time.

Of the 20 participants who submitted a registration form and completed the activity, nine clicked to access at least one gloss while reading, leaving 11 who did not click a gloss. Because the treatment was given only when at least one gloss was clicked, the cases where gloss clicks were less than one were excluded from the next stages of

data analysis. Of the nine qualified participants, six had been assigned to Treatment A, and three to Treatment B, as shown in Table 3.

Table 3

Frequency Statistics of Group Assignment

Treatment	Description	*n*
A	Contiguous gloss	6
B	Non-contiguous gloss	3

For the nine cases, the mean age was 20.8 years (*SD* = 2.36), mean technology proficiency was 6.56 (*SD* = 1.87), and mean gloss proficiency was 7.00 (*SD* = 1.94), as shown in Table 4.

Table 4

Descriptive Statistics of Participant Characteristics (n = 9)

	M	*SD*
Age (years)	20.8	2.37
Technology Proficiency (1-10)	6.5	1.88
Gloss Proficiency (1-10)	7.0	1.94

The group of nine had a mean comprehension score of 53.4% (*SD* = 26.13%), a mean read time of 512.8 seconds (*SD* = 345.03s), and a mean gloss time of 140.33 seconds (*SD* = 138.60s). Each participant clicked an average of five gloss entries (*SD* = 4.1), as shown in Table 5.

Table 5

Descriptive Statistics of Dependent Measures for Sample (n = 9)

	Read Time (seconds)	Gloss Time (seconds)	Gloss Clicks	Quiz Score (percent)
M	512.88	140.33	5.111	53.40
SD	345.033	138.60	4.10	26.13

Participants in Treatment Group A had a mean quiz score of 52.4% (*SD* = 28.8), while those in Group B scored 55.5% (*SD* = 25.4). Group A participants read for a mean time of 377.3 seconds (*SD* = 253.5s), and Group B participants read for a mean time of 784 seconds (*SD* = 387.4s). In terms of time spent glossing, the values were *M* = 91.8s (*SD* = 99.9s) for Group A and *M* = 237.3s (*SD* = 175.1s) for Group B. Group A members clicked glosses an average of 5 times during the experiment (*SD* = 5.1); comparatively, Group B participants clicked an average of 5.3 times (*SD* = 0.58), as shown in Table 6. Seemingly large differences in reading time and gloss time between the two groups were observed – Group B spent nearly twice as long reading and nearly twice as long glossing. The large differences might suggest that the treatment condition may have affected these dependent variables, but the potential for this relationship should not be over-interpreted because of the small sample size.

Table 6

Descriptive Statistics of Dependent Variables, Between Groups

Treatment	Statistic	Read Time (seconds)	Gloss Time (seconds)	Gloss Clicks	Quiz Score (percent)
A	*M*	377.33	91.83	5.00	52.44
	SD	253.55	99.96	5.17	28.82
B	*M*	784.00	237.33	5.33	55.33
	SD	387.41	175.18	.58	25.42

To test whether there was a significant difference between Treatment A and Treatment B in terms of reading comprehension, T-tests were run using the integrated T-Test for Independent Samples function of IBM SPSS Statistics v.22. A test for normality using the Shapiro-Wilk test indicated that the data were statistically normal. The comprehension scores for Treatment Group A showed a normal distribution, $p = .223$, as did the scores in Group B, $p = .651$. Likewise, the p-values for the other measures, as shown in Table 7, were each above the significance threshold of .05, with the exception of gloss clicks, $p = .028$.

Table 7

Shapiro-Wilk Test for Normality

	Treatment	*W*	*df*	*p*
Quiz Score (percent)	A	.869	6	.223
	B	.967	3	.651
Read Time (seconds)	A	.960	6	.823
	B	.895	3	.370
Gloss Time (seconds)	A	.821	6	.090
	B	.829	3	.186
Gloss Clicks	A	.765	6	.028
	B	.750	3	.000

This indicated that this underlying assumption of the T-test was met, with the exception of the gloss click counts, which was a small set of values, containing only three in Group B, for example. As shown in Table 8, heterogeneity of variance was not significant for the dependent variables of quiz score, *Levene's* $F(1,7) = .969$, $p = .358$, reading time, $F(1,7) = .976$, $p = .356$, gloss time, $F(1,7) = 2.34$, $p = .170$, or gloss clicks, $F(1,7) = 1.73$, $p = .229$.

Table 8

Levene Test of Homogeneity of Variance

	F	df_1	df_2	*p*
Quiz Score (percent)	.969	1	7	.358
Read Time (seconds)	.976	1	7	.356
Gloss Time (seconds)	2.345	1	7	.170
Gloss Clicks	1.733	1	7	.229

With little evidence in the literature about the nature of the relationship between glosses, spatial contiguity, and reading comprehension, there was no expectation about the strength, directionality, or other aspects of the relationship that either the gloss presentation or user background characteristics have with reading comprehension, gloss clicks, gloss time, or reading time; therefore, two-tailed tests were used.

An alpha level of .05 was used for all statistical tests. No significant difference was observed between the reading comprehension scores of Group A ($M = 52.45\%$, $SD = 28.82$) and Group B ($M = 53.33\%$, $SD = 25.42\%$); $t(7) = -.146$, $p = .888$. To test whether

significant differences existed among the other dependent variables, additional T-tests were run using the same procedure in SPSS. As shown in Table 9, no significant difference was observed for gloss time by Group A (M = 91.83s, SD = 99.96s) vs. Group B (M = 237.33s, SD = 175.19s); $t(7) = 1.631$, $p = .147$, and no significant difference was observed for reading time by Group A (M = 377.33s, SD = 272.54s) vs. Group B (M = 784.0s, SD = 387.42s); $t(7) = -1.93$, $p = .095$. Likewise, no significant difference was observed for gloss clicks by Group A (M = 5, SD = 5.17) vs. Group B (M = 5.3, SD = .574); $t(7) = -.107$, $p = .917$.

Table 9

Independent Samples T-test for Equality of Means

	T	*df*	*p* (two-tailed)
Quiz Score (percent)	-.146	7	.888
Read Time (seconds)	-1.930	7	.095
Gloss Time (seconds)	-1.631	7	.147
Gloss Clicks	-.107	7	.917

To examine the relationships between the dependent variables of reading time, comprehension, time spent glossing, and gloss clicks, Pearson product-moment correlation coefficients were computed using the integrated Correlation function in SPSS. One assumption of this test, as with the T-tests, is normal and homoscedastic data. To test the normality and homoscedasticity of the data, Shapiro-Wilk and Levene's tests were used. Based on the results of the Shapiro-Wilk test for normality and Levene's F

tests conducted to test assumptions of the T-test (Tables 7 and 8), these assumptions were met for Pearson's product-moment correlations.

Learner Control of Glossary Features

Of the 20 participants who submitted a registration form and completed the activity, nine participants (45%) clicked to access at least one gloss while reading, leaving 11 (55%) who did not click a gloss. This suggests that language learners will, indeed, utilize glossary features when offered during a reading activity that is matched to their English reading proficiency. However, the lack of complementary data makes it difficult to speculate why participants did or did not choose to utilize the gloss functionality.

Relationships between Treatment, Reading Comprehension, and User Behavior

No significant correlations were observed between gloss clicks and gloss time, $r(7) = .399$, $p = .288$, but were observed for read time and gloss time, $r(7) = .874$, $p = .002$. This correlation was expected, as time spent interacting with the glosses increases the time spent between the start and end of the reading portion of the activity. There was also a significantly positive correlation between gloss self-efficacy and technology self-efficacy, $r(7) = .790$, $p = .011$, suggesting that participants rated themselves similarly on both scales, that is, students with a high perceived level of technology self-efficacy also reported a high perceived level of self-efficacy with glosses, as shown in Table 10.

Table 10

Pearson Correlation Matrix for Participant Background Characteristics and Dependent Variables

Measure	Statistic	Technology Proficiency	Gloss Proficiency	Quiz Score	Read Time	Gloss Time	Gloss Clicks
Technology	*r*	--	.			.	
Proficiency	*p*						
Gloss	*r*	.790*	--			.	
Proficiency	*p*	.011					
Quiz Score	*r*	-.259	-.002	--			
(percent)	*p*	.501	.995				
Read Time	*r*	.058	-.033	.029	--		
(seconds)	*p*	.883	.933	.941			
Gloss Time	*r*	.097	.107	.121	.874**	--	
(seconds)	*p*	.804	.784	.756	.002		
Gloss Clicks	*r*	-.139	.141	.443	.235	.399	--
	p	.722	.717	.233	.543	.288	

Note: *significant for α = .05; ** significant for α = .01. All p-values two-tailed.

Relationships between Background Variables, Reading Comprehension, and User Behavior

No significant correlations were observed between gloss proficiency and reading score, $r(7) = -.002$, $p = .995$, two-tailed, between tech proficiency and reading score, $r(7) = -.259$, $p = .501$, nor between tech proficiency and gloss clicks, $r(7) = -.139$, $p = .722$, or gloss proficiency and gloss clicks, $r(7) = -.141$, $p = .717$. Similarly, tech proficiency was not significantly correlated with reading time, $r(7) = .058$, $p = .883$ nor was gloss proficiency significantly correlated with reading time $r(7) = -.033$, $p = .933$. Finally,

gloss proficiency was not significantly correlated with gloss time, $r(7) = .107$, $p = .784$, or reading time, $r(7) = -.033$, $p = .933$, as also shown in Table 10.

Chapter 5: Conclusions and Implications

This study designed, developed, and implemented an experimental interface for electronic glossaries, using it to test if English language learners would utilize integrated vocabulary annotations when they are made available, and to test if different modes of presentation would impact reading comprehension or affect user behavior in terms of gloss clicks, time spent glossing, or time spent reading. No significant differences were found between treatment groups in terms of reading score, gloss clicks, gloss time, or reading time, nor were any correlations found between technology self-efficacy or dictionary self-efficacy and the dependent variables. As in the pilot study described briefly in Chapter 1, this study has continued to promote questions about how and why students utilize vocabulary supports in academic settings. Given the relevance of learner control and spatial contiguity to glossing, further qualitative inquiry might explore how students spatially configure associated materials when they access annotations autonomously using secondary computing devices like smartphones.

The study has made a contribution to the field in several ways. 1) It has provided a longitudinal review of research into dictionary interfaces over a 20-year period, marking the shift from a largely paper-based paradigm to a multimedia-capable, then Web-based one. 2) It has described a novel instrumentation that capitalizes on the

features described in previous dictionary research systems, as well as current affordances in Web development technologies; 3) A platform for future replications of this quasi-experiment was made live as a freely available Web application, and, the source code was donated to the research community using software repository GitHub so that it can be validated or modified (forked). 4) It presents original findings on learner control, specifically the relationship between the spatial contiguity of glosses and reading comprehension.

For the first research question, which asked if English language learners will utilize an integrated vocabulary annotation when it is made available, it was found that students did indeed make use of the gloss functionality. Almost half of the participants who completed the activity activated at least one gloss entry. It is possible that the choice of whether or not to gloss could be reflective of students' extant vocabulary knowledge of the words in the reading; that is, a decision to click a gloss may depend, at least in part, on whether the student needs more information about a term. However, the present study selected participants for whom the reading material and gloss content was developed by practitioners who knew the students well. Thus, there is some reason to believe that the gloss entries and the reading passage were reasonably well matched to students' English language levels, and that each participant was equally likely to utilize glosses at least in terms of their need to do so.

Still, a host of other potentially intervening variables may influence learners' decision to utilize interactive features of software. For example, it is possible that the orientation session was not equally effective for all students, i.e., that factors centered on

the orientation session may themselves have some impact on learners' decisions about utilization of the software features. In a similar sense, it may be asked if there an affective dimension contributing to this decision. For example, it may be that some students avoided glosses because of some perception that it would be inappropriate to utilize them in an instructional setting. Participant expectations is a known threat to validity (Shadish, Cook, & Campbell, 2002), and is discussed further in the Limitations section.

As suggested in critiques of learner control research (e.g., Gerjets & Scheiter, 2007; Ross & Morrison, 1989), the present study not only measured the degree to which learners utilized optional supports, but also tested the relationship of several background characteristics and behavior measures such as number of gloss clicks with reading comprehension outcomes. Although a well-developed randomization mechanism was implemented to mitigate the moderating effects of unknown intervening variables, the absence of an observed effect for user behavior variables could have been attributable partly to methodological problems not limited to the small sample (Gerjets & Scheiter, 2007).

For the second research question, which asked if a variation in the spatial contiguity of a glossary pop-up window affects reading comprehension, no significant differences were found between groups, but the effect of the treatment on reading comprehension was not evaluated with robust inferential statistics because of data quality conditions, as explained in the Limitations section.

For the third research question, which asked if technology proficiency or gloss proficiency correlate with reading comprehension, time spent glossing, gloss clicks, or

reading time, the results of the T-tests indicated no significant differences in reading comprehension, as measured by quiz score, or in user behaviors (reading time, glossing time, and gloss clicks) between groups, as measured by timestamp calculations. As well, correlations did not indicate any significant relationship between any of the variables other than a strong positive relationship between gloss time and read time, as discussed.

Limitations

The main limitation of the analysis was the small sample. Larger samples make it possible to implement more robust test statistics. This study could not analyze the effect of the treatment on the dependent variables because data quality precluded the use of robust inferential tests.

As mentioned in Chapter 3, the data revealed limitations in the group assignment mechanism that was used. An effective randomization method can help ensure that selection bias does not disproportionately affect any one group's dependent measures. A preferable group assignment mechanism would not only reference the previous registrant's treatment group assignment, and alternate, in order to ensure balanced groups, but would also reference the total number of registered participants for a site who had completed an activity and who logged one or more gloss clicks during that session. This could help ensure that randomization is, indeed, involved in the group assignment, but that registrants to one site on the system will be balanced into similarly sized groups; and further, ensure that the groups also contain equal numbers of participants who clicked on a gloss and those who did not. This adjustment was made in a post-experimental "bug fix" to the PHP code on the registration page that handles the group assignment function.

Although the participants in the study were a part of a cohort, meaning that their English language proficiency profiles were likely similar, the study had endeavored to ascertain an objective measure of each participant's English language proficiency level to mitigate the threat of heterogeneity of units (Shadish, Cook, & Campbell, 2002). Using this information, a researcher could correct for differences in the performance of the participants stemming from differential English proficiency, effectively removing error from the attempt to detect the effect of the gloss treatment. However, the only measure available for this variable was the 1-4 level of the course the participants were enrolled in, and all of the participants were grouped into level four.

As mentioned, a range of participant behaviors is often associated with observer effects, including attempts to positively or negatively influence the study's findings (whatever those are perceived to be). These effects can mask treatment effects and thereby confound their measurement. In the orientation materials, the guidelines state that students are encouraged to utilize the glossary feature if they think it will help them comprehend the reading. As such, it is possible that any glossing activity on the part of participants could have been due to the fact that they felt that it was an expectation. Conversely, participants who did not utilize the glossing features may have also been responding to a perceived expectation that they should do so. A post-experimental survey about perceptions about the gloss features was sent to the participating instructor, but because no responses were received on the survey, the study has an unfortunate gap in this area. Future studies may consider embedding the post-experimental survey into the lesson activity within the system as one method to gather data on these important aspects

of learner's control, interest, and motivation as they relate to the glossary features in particular.

Because the researcher was not on site for the implementation of the study, the fidelity of implementation could not be directly verified. The "unreliability of treatment implementation" is a potential threat to statistical conclusion validity (Shadish et al., 2002, p. 74). Aspects of the environment where the study was conducted that may have impacted implementation are unknown, because the events were not observed.

Finally, as mentioned in Chapter 3, no data were available for measures of internal consistency of the reading comprehension assessment; and likewise, the survey items prompting participants to indicate their level of self-efficacy with glossaries and information technology were not validated measures. This lack of validation is a threat to the construct validity of the measurements, and applies to both reading comprehension assessment and the self-efficacy measures. The consequences of utilizing an assessment instrument that does not have inter-item consistency is that variation in the effect of any problematic items could confound measurement of the intended dependent variable of reading comprehension. This is a substantial potential threat to the study's conclusion validity, as its main constructs were measured by the reading comprehension instrument.

Implications for Practice

Some potential implications for practice were realized by conducting this study. Trainers, teachers, administrators, designers, publishers, and technologists have a stake in ensuring a match between technology resources and individual students. As technological capabilities continue to develop, particularly in the mobile computing sectors, the

increased flexibility of technology will correspond to a wider range of potential applications. However, suitable matching of technology features with participant background variables will require more precise measurements of skill and more reliance on empirical findings for decision-making. If this is done effectively, new technologies may hold enormous potential as vocabulary support devices, or as a number of possible assistive technologies.

Practitioners should consider the quality of instructional materials before using them to make major decisions about instructional programming. As mentioned, the psychometric qualities of the reading assessment could have a major impact on the validity of its results, and likewise with the content alignment of readings, glosses, and assessments.

Suggestions for Future Research

As the present study considered the first two decades of experimental glossary interfaces as investigated with language learners, and related methodologies, in the design of its interface and procedures, so can future researchers use advancements made in the present study to bolster their own efforts. Suggestions are offered in the following sections.

Although the small sample size presented challenges in analyzing the data, there is some boon to the ecological validity of education research when a sample size is about 20. This is a number that more or less matches the number of students usually grouped together as a class or section in normal academic settings. With some statistical corrections that relate mainly to reading comprehension level and corresponding

difficulty level of the reading passage, as well as consideration for other relevant differences in participant background variables that could impact the analysis, data from multiple sites, perhaps using a multiple-baseline design, could be aggregated as one strategy to increase the sample size. With a larger sample, adaptations of the present study could yield more valid insights into whether certain variables related to reading and glossing online are correlated. But with an even larger sample, the effect of the positioning of the gloss pop-up, as well as background variables like glossary and technology proficiency, could be evaluated for their impact on performance measures like reading comprehension and behavior measures like time spent reading, glossing, and the number of times a participant clicks a gloss. Moreover, potentially moderating variables such as gloss and technology self-efficacy could be isolated in order to both measure their relationships with dependent variables and to prevent them from confounding measurements of the main effects of treatments. Furthermore, in order to generate experimental results that have an acceptable level of power, an algorithm in the statistics software program G-Power suggests that this experiment would require at least 100 participants, or 50 in each of the treatment groups (Faul, Erdfelder, Lang, & Buchner, 2007); hence, a study with recruiting efforts of a suitable scope, and, e.g., situated within English language programs at colleges and universities, would likely involve 10 or more sites if it were to cope with the challenges that the present study faced with regard to recruiting and the resulting constraints that precluded the inference of causal relationships.

Extensions of the present study could utilize the “manage equivalents” feature that is included in the software. This feature allows glossaries to contain individual translations of glossed words that are dynamically populated according to each participant’s native language, meaning that a singular teacher can provide L1 support to a plurality of native languages in a single classroom where Spanish, French, German, Arabic, Mandarin, and other target languages are studied, including English as a native language.

The interface used in the experiment includes features that will allow other researchers to include sound, images, and video in their stories and glossaries, enabling them to address research questions about the inclusion of multimedia in glosses. This line of inquiry related to the media elements of glossaries has been popular since the 1990s, as explored in Chapter 2, but the community has not had a consistent set of tools with which to replicate one another’s studies.

Future research efforts could test the effect of variations to aspects of the user interface such as the header, color scheme, line height, font size and style, and so on. Of particular interest should be the styling of the hyperlink of the glossed word. In the present study, the glossed words appeared as normal blue hyperlinks. This is a standard style for navigational hyperlinks, but other styling may be necessary to express that the link is for an annotation, not for navigating to another page. For example, glossed words viewed in the Amazon Kindle e-reader interface are marked with dotted, not solid, underlining.

In a wider implementation, similar systems should employ advanced functionality to measure the relative importance of each word in a reading passage to both the reading comprehension assessment and the level of English vocabulary knowledge of each individual participant, as mentioned in Chapter 3. With a formal way to demonstrate the alignment between the proficiency level in general, and the vocabulary level in particular, more exact statistical methods could account for the influence of characteristics of this alignment (or non-alignment). This concern speaks to the general notion of internal validity, i.e., that the chain of causal reasoning breaks down as successive links in that chain become weaker. But in terms of statistical methods, this lack of validation on these internal validity issues introduces error into the calculations of relationships between independent and dependent variables. Nevertheless, for future research, it is recommended that any off-the-shelf materials be validated with some measure of alignment between participant background characteristics (notably, reading comprehension level, which words are offered in the integrated glossary, the difficulty and internal consistency of the comprehension assessment, and other factors). While many language education administrators find measuring these variables and placing them into context to be a difficult task, several dynamic assessment systems are being developed by major electronic textbook publishers, which may make the prospect of gathering more exact data, and consequently, a potentially more precise application of automated formative assessment, an easier one to realize.

References

Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science, 35*(6), 499.

Aleven, V., & Koedinger, K. (2000). Limitations of Student Control: Do Students Know when They Need Help? In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Intelligent Tutoring Systems* (Vol. 1839, pp. 292–303). Springer Berlin Heidelberg. Retrieved from http://dx.doi.org/10.1007/3-540-45108-0_33

Alexander, R. C. (2009). *Fostering student engagement in history through student-created digital media: A qualitative and quantitative study of student engagement and learning outcomes in 6th-grade history instruction.* (Unpublished doctoral dissertation, University of Virginia).

Al Ghafli, M. H. (2011). *The effect of mediated glosses on vocabulary retention and reading comprehension with English language learners in Saudi Arabia.* (Unpublished doctoral dissertation, University of Kansas).

Ali Farhan, A. (2011). Hypermedia annotation presentation: The effect of location and type on the EFL learners' achievement in reading comprehension and vocabulary acquisition. *Computers & Education, 57*(1), 1281-1291.

Al-Seghayer, K. (2001). *The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study.* (Unpublished doctoral dissertation, University of Pittsburgh).

Ariew, R., & Ercetin, G. (2004). Exploring the potential of hypermedia annotations for second language reading. *Computer Assisted Language Learning, 17*(2), 237-259.

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice (Blackwell Publishing Limited), 20*(1), 50-57.

Aust, R., Kelley, M., & Roby, W. B. (1993). The use of hyper-reference and conventional dictionaries. *Educational Technology Research and Development, 41*(4), 63-74.

Ben Salem, E. (2007). *The influence of electronic glosses on word retention and reading comprehension with Spanish language learners.* (Unpublished doctoral dissertation, The University of Kansas).

Bird, S., & Liberman, M. (2001). A formal framework for linguistic annotation. *Speech Communication, 33*(1), 23-60.

Brown, F. (2010). Vocabulary knowledge and comprehension in second language text processing: A reciprocal relationship? *Asian EFL Journal, 12*(1), 88-133.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*(2), 233-246.

Chen, Y. (2011). Dictionary use and vocabulary learning in the context of reading. *International Journal of Lexicography.*

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80*(2), 183-198.

Cummins, J., Brown, K., & Sayers, D. (2007). *Literacy, technology, and diversity: Teaching for success in changing times*. Boston: Pearson/Allyn and Bacon.

Downey, B., & Eppard, J. (2010, March). The autonomous integration of technology by non-native speakers of English in a university context. *Annual Conference of the Mid Atlantic Language Learning and Technology Consortium,* Washington, D.C.

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

Ferster, B. (2013). *Interactive visualization: Insight through inquiry*. Cambridge, Mass.: MIT Press.

Fraser, C. (1999). The role of consulting a dictionary in reading and vocabulary learning. *Canadian Journal of Applied Linguistics (CJAL)/Revue Canadienne De Linguistique Appliquée (RCLA), 2*(1-2).

Gerjets, O., Scheiter, K., Opferman, M., Hesse, F. & Eysink, T. (2008). Learning with hypermedia: The influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior*, *2*(25), 360-370.

Golkar, M., & Yamini, M. (2007). Vocabulary, proficiency and reading comprehension. *Reading Matrix: An International Online Journal, 7*(3), 88-112.

Hasler, B., Bernd, K., & Sweller, J. (2007). Learner control, cognitive load, and instructional animation. *Applied Cognitive Psychology*, *21*(6), 713-729.

Hayati, M., & Fattahzadeh, A. (2006). The effect of monolingual and bilingual dictionaries on vocabulary recall and retention of EFL learners. *Reading Matrix: An International Online Journal, 6*(2), 125-134.

Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal, 86*(4), 546-561.

Kalyuga, S., Chandler, P., & Sweller, J. (2011). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 25*(1), S123-S144.

Kinzie, M.B. (1990). Requirements and benefits of effective interactive instruction: Learner control, self-regulation, and continuing motivation. *Educational Technology Research and Development*, *38*(1), 5-21.

Kinzie, M. B., Foss, M. J., & Powers, S. M. (1993). Use of dissection-related courseware by low-ability high school students: A qualitative inquiry. *Educational Technology Research and Development, 41*(3), 86-101.

Kinzie, M. B., Sullivan, H. J., & Berdel, R. L. (1988). Learner control and achievement in science computer-assisted instruction. *Journal of Educational Psychology*, *80*(3), 299.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*(3), 239-264.

Kopriva, R. J., Emick, J. E., Hipolito-Delgado, C., & Cameron, C. A. (2007). Do proper accommodation assignments make a difference? Examining the impact of improved decision making on scores for English language learners. *Educational Measurement: Issues & Practice, 26*(3), 11-20.

Krathwohl, D. R. (1993). *Methods of educational and social science research: An integrated approach*. New York; London: Longman.

Larsen, V. A. (1997). *Exploring user behavior utilizing statistical analysis of WWWhttpd access logs: An enquiry employing net-frog.* (Unpublished doctoral dissertation, University of Virginia).

Lesaux, N. K., Kieffer, M. J., Faller, S. E., & Kelley, J. G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly, 45*(2), 196-228.

Mayer, R. E. (2005). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. *The Cambridge Handbook of Multimedia Learning,* 183-200.

Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction, 13*(2), 125-139.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84*(4), 444-452.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*(1), 107-119.

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

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

Paas, F., & Van Merriënboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122-133.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*(2), 221-243.

Rashidi, N., & Khosravi, N. (2010). Assessing the role of depth and breadth of vocabulary knowledge in reading comprehension of Iranian EFL learners. *Journal of Pan-Pacific Association of Applied Linguistics, 14*(1), 81-108.

Roby, W. B. (1991). *Glosses and dictionaries in paper and computer formats as adjunct aids to the reading of Spanish texts by university students.* (Unpublished doctoral dissertation, University of Kansas).

Roby, W. B. (2005). The Internet, autonomy, and lexicography: A convergence? *Mélanges - Centre De Recherches Et D'Applications Pédagogiques En Langues, 28*, 47-66. Retrieved from http://cat.inist.fr/?aModele=afficheN&cpsidt=18285361

Ross, S. & Morrison, G. In search of a happy medium in instructional technology

research issues concerning external validity, media replications, and learner control. *Educational Technology Research and Development*, *37*(1), 19-33.

Scheiter, K. & Gerjets, P. (2007). Learner Control in Hypermedia Environments. *Educational Psychology Review*, *19*(3), 285-307.

Schnotz, W. & Heib, A. (2009). Semantic scaffolds in hypermedia learning environments. *Computers in Human Behavior, 25*(2), 371-380.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction, 8*(4), 351-362.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General, 119*(2), 176-192.

Taylor, A. (2006). The effects of CALL versus traditional L1 glosses on L2 reading comprehension. *CALICO Journal, 23*(2), 309-318. Retrieved from https://calico.org/html/article_119.pdf

Thomas, W. P., & Collier, V. P. (1997). *School effectiveness for language minority students.* (NCBE Resource Collection Series No. 9). Washington, D.C.: National

Clearinghouse for Bilingual Education, George Washington University, Center for the Study of Language and Education.

Welker, H. A. (2010). In Walker R. K. (Ed.), *Dictionary use: A general survey of empirical studies* (Author's Edition). Brazil: Brasilia.

Wongtrirat, R. (2010). *English language proficiency and academic achievement of international students: A meta-analysis.* (Unpublished doctoral dissertation, Old Dominion University).

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language, Learning & Technology, 10*, 85.

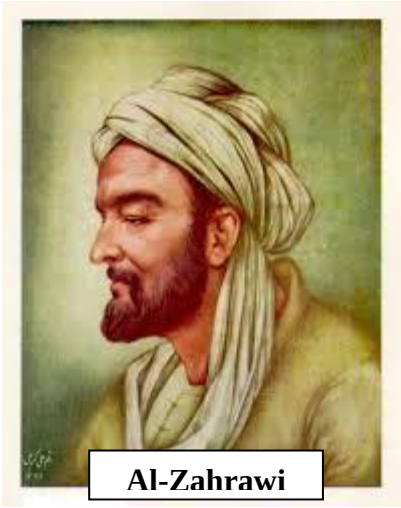
Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of picture and annotation types. *CALICO Journal, 20*(1), 33-58.

Appendix

Reading Passage

The Healer of Cordoba

(from *Pathways 3*)


Al-Zahrawi

It is the year 1005. In the Andalusian[1] city of Medina Azhara, a woman is giving birth. Through the window of the delivery room, she can see the city's elaborate[2] columns, fountains, and finely polished marble terraces[3]. Her heart is pounding because she fears this is the last time she will see them. However, she has great faith in her doctor.

The doctor's name is al-Zahrawi, and, in later years, he will be known to Europeans as Abulcasis, one of the great **pioneers** of surgery. At the moment, all of al-Zahrawi's attention is focused on the difficult birth. He sees that the baby must be turned before it can pass through the birth canal. From his medical bag, he takes out a tool that he made himself – a pair of forceps with the semicircular end designed to pull the fetus from the mother. In fact, he pioneered the use of forceps about 50 years earlier, when he was just starting his medical career.

"Will my baby live?" the desperate mother **manages to** ask between contractions.[4] "Almost certainly," the doctor answers. "You have a healthy

---

[1] Andalusia is a region of southern Spain; during the medieval period of Muslim influence in Spain, the area was known as Al-Andalus.

[2] If something is elaborate, it is richly decorated with a lot of detail.

[3] A terrace is a flat area of stone or grass next to a building.

[4] Contractions are the tightening of the muscles of the uterus during childbirth.

boy. But this next moment is going to be painful." The mother is happy to hear that her baby will live, but, as the doctor warned, the pain is terrible. It is so strong that that she loses consciousness for a few moments, but soon she is awakened by her baby's healthy cry.

The forceps that al-Zahrawi used in the successful delivery are just one of 200 surgical instruments described in his work *Al-Tasrif*, or *The Method of Medicine*. Many of the instruments and techniques described in its pages were invented by al-Zahrawi himself. Born in Cordoba in 936, al-Zahrawi worked as a royal court physician at the height of Muslim civilization in Spain. During his decades-long career, he **compiled** huge amounts of medical knowledge based on existing texts and his own experience.

Al- Zahrawi brought all his knowledge together in the 30 **volumes** of *Al-Tasrif*, a compilation of everything that was known about medicine at the time. The collection begins with **general** concepts, then goes on to describe hundreds of topics including food and nutrition, skin diseases, and poisons. The final, and longest, volume deals with surgery and includes treatments for head and spinal injuries as well as techniques for amputating[5] a limb without killing the patient.

The compilation also includes the world's first illustrations of surgical instruments – sketches of various surgical hooks, knives, scissors, and forceps – many of which look very familiar today. Although surgery was still dangerous and painful, al-Zahrawi's tools would have helped to treat patients suffering from bone diseases, bladder[6] stones, and wounds, as well as **assisting** in childbirth. One of al-Zahrawi's most significant inventions was the systematic use of catgut[7] for stitching[8] a patient internally after surgery. Catgut was found to be the only natural substance **capable** of

---

[5] Amputating a person's arm or leg means cutting all or part of it off in an operation.
[6] Your bladder is the part of your body where urine is stored.
[7] Catgut is a strong cord or thread made from the intestines of animals, usually sheep.
[8] Stitching is using a needle and thread to close a wound or join two pieces of something together.

dissolving[9] and being accepted by the body, and it is still used in surgeries today.

Al-Zahrawi described his instruments and methods in order to share his knowledge with others, including doctors who came after him. However, he may not have been aware of the extent to which his carefully documented knowledge would educate and inform surgeons centuries after his death. Amazingly, given its importance and influence, al-Zahrawi's single, handwritten copy of *Al-Tasrif* was almost lost forever during an attack on Medina Azahara in 1010, when many buildings, and **documents** were destroyed. Fortunately, al-Zahrawi's work was saved. Over the next several decades, it was secretly passed from person to person. Eventually al-Zahrawi's writings were **translated** into Latin from its original Arabic, and, more than four centuries after they were written, parts of the work were finally printed in 1471.

The printed translation **enabled** al-Zahrawi's innovations and observations to spread throughout Europe, where they had an enormous influence on medicine and surgery. The Method of Medicine was used as a **manual** for surgery in medical schools for centuries. Al-Zahrawi's legacy can still be seen in many of the techniques[10] and tools used in modern hospitals, and he continues to be regarded today as the "father of modern surgery."

---

[9] Dissolving is melting away or disappearing.

[10] A person's legacy is something that a person does or creates that will continue to exist after he or she is dead.

Glossary

1. Andalusia is a region of southern Spain; during the medieval period of Muslim influence in Spain, the area was known as Al-Andalus.
2. If something is elaborate, it is richly decorated with a lot of detail.
3. A terrace is a flat area of stone or grass next to a building.
4. Contractions are the tightening of the muscles of the uterus during childbirth.
5. Amputating a person's arm or leg means cutting all or part of it off in an operation.
6. Your bladder is the part of your body where urine is stored.
7. Catgut is a strong cord or thread made from the intestines of animals, usually sheep.
8. Stitching is using a needle and thread to close a wound or join two pieces of something
9. Dissolving is melting away or disappearing.
10. A person's legacy is something that a person does or creates that will continue to exist after he or she is dead.

Comprehension Assessment

The Healer of Cordoba – Comprehension Questions

Which of the following did al-Zahrawi NOT do?

a. design many surgical instruments
b. invent painless ways to amputate limbs
c. systematically use catgut in surgeries
d. record all his medical knowledge in 30 volumes.

Which of the following is the main idea for the second paragraph?

a. A doctor named al-Zahrawi helps a woman through a difficult birth using forceps he designed.
b. Al-Zahrawi described medical techniques and instruments such as forceps in *The Method of Medicine*.
c. A doctor named al-Zahrawi pioneered the use of forceps when he began his medical career.
d. Al-Zahrawi, also known as Abulcasis, was one of the great pioneers of surgery and developed many medical techniques doctors use today.

Please place the following in the correct order from first (1) to last (4).

	Medina Azahara is attacked and *Al-Tasrif* is almost destroyed.
	Al-Tasrif is translated into Latin from Arabic
	Muslim civilization begins in Spain.
	Al-Tasrif is published.

1) If the baby was rotated in the wrong direction al- Zahrawi created a tool to help the baby move in the right direction. Which of the following was the tool that al-Zahrawi created?
   a) Surgical hook
   b) Scissors
   c) Forceps
   d) Catgut

2) True or False: al-Zahrawi's work was written in Arabic and Latin.
   T
   F

3) *The Method of Medicine* was written by:
   a) Abulcasis Medina
   b) Cordoba
   c) Al-Zahrawi
   d) All the above

According to the information found in the reading, which of these patients would al-Zahrawi's medical books NOT help?

a. a patient who needed his leg cut off because of infection
b. a patient who was stabbed with a knife
c. a patient who was deaf
d. a patient who had high blood pressure caused by an unhealthy diet

What language was the printed version of al-Zahrawi's work in?

a. English
b. Arabic
c. Latin
d. Spanish

Which of the following sentences best illustrates the main idea of the text?

a. With his invention of the forceps, al-Zahrawi saved many women's lives during childbirth.
b. Al-Zahrawi is still regarded today as the "father of modern surgery."
c. Al-Zahrawi's *The Method of Medicine* is considered one of the most important and influential medical manuals in history.
d. Al-Zahrawi invented over 200 surgical instruments, many of which are still used in modern hospitals.

IRB Documentation

Pertaining to SBS Number 2012044300

4 messages

mjm6ny@virginia.edu <mjm6ny@virginia.edu> Mon, Jan 7, 2013 at 2:29 PM
To: bod4b@virginia.edu, rf2c@virginia.edu
Cc: mjm6ny@virginia.edu

In reply, please refer to: Project # 2012-0443-00

January 7, 2013

Brendan Downey and Ruth Ferree
CISE (Curriculum, Instruction & Special Ed)
11 Heron Ct.
Charles Town, WV 25414

Dear Brendan Downey and Ruth Ferree:

Thank you for submitting your project entitled: "The influence of the spatial contiguity of on-screen electronic glosses on reading comprehension with L2 learners of English in the United States of America" for review by the Institutional Review Board for the Social & Behavioral Sciences. The Board reviewed your Protocol on January 7, 2013.

The first action that the Board takes with a new project is to decide whether the project is exempt from a more detailed review by the Board because the project may fall into one of the categories of research described as "exempt" in the Code of Federal Regulations. Since the Board, and not individual researchers, is authorized to classify a project as exempt, we requested that you submit the materials describing your project so that we could make this initial decision.

As a result of this request, we have reviewed your project and classified it as exempt from further review by the Board for a period of four years. This means that you may conduct the study as planned and you are not required to submit requests for continuation until the end of the fourth year.

This project # 2012-0443-00 has been exempted for the period January 7, 2013 to January 6, 2017. If the study continues beyond the approval period, you will need to submit a continuation request to the Board. If you make changes in the study, including a change in faculty advisor, you will need to notify the Board of the changes.

Sincerely,

Tonya R. Moon, Ph.D.
Chair, Institutional Review Board for the Social and Behavioral Sciences

Pertaining to SBS Number 2012044300

mjm6ny@virginia.edu <mjm6ny@virginia.edu> Thu, Feb 7, 2013 at 4:30 PM
To: bod4b@virginia.edu, rf2c@virginia.edu
Cc: mjm6ny@virginia.edu

In reply, please refer to: Project # 2012-0443-00

February 7, 2013

Brendan Downey
Ruth Ferree
CISE (Curriculum, Instruction & Special Ed)
11 Heron Ct.
Charles Town, WV 25414

Dear Brendan Downey and Ruth Ferree:

The Institutional Review Board for the Behavioral Sciences has approved your January 31, 2013 modification request to your exempted research project entitled "The influence of the spatial contiguity of on-screen electronic glosses on reading comprehension with L2 learners of English in the United States of America." You may proceed with this study.

This project # 2012-0443-00 has been exempted for the period February 7, 2013 to January 6, 2017. If the study continues beyond the exemption period, you will need to submit a continuation request to the Review Board. If you make changes in the study, you will need to notify the Board of the changes.

Sincerely,

Tonya R. Moon, Ph.D.
Chair, Institutional Review Board for the Social and Behavioral Sciences

Pertaining to SBS Number 2012044300

mjm6ny@virginia.edu <mjm6ny@virginia.edu> Mon, Jan 7, 2013 at 2:29 PM
To: bod4b@virginia.edu, rf2c@virginia.edu
Cc: mjm6ny@virginia.edu

In reply, please refer to: Project # 2012-0443-00

January 7, 2013

Brendan Downey and Ruth Ferree
CISE (Curriculum, Instruction & Special Ed)
11 Heron Ct.
Charles Town, WV 25414

Dear Brendan Downey and Ruth Ferree:

Thank you for submitting your project entitled: "The influence of the spatial contiguity of on-screen electronic glosses on reading comprehension with L2 learners of English in the United States of America" for review by the Institutional Review Board for the Social & Behavioral Sciences. The Board reviewed your Protocol on January 7, 2013.

The first action that the Board takes with a new project is to decide whether the project is exempt from a more detailed review by the Board because the project may fall into one of the categories of research described as "exempt" in the Code of Federal Regulations. Since the Board, and not individual researchers, is authorized to classify a project as exempt, we requested that you submit the materials describing your project so that we could make this initial decision.

As a result of this request, we have reviewed your project and classified it as exempt from further review by the Board for a period of four years. This means that you may conduct the study as planned and you are not required to submit requests for continuation until the end of the fourth year.

This project # 2012-0443-00 has been exempted for the period January 7, 2013 to January 6, 2017. If the study continues beyond the approval period, you will need to submit a continuation request to the Board. If you make changes in the study, including a change in faculty advisor, you will need to notify the Board of the changes.

Sincerely,

Tonya R. Moon, Ph.D.
Chair, Institutional Review Board for the Social and Behavioral Sciences

Project Title: The influence of the spatial contiguity of on-screen electronic glosses on reading comprehension with L2 learners of English in the United States of America

Informed Consent Agreement -Instructor

Please read this consent agreement carefully before you decide to participate in the study.

Purpose of the research study: The purpose of the study is to investigate the effects of the location of vocabulary support on reading comprehension.

What you will do in the study: You will assign a random identification number to each student in the ESL course to be used when they take reading assessments online. You will need to keep the list and not share it with the researcher. The results of the assessments will be returned to you using those numbers. To record the results for students, you will have to match the id numbers back to the student names. At the end of the semester, you will destroy the matched list.

Time required: The study will require about 4 hours outside of the regular class time.

Risks: There are no anticipated risks in this study. Creating the ID codes will take some extra time.

Benefits: There are no direct benefits to you for participating in this research study. The study may help us understand the relationship between vocabulary support and reading comprehension.

Confidentiality: The information that you give in the study will be handled confidentially. Any reports of the study will use a pseudonym for you and your school.

Voluntary participation: Your participation in the study is completely voluntary.

Right to withdraw from the study: You have the right to withdraw from the study at any time without penalty.

How to withdraw from the study: If you want to withdraw from the study, notify the researcher.

Payment: You will receive no payment for participating in the study.

If you have questions about the study, contact:
Brendan Downey
11 Heron Ct. Charles Town WV 25414
Tel: (540) 905-8113
Email: bod4b@virginia.edu

Faculty Advisor's Name: Ruth Ferree
Instructional Technology, 325 Bavaro Hall
University of Virginia, Charlottesville, VA 22903.
Tel: (434) 924-0853
Email: rf2c@virginia.edu

IRB-SBS Office Use Only		
Protocol #		
Approved	from:	to:
SBS Staff		

Project Title: The influence of the spatial contiguity of on-screen electronic glosses on reading comprehension with L2 learners of English in the United States of America

Informed Consent Agreement-Student

Please read this consent agreement carefully before you decide to participate in the study.

Purpose of the research study: The purpose of the study is to measure the effects of vocabulary support on reading comprehension.

What you will do in the study: You will answer a short questionnaire and then read an English text. While reading, you will have access to an electronic dictionary. After reading, you will answer a series of questions about the reading and the interface you used.

Time required: The study will require about 1 hour of your time, and will take place during your regularly scheduled class meetings.

Risks: There are no anticipated risks in this study.

Benefits: There are no direct benefits to you for participating in this research study. The study may help us understand the relationship between vocabulary support and reading comprehension.

Confidentiality: The information that you give in the study will be handled confidentially. Your data will be anonymous which means that your name will not be collected or linked to the data. Because of the nature of the data, it may be possible to deduce your identity; however, there will be no attempt to do so and your data will be reported in a way that will not identify you.

Voluntary participation: Your participation in the study is completely voluntary. Participation in this study will not affect your course grade in any way. The reading assessments are part of the course work, but the data will only be used for research if you agree to participate.

Right to withdraw from the study: You have the right to withdraw from the study at any time without penalty.

How to withdraw from the study: If you want to withdraw from the study, simply tell the instructor that you wish to withdraw and your data not to be used. There is no penalty for withdrawing.

Payment: You will receive no payment for participating in the study.

If you have questions about the study, contact:
Brendan Downey
11 Heron Ct. Charles Town WV 25414
Tel: (540) 905-8113
Email: bod4b@virginia.edu

Revision date: 11/01/11
Page 1

IRB-SBS Office Use Only		
Protocol #		
Approved	from:	to:
SBS Staff		

Project Title: The influence of the spatial contiguity of on-screen electronic glosses on reading comprehension with L2 learners of English in the United States of America

Faculty Advisor's Name: Ruth Ferree
Instructional Technology, 325 Bavaro Hall
University of Virginia, Charlottesville, VA 22903.
Tel: (434) 924-0853
Email: rf2c@virginia.edu

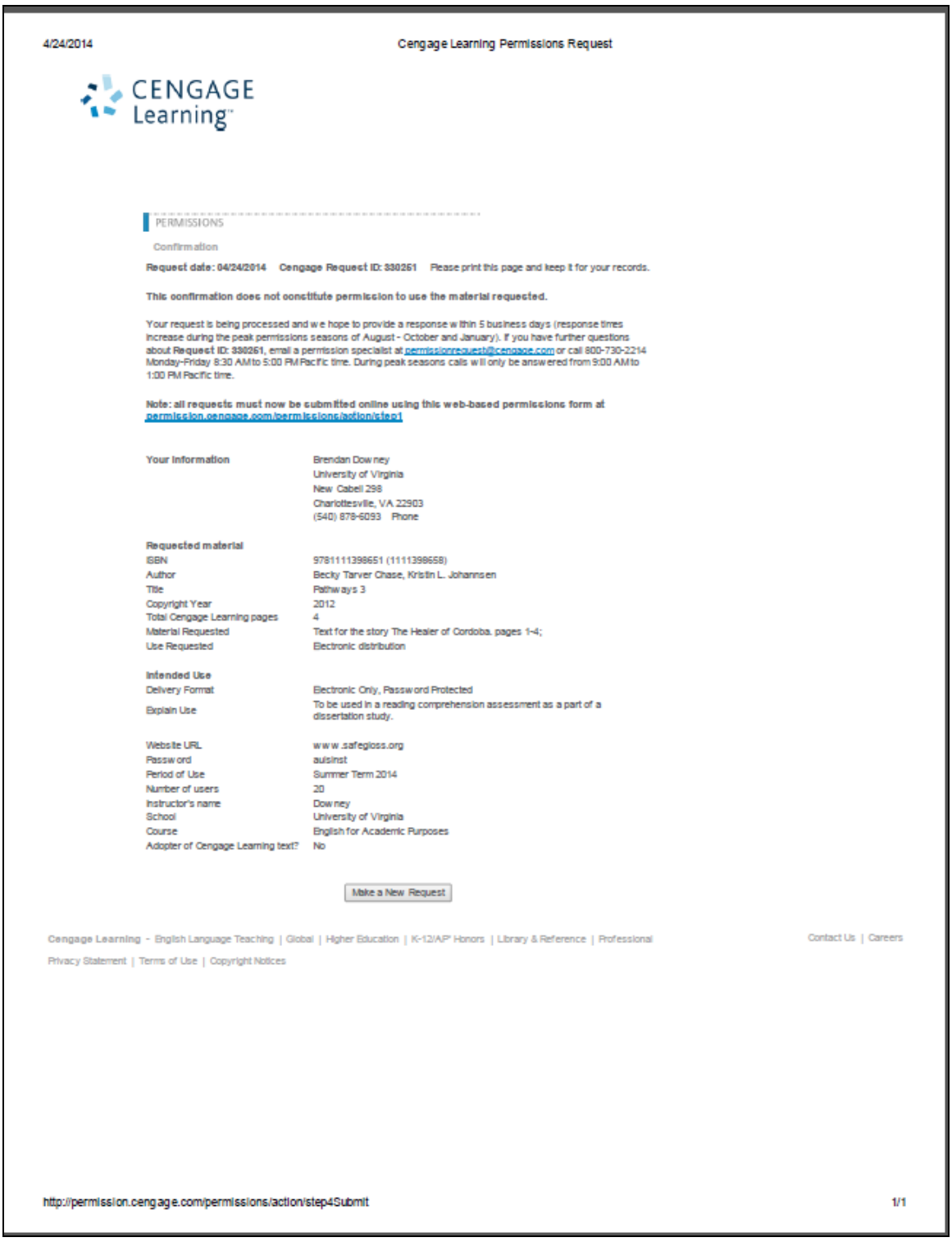
If you have questions about your rights in the study, contact:

Tonya R. Moon, Ph.D.
Chair, Institutional Review Board for the Social and Behavioral Sciences
One Morton Dr Suite 500
University of Virginia, P.O. Box 800392
Charlottesville, VA 22908-0392
Telephone: (434) 924-5999
Email: irbsbshelp@virginia.edu
Website: www.virginia.edu/vpr/irb/sbs

Agreement:
I agree to participate in the research study described above.
If you do not agree, please close the window and log-off your device.

IRB-SBS Office Use Only		
Protocol #		
Approved	from:	to:
SBS Staff		

Copyright Request and Permission



4/24/2014 Cengage Learning Permissions Request

CENGAGE Learning

PERMISSIONS

Confirmation

Request date: 04/24/2014 **Cengage Request ID: 330261** Please print this page and keep it for your records.

This confirmation does not constitute permission to use the material requested.

Your request is being processed and we hope to provide a response within 5 business days (response times increase during the peak permissions seasons of August - October and January). If you have further questions about **Request ID: 330261**, email a permission specialist at permissionrequest@cengage.com or call 800-730-2214 Monday-Friday 8:30 AM to 5:00 PM Pacific time. During peak seasons calls will only be answered from 9:00 AM to 1:00 PM Pacific time.

Note: all requests must now be submitted online using this web-based permissions form at permission.cengage.com/permissions/action/step1

Your Information	Brendan Downey University of Virginia New Cabell 298 Charlottesville, VA 22903 (540) 878-6093 Phone
Requested material	
ISBN	9781111398651 (1111398658)
Author	Becky Tarver Chase, Kristin L. Johannsen
Title	Pathways 3
Copyright Year	2012
Total Cengage Learning pages	4
Material Requested	Text for the story The Healer of Cordoba, pages 1-4;
Use Requested	Electronic distribution
Intended Use	
Delivery Format	Electronic Only, Password Protected
Explain Use	To be used in a reading comprehension assessment as a part of a dissertation study.
Website URL	www.safegloss.org
Password	auisinst
Period of Use	Summer Term 2014
Number of users	20
Instructor's name	Downey
School	University of Virginia
Course	English for Academic Purposes
Adopter of Cengage Learning text?	No

Make a New Request

Cengage Learning - English Language Teaching | Global | Higher Education | K-12/AP/ Honors | Library & Reference | Professional

Contact Us | Careers

Privacy Statement | Terms of Use | Copyright Notices

http://permission.cengage.com/permissions/action/step4Submit 1/1

CENGAGE Learning™

NELSON EDUCATION

IP Granting Dept
500 Terry A Francois Blvd, 2nd Floor, San Francisco, CA 94158
Phone: 800-730-2214 Fax: 800-730-2215
Email: permissionrequest@cengage.com

Submit all requests online at www.cengage.com/permissions.

Request # **330251**

04/28/2014

Brendan Downey
University of Virginia
School of Education - ASCIT
New Cabell 298
Charlottesville, VA 22903 United States

Thank you for your interest in the following Cengage Learning/Nelson Education, or one of their respective subsidiaries, divisions or affiliates (collectively, "Cengage/Nelson") material.

Title:	Pathways 3: Listening, Speaking, and Critical Thinking 1E		
Author(s):	Chase/Johannsen	ISBN:	9781111398651 (1111398658)
Publisher:	Heinle/ELT	Year:	2012
Specific material:	Text for the story The Healer of Cordoba. pages 1-4; To be used in a reading comprehension assessment as a part of a dissertation study.		
Total pages:	4		

For use by:

Name:	Downey
School/University/Company:	University of Virginia
Course title/number:	English for Academic Purposes
Term of use:	Summer Term 2014

Intended use:
For inclusion in a research project, master's thesis, or doctoral dissertation. May also be stored electronically for on-demand delivery through a dissertation storage system such as UMI system or as listed above. This permission is for non-exclusive rights for the US and Canada in English. Permission extends only to the work specified in this agreement, not to any future editions, versions, or publications. Applicant will not attempt to assign rights given herein to others, and the publication of this material in the work herein approved does not permit quotation therefrom in any other work. If, at a later date, a publishing contract is achieved, additional permission will be required.

The non-exclusive permission granted in this letter extends only to material that is original to the aforementioned text. As the requestor, you will need to check all on-page credit references (as well as any other credit / acknowledgement section(s) in the front and/or back of the book) to identify all materials reprinted therein by permission of another source. Please give special consideration to all photos, figures, quotations, and any other material with a credit line attached. You are responsible for obtaining separate permission from the copyright holder for use of all such material. For your convenience, we may also identify here below some material for which you will need to obtain separate permission.

This credit line must appear on the first page of text selection and with each individual figure or photo:

> From Chase/Johannsen. *Pathways 3*, 1E. © 2012 Heinle/ELT, a part of Cengage Learning, Inc. Reproduced by permission. www.cengage.com/permissions

Sincerely,

Jillian Shafer
Permissions Coordinator

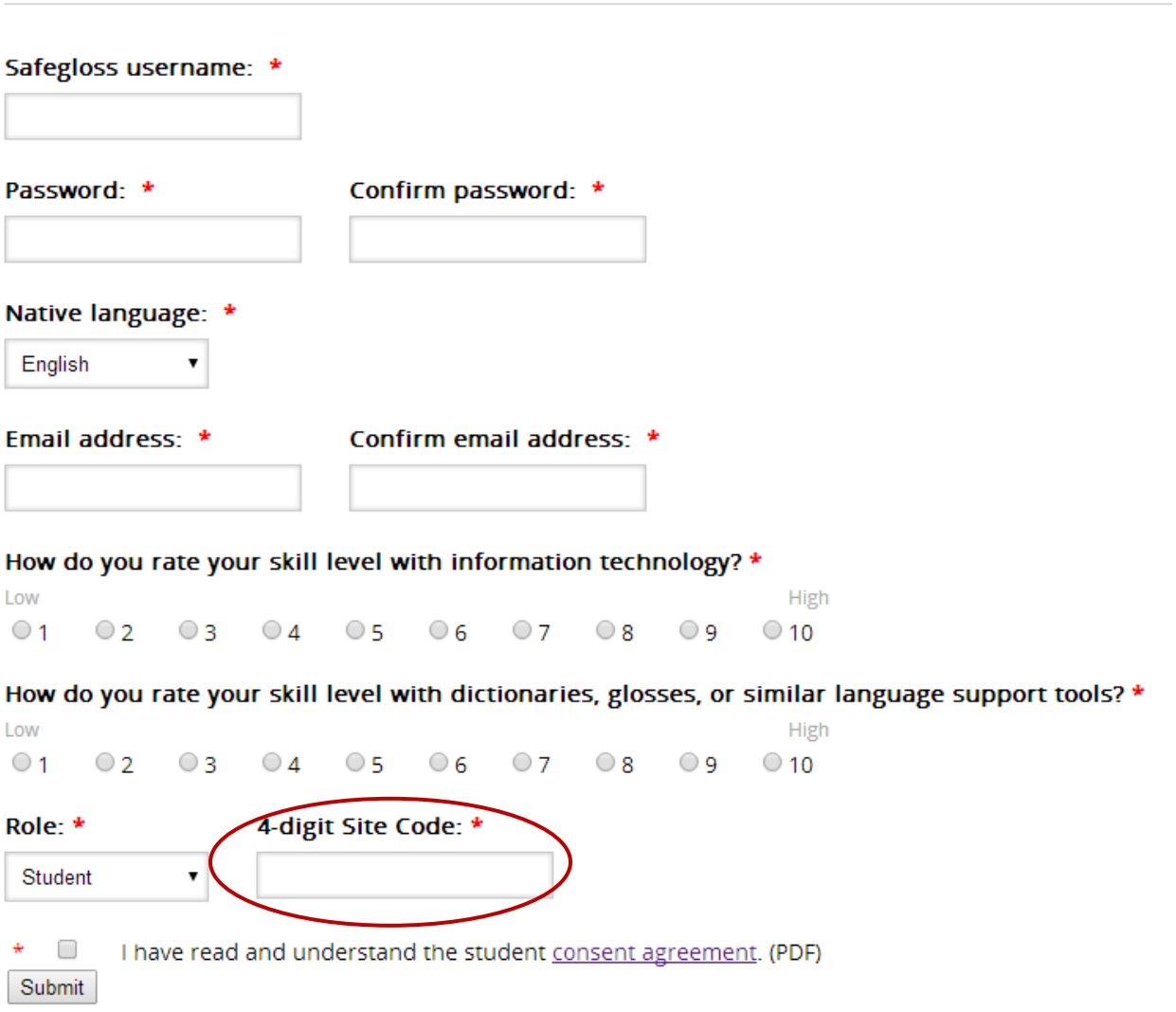
Safegloss.org orientation

Safegloss.org

Navigate to **http://www.safegloss.org** to access your activities. Before you can get started, you need to create an account. In order to create your account you will need to enter a username (which you will receive from your instructor) and password, answer a few questions, and then enter a Site Code (which you will also receive from your instructor). Your account on this system is anonymous. Click the Register tab to create your account.

Registering for an account

When you click the Register tab, you will see this form.



Enter your username

Enter the username provided by your instructor.

Choose a password

Enter a password, and then enter it again to confirm it. Note: you should create a strong, unique password.

Enter an email address

Enter an email address, then enter it again to confirm it.

Answer the two questions

Answer the two survey questions about your technology and language skill by selecting a number from 1 to 10 for each question, with 1 being a low level, and 10 being a high level.

Enter a Site Code

Enter the Site Code provided by your teacher.

Role

Select “student” from the drop down list.

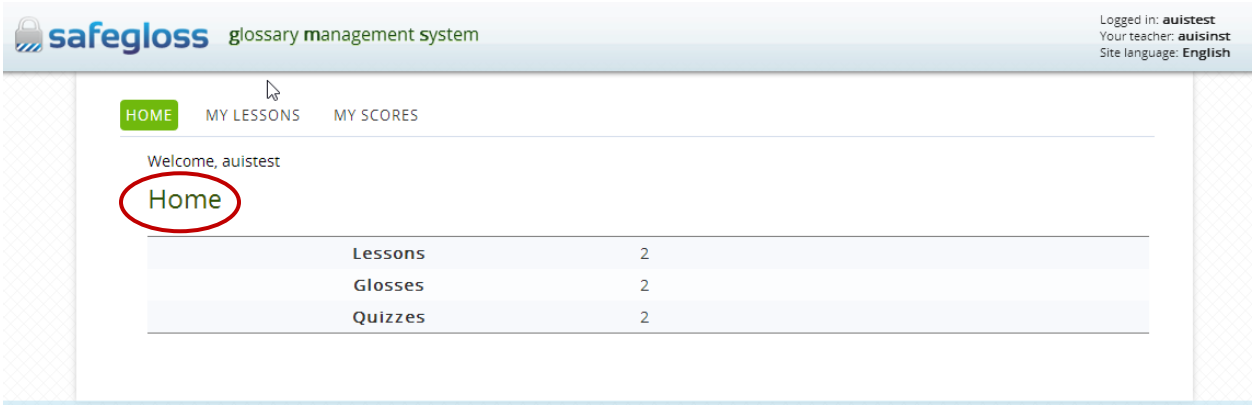
Acknowledge the terms of service

Read the informed consent agreement and print a copy for your records, if you wish.

Click the box to confirm your agreement. Click Submit.

Home

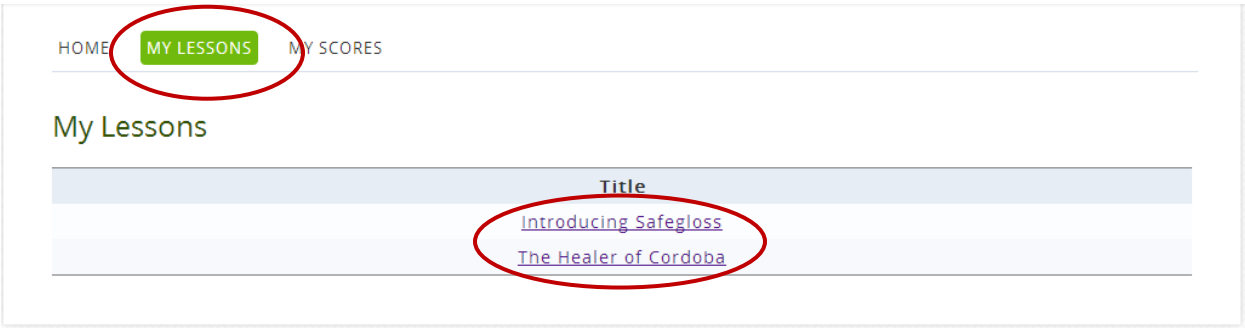
When you have logged in, you will see the Home screen of your account.



Your lessons

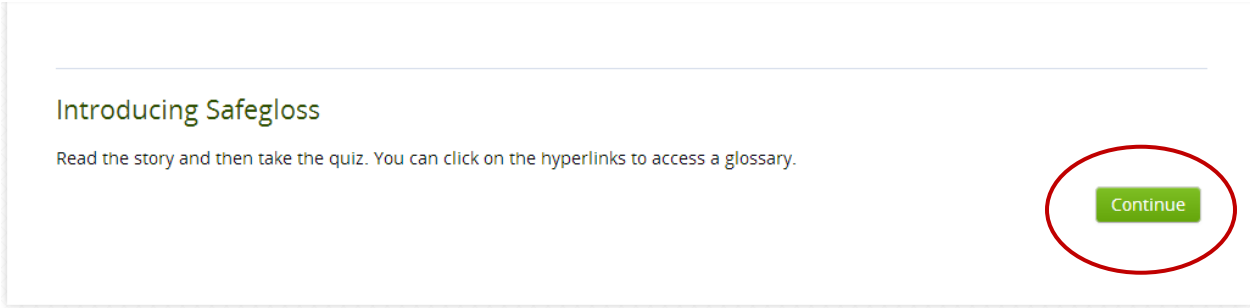
Click the My Lessons tab. Your lessons will appear in a list in the middle of the screen.

Click an activity to start. The first lesson, *Introducing Safegloss*, is optional.



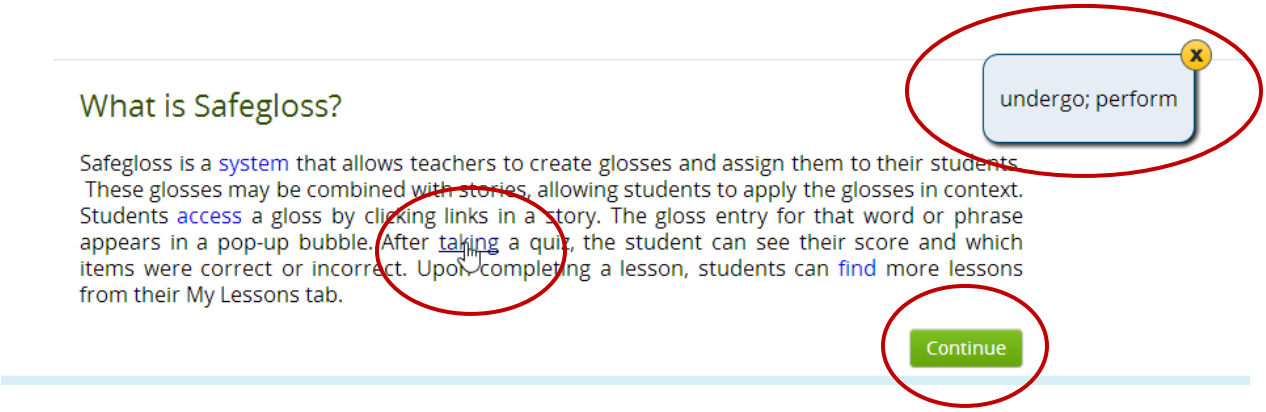
The lesson introduction

When you start a lesson, you will see a short introduction. Read the introduction and then click the Continue button to start the reading.



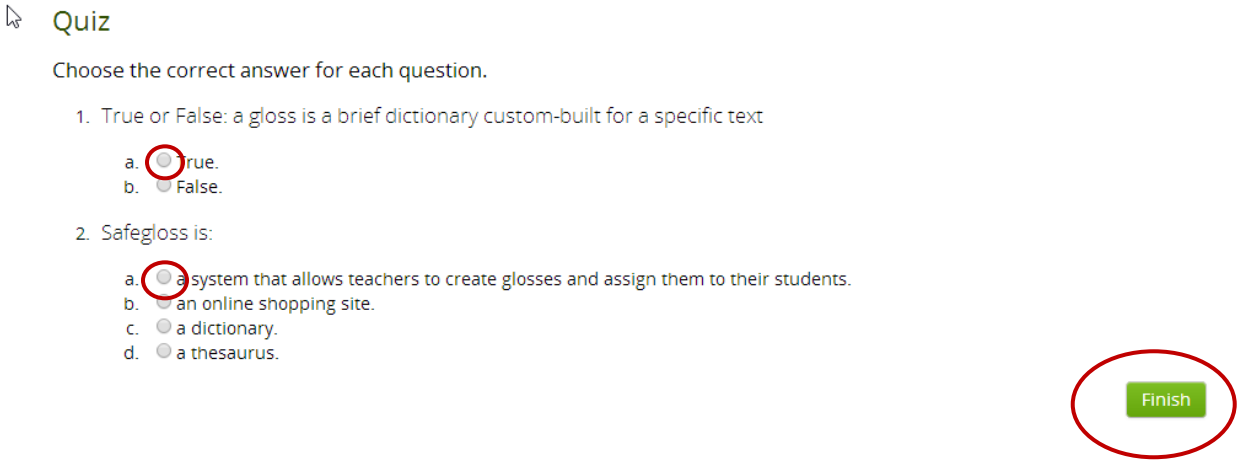
The reading and glossary

After the lesson introduction, you will see the reading. The reading contains special hyperlinks that can be clicked. When you click a hyperlinked word, you will see more information about that word. **You are encouraged to use this feature if it will help you understand what you are reading**. When you are finished reading, click the Continue button at the bottom of the page. This will take you to a comprehension quiz. You will not be able to return to the reading while you are taking the quiz, so do not click the button until you are ready.



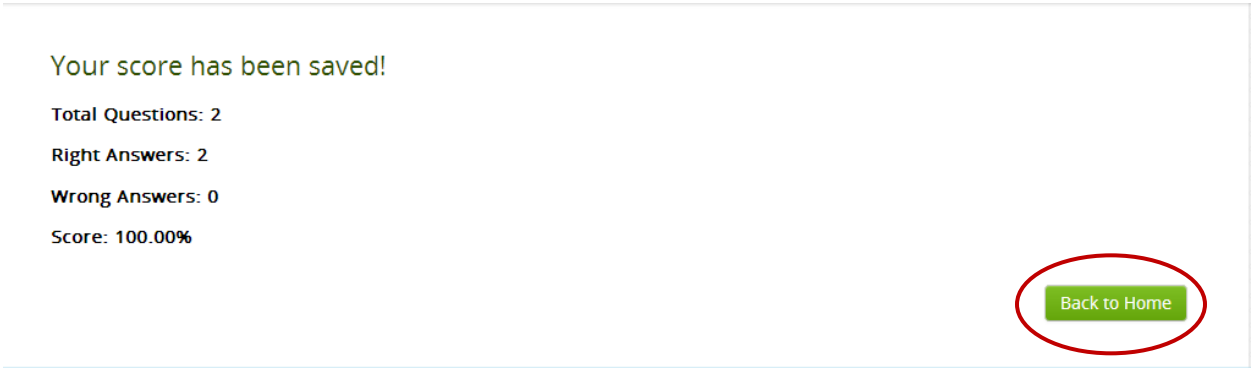
The quiz

The quiz will have several questions about the reading. Read each question and mark your answer. When you have completed the quiz, click the Finish button at the bottom of the screen.

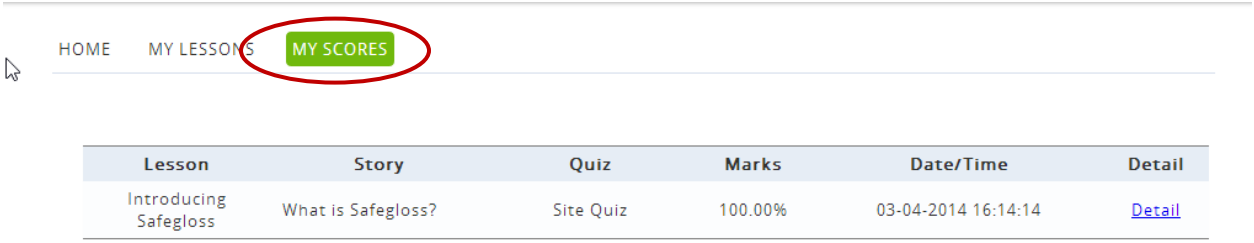


Your scores

After completing the quiz, you will see your score. Click the Back to Home button at the bottom of the screen to continue.

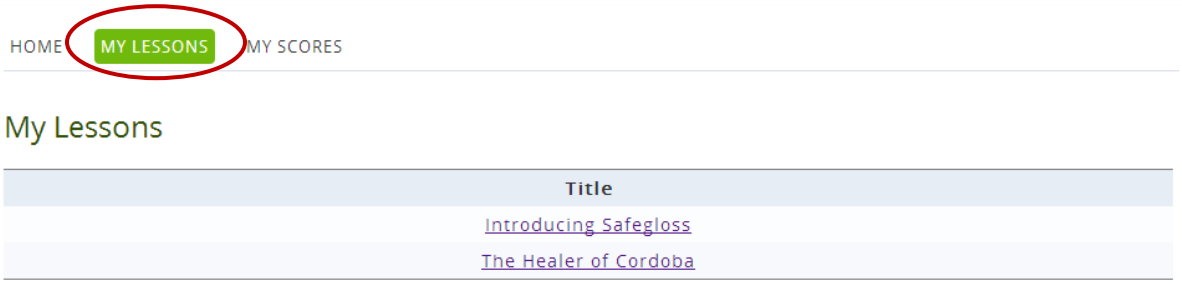


Your score will also be available from the My Scores tab of your home screen.



Starting another activity

If you wish to begin another lesson, click the lesson's name from the My Lessons tab.



Logging out

When you have finished your session, click the Logout button at the upper right of the screen.