Tibor Beres. DAIRSACC - Do Acronyms Influence Reading Speed and Content Comprehension? A Master's Paper for the M.S. in I.S degree. December, 2007. 105 pages. Advisor: Stephanie W. Haas

Acronyms, initialisms and other types of abbreviations are frequently used in scientific, academic, governmental and administrative setting to shorten lengthy terminology and nomenclature. While they can make a text easier to read for people familiar with the abbreviations, they can add to the text's inherent difficulty and impede comprehension for those who are not familiar with their meaning. The phenomenon of acronym polynymy (multiple definitions associated with the same acronym) can create confusion and add to the cognitive load associated with understanding the text. The current practice of defining acronyms only once, when introduced can result in readers scrolling back and forth in the text looking for acronym definitions, increasing the cognitive load and negatively affect reading speed and content comprehension. The purpose of this research was to study if the presence of a large number of acronyms in a text impedes reading performance. The current study also investigated if providing easy access to acronym definitions via hover text would alleviate comprehension problems caused by unknown acronyms in the text. The hypothesis was that by enabling fast acronym disambiguation, and eliminating the need to scroll for acronym definitions, the hover functionality would enhance reading speed and content comprehension. The results of the experiment are analyzed and recommendations for future investigations of the acronym problem are formulated.

Headings:

Acronyms

Acronym disambiguation

Reading performance

Content comprehension

DAIRSACC – DO ACRONYMS INFLUENCE READING SPEED AND CONTENT COMPREHENSION?

by
Tibor Beres

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

December 2007

Approved by

_____

Stephanie W. Haas

Introduction

Acronyms, initialisms and other types of abbreviations are frequently used in scientific, academic, governmental and administrative settings to shorten lengthy terminology and nomenclature. While they can make a text easier to read for those familiar with the meaning behind abbreviations, they are meaningless or ambiguous for those who are unfamiliar with them. Ambiguous terminology increases the cognitive load associated with understanding a text (Mason & Just, 2007).

The current practice of defining an acronym in written text only once, when first introduced, can also add to the inherent difficulty factor of the text, as readers might have to scroll or page for the definition. Scrolling or paging increases the demand on the short term memory (also referred to as *working memory*) as readers will have to simultaneously store the acronym they are trying to disambiguate, its associated definition or expansion, and their place in the text (Mason & Just, 2007). The effects of various factors influencing the cognitive load is additive and the more resources invested in one sort of cognitive activity (deciphering acronyms, for instance) the less resources available for other cognitive processes, such as understanding the meaning of text (Shiffrin & Nosofsky, 1994; Mousavi, Low & Sweller, 1995; Baddeley, 2003; DeStefano & LeFevre, 2007). In other words, the more readers have to employ their short term memory when reading a text, the fewer resources can be allocated for the complex activities of reading and comprehending. According to the cognitive load theory, the

increased cognitive load due to the large number of unknown acronyms in the text (more cognitive resources allocated to acronym disambiguation and short term memory) should result in decreased reading performance and inferior text comprehension.

The purpose of this study was to determine if the presence of a large number of unknown acronyms and initialisms in a text, along with the practice of defining them only once when first introduced would negatively affect reading performance and content comprehension. In addition, we set out to test if providing instant access to acronym definitions as hover text upon mouse-over (eliminating the need for scrolling for acronym definitions or remembering acronym definitions while reading) would reduce the cognitive load and therefore improve reading speed and content comprehension, as suggested by the cognitive load theory.

Use of Abbreviations in Text and Associated Factors

Before analyzing the factors associated with abbreviations in text we need to define what abbreviations (and its different subclasses) are. The Terminology subsection contains formal definitions for abbreviations, acronyms and initialisms, and introduces additional terms that can be used interchangeably when talking about acronyms or initialisms.

Terminology

The Oxford English Dictionary contains the following definitions for abbreviations, acronyms and initialisms.

> *Abbreviation* - 1. The act of shortening, reducing in length; 2. The result of abbreviating; an abbreviated or reduced form; short summary, abridgement[1].
>
> *Acronym* - A word formed from the initial letters of other words[2].
>
> *Initialism* - The use of initials; a significative group of initial letters. Now *spec.* a group of initial letters used as an abbreviation for a name or expression, each letter or part being pronounced separately (contrasted with ACRONYM)[3].

While there are differences in the formal definition of abbreviations, acronyms, and initialisms, for the purposes of our study it makes no difference into which category the abbreviation falls. They present the same challenges for the readers when it comes to content comprehension. Throughout this paper we will blur the differences between these terms and use the terms interchangeably. As acronyms and initialisms are contractions of a longer name or expression, and in accordance with the terminology of Chang, Schutze

& Altmann (2002) and Schwartz & Hearst (2003), throughout this paper we will also use the terms *short form* when referring to acronyms, and *long form* when referring to acronym definition or expansion. For example, *DAIRSACC* (short form) is the abbreviation for *Do Abbreviations Influence Reading Speed and Content Comprehension* (long form).

Acronyms in Written Text

The following section discusses the problematic of using acronyms in written text. This section is important as it sheds some light on why the usage of acronyms in a text can create comprehension difficulties for readers who are not familiar with their meaning.

Acronyms and initialisms do not necessarily have semantic meaning on their own although they can take the form of genuine words, such as *RACE* for *Registry of Approved Continuing Education* or *PEACE* for *Parental Encouragement for Autistic Children Everywhere*. They are dynamically assigned meaning, usually with a definition or an expansion – for example *GUI stands for Graphical User Interface*. In writing, acronyms or their expansions are usually enclosed within parentheses, such as *Federal Emergency Management Agency (FEMA)* or *MLB (Intermediate Life Boat)*.

As there are no strict rules governing acronym creation and authors are more concerned with functionality than anything else (making the acronyms easy to remember), acronyms are created using a variety of methods. Acronyms can be formed by combining the initial letters of words (*ROM* for *Read Only Memory*); combining parts of words (*radar* for *radio detecting and ranging*); combining initial letters with parts of the words (*CHEMTREC* for *Chemical Transportation Emergency Center*), or combining letters from anywhere in the name (*XML* for *eXtensible Markup Language*). They can be

nested – one acronym incorporating another acronym (*AIM* for *AOL Instant Messenger* where *AOL* is an acronym for *America Online*). They can ignore parts of the long form (*ATL* for *Adult T-cell Leukemia*). Acronyms can contain letters or numbers that are not part of the long form (*CSNBX* for *Congenital Stationary Night Blindness*; *RB1* for *retinoblastoma*). Letters in the acronym can have a different order than the words in the definition (*AW* for *water activity*). Acronyms can also be recursive (*PHP* for *PHP Hypertext Preprocessor*).

As acronyms are defined in a given context, they are also ambiguous, meaning that there can be several definitions associated with the same short form in different contexts (phenomenon known as *acronym polynymy*). Conversely, several acronyms can be associated with the same definition or expansion. Repeated usage of an acronym can result in the acronym gradually acquiring the semantic meaning of the long form and establishing itself as a genuine word, eventually replacing the long form in the language altogether. Examples of this include *ZIP* for *Zoning Improvement Plan* and *laser* for *Light Amplification by Stimulated Emission of Radiation*. The association between long form and short form is no longer required and gradually the long form might even confer less semantic meaning than the short form - *IBM* is probably better known and its meaning better identified throughout the world than *International Business Machines* and probably few people know the first names of the four artists who formed the Swedish pop-group *ABBA*, the source for the ABBA acronym.

There are also instances where the acronyms are kept but the associated expansions, or definitions are changed (*UAE – Unix Amiga Emulator*, later *Universal Amiga Emulator*). Sometimes acronyms are kept and the associated long forms

eliminated altogether (*SAS* – originally an acronym for *Statistical Analysis Software* became the name of the company who developed the software).

This section illustrated why it is important to provide acronym definitions when using acronyms in a text. Given the wide variety of practices in place for acronym creation and the phenomenon of acronym polynymy, authors cannot expect readers to be able to easily decode the meaning behind the acronyms. Acronyms can be meaningless or ambiguous to readers and writers should ensure that they always provide the means to disambiguate the meaning of all of the acronyms in the text.

Acronyms, Short Term Memory and Cognitive Load

As acronyms accrue semantic content through the associated definitions the process of becoming familiar with the meaning of an acronym does not differ significantly from the process of learning new words. A most simplistic view of this process is that when encountering unknown acronyms, people rely on their short term memory to temporarily store the acronym, its definition and the associated information.

The problem of short term memory capacity influencing information processing has been extensively researched in the field of psychology ever since Miller's groundbreaking article "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information" was published in 1956. Miller coined the term of *amount of information* to quantify informational content that gets processed, and the term of *chunks of information* to quantify the capacity of short-term memory and argued that the capacity of our working memory greatly affects the amount of information we are able to process at any point of time.

Although Miller (1956) pointed to the *magical number 7 plus minus 2*, as the number of information chunks our immediate memory can simultaneously process, subsequent studies suggest that good performance is tied to a smaller number - *magical number 4 plus minus 1* (Cowan, 2001a), *magical number two or three* (Lewis, 1996), or tied to the degree to which informational chunks integrate with prior knowledge (Baddeley, 1994). The process of *recoding* in Miller's terminology refers to our ability to combine individual pieces of information into larger, more meaningful chunks where the emerging construct's semantic meaning is more than the sum of the individual pieces of information. By recoding the individual stimuli we are able to improve the performance of our working memory. This process of recoding is enhanced by prior knowledge (Miller, 1956).

Short term memory span seems to be influenced by word length as well; the longer the words, the more limited our capacity to process (Baddeley, 1994). The working memory is also affected by interference from additional stimuli, the quality of the stimuli, the sequence or order of different stimuli and by the passage of time (Shiffrin & Nosofsky, 1994). Additional information coming into our short term memory (a new acronym for example) would push earlier chunks out.

The mere passage of time affects our ability to keep information in our working memory, although it is debated if we are replacing informational content because it expires with time or because constantly interfering factors get our attention (Shiffrin & Nosofsky, 1994).

The multitude of mental processes involved in recoding and processing of information, (thinking or reasoning in general) has been labeled *cognitive load* in the

field of psychology. According to cognitive load theory, the performance of all complex mental processes such as reading, problem solving, thinking in general, is affected by cognitive load (Shiffrin & Nosofsky, 1994; Mousavi et al., 1995; Baddeley, 2003; DeStefano & LeFevre, 2007).

When talking about the factors influencing the cognitive load accompanying complex mental processes such as information processing, the literature mentions *intrinsic* and *extraneous* factors. Intrinsic factors refer to properties of the informational content, such as the informational load – the amount of information incorporated in the text or stimuli. For example, this master's paper's informational load is obviously higher than the informational load of the sentence *Today is my birthday*. The cognitive load attributed to intrinsic factors is determined by subjective factors such as prior knowledge, verbal skills, analytical skills, etc. (Gillström & Rönnberg, 1995; McCrudden, Schraw, Hartley & Kiewra, 2004; Braten & Stromso, 2006).

Extraneous factors contributing to the cognitive load lie in the presentational method of the information (written document or a presentation, typography, pitch, "noise" in the environment or the medium, etc). The effect of each individual factor contributes to the overall cognitive load and the effect of these factors is additive (McCrudden et al., 2004; DeStefano & LeFevre, 2007).

The rather limited capacity of our working memory and the multitude of factors influencing its performance suggest that text containing a multitude of unknown acronyms could potentially pose significant comprehension problems by increasing the cognitive load. The current practice of defining acronyms only once, when first

introduced, might not serve readers' best interests of quickly learning the acronyms and the associated definitions, as it potentially can contribute to increased cognitive load.

The Acronym Problem in the Bioinformatics Literature

There is increased interest in the recent bioinformatics literature towards the problem of automatic identification of acronyms in electronic text. This interest is fueled by the information retrieval problems caused by the unprecedented rate at which new knowledge is published in fields such as medicine, biotechnology, etc. and the proliferation of acronym usage accompanying this process. Chang, Schutze & Althman, (2002) estimated that the number of citations in MEDLINE grows by 400,000 every year and the growth rate of both abstracts and new abbreviation definitions in MEDLINE was increasing. They estimated that in 2001, 64,262 new abbreviations were introduced into MEDLINE. Zhou, Torvik & Smalheiser (2006), citing Stead, Kelly & Kolodner (2005) put the number of new articles published in the field of biomedicine to over 900,000 a year. Chang et al. (2002) reported that their algorithm extracted 1,948,246 abbreviations from MEDLINE when scanning only abstracts published until the end of 2001. Adar (2004) reported that the acronym extraction engine of the Simple and Robust Acronym Dictionary (SaRAD)[1] extracted 3,960,168 abbreviation-definition pairs from MEDLINE documents published up to January 2002. As of July 2007, the Stanford Biomedical Abbreviation Database[2] contained 2,074,367 abbreviations from the biomedical field, according to information posted on their website.

The phenomenon of acronym polynymy further escalates the problems created by the acronym proliferation. Several published articles estimate the percentage of acronyms

[1] SaRAD. Demo page available at http://www.hpl.hp.com/research/idl/projects/abbrev.html
[2] The Stanford Biomedical Abbreviation Database. http://abbreviation.stanford.edu

with more than one definition associated at more than 20%. Chang et al. (2002) estimated that 21.7% of the 1,948,246 acronyms in MEDLINE were associated with more than one definition. Wren & Garner (2002) estimated that acronym polynymy in MEDLINE was as high as 30% and that about 10% of the definitions were associated with multiple acronyms. This semantic ambiguity creates major problems from information retrieval point of view.

The ambiguity in terminology negatively influences recall and precision of information and knowledge management systems (Wren & Garner, 2002; Liu, Aronson & Friedman, 2002; Adar, 2004; Ao & Takagi, 2005; Zhou et al., 2006). In the absence of a unified or normalized terminology, keyword and index normalization is rather difficult if not impossible. As a consequence, documents on the same topic using different terminology (e.g., different acronyms to denote the same concept, such as different acronyms for a chemical compound) will not be indexed with the same keywords. Conversely, documents on different topics which contain the same abbreviations might get indexed and clustered based on these keywords. As an example, when searching for *microlaryngotracheobronchoscopy* with its acronym *MLB* as the query keyword, one might indeed retrieve documents on microlaryngotracheobronchoscopy but also documents on Major League Baseball, Minor League Baseball, Mother Love Bone band, Motor Life Boat, Main Logic Board, Mark Lanegan Band, Middle Linebacker (football), Maritime Lumber Bureau, Multi-Layer Board, Melbourne Regional Airport (Airport Code), Multi-Line Business, Metallic Link Belt, March of the Light Brigade, Michael Brittin Enterprises (NASA), Movable Lane Barrier, Multinational Logistics Base, Intermediate Life Boat, Mobile Logistics Support, and Median Longitudinal Bundle.

(This list includes only the verified definitions associated with the MLB acronym featured in the human verified list of acronyms on Acronym Finder[3].)

Even when searching specialized databases or text collections such as MEDLINE, the lack of normalized terminology will affect recall and precision – someone searching for *oral allergy syndrome* with the acronym *OAS* will also retrieve documents containing OAS for *O-acetyl-L-serine*, OAS for *2'-5' oligoadenylate synthetase* or OAS for *Overt Aggression Scale*. On the other hand, someone searching for *neuropilin-1* with its *np1* acronym might not retrieve those documents where neuropilin-1 is abbreviated as *NRP-1*, *NRP1*, *Nrp-1*, *Nrp1*, *NP-1*, *Np-1*, *npn-1* or *Npn1* (all these are valid acronyms for neuropilin-1 that can be found in MEDLINE).

The existence of undefined acronyms makes it difficult to disambiguate acronyms and to automatically identify and extract acronyms from text, as the acronym extraction algorithms assume that the acronyms are associated with a definition at least once (Adar, 2004). Undefined acronyms are regarded as false positives by these algorithms and are discarded.

Acronym polynymy negatively influences prospective data mining as well (Wren et al., 2002; Chang et al., 2002; Liu et al., 2002). Documents on oral allergy syndrome containing the acronym OAS might get clustered together with documents containing OAS for Overt Aggression Scale while documents on completely different topics might get clustered together if their terminology incorporates the same acronyms.

Another problem worth mentioning is that optical character recognition (OCR) devices are unable to properly deal with acronyms and abbreviations in a text (Taghva & Gilbreth, 1999). OCR devices can be used to digitize resources or as reading aid by

[3] Acronym Finder. http://www.acronymfinder.com

people with visual impairments. OCR devices scan the documents page by page and convert the images into ASCII text. Errors can occur during the conversion process, due to erroneous character recognition. Although using a comprehensive dictionary can help OCR devices in recognizing words, because acronyms and abbreviations are not featured in dictionaries, their presence in a text makes the correction procedure more difficult (Taghva & Gilbreth, 1999).

Manually developing and maintaining up-to-date acronym and abbreviation glossaries, taxonomies and ontologies is not feasible. Given the unprecedented rate at which new knowledge gets published and the acronym proliferation and polynymy characterizing these publications, it is virtually impossible to manually develop and maintain up-to-date acronym glossaries or taxonomies. (Wren et al., 2005; Chang et al., 2002; Liu et al., 2002; Yeates, 1999; Schwartz & Hearst, 2003; Zhou et al., 2006; Zahariev, 2004; Xu & Huang, 2007).

Finally and especially related to this study, the presence of acronyms and other abbreviations in the text can cause comprehension difficulties for readers not familiar with their meaning (Larkey, Ogilvie, Price & Tamilio, 2000; Chang et al, 2002; Mason & Just, 2007; Xu & Huang, 2007).

The increased interest in developing an automated tool to identify acronyms and associated definitions is fueled by the potential benefits both for the information, knowledge management systems and for the readers. Developing an automatic acronym extraction engine could be a low cost alternative for the costly process of manually developing and maintaining acronym glossaries, taxonomies, ontologies (Taghva & Gilbreth, 1999; Chang et al., 2002; Wren et al., 2005; Liu et al., 2002; Yeates, 1999;

Zahariev, 2004; Adar, 2004). An acronym extraction engine could be used as a foundation for developing a knowledge management system to provide helpful functionalities such as automatic acronym sense disambiguation (Xu & Huang, 2007), display of acronym definitions (Taghva & Gilbreth, 1999; Yeates, 1999), search-by-acronyms indexes (Yeates, 1999), or hyper linking documents containing the same acronym (Taghva & Gilbreth, 1999). Automatically extracting acronyms and their definitions would enable the correct mapping of the various long forms associated with the various short forms and would offer additional arguments for data-mining tools, enhancing their productivity (Taghva & Gilbreth, 1999; Chang et al., 2002; Ao & Takagi, 2005; Liu et al., 2002; Zhou et al., 2006).

The remainder of this section discusses some of the notable efforts for developing acronym finding algorithms and automated tools for acronym glossary and dictionary generation. Most algorithms and databases presented in this section were developed in an effort to enable the automatic extraction of acronyms and their associated definitions from MEDLINE, in order to enhance search engine recall and precision.

The Acronym Finding Program (AFP), developed by Taghva & Gilbreth (1999) is one of the earliest algorithms developed. AFP considered a string of uppercase letters as a candidate acronym and based on probabilistic matching, tried to find matching words for a definition. Taghva & Gilbreth reported a precision of 98% and a recall of 86% when tested on a collection of governmental collection documents consisting of 1,328 documents.

The Three Letter Acronym (TLA) system was developed at the University of Waikato in New Zealand (Yeates, 1999). TLA identified candidate acronyms and long

forms by looking for parentheses and periods in the text. TLA then broke up the text in chunks, where the parentheses and periods marked the beginning and the end of each chunk. Each word in each chunk was considered a potential acronym and was compared to the preceding and following chunk in an attempt to find matching definitions. TLA tried to eliminate false positives based on a set of heuristic rules, such as that acronyms are shorter than their definitions, they contain initials of the words in their definitions, or they are given in upper case. TLA's algorithm produced a recall of 91% and a precision of 68% with testing material from the New Zealand Digital Library (Yeates, 1999).

Acrophile[4] is an online database of acronyms and their definitions, developed at the University of Massachusetts (Larkey et al., 2000). At the core of Acrophile is an automatic acronym extracting tool that parses web pages or document collections for candidate acronyms and acronym definitions. The database can be constantly updated by passing candidate acronyms as queries to a search engine (AltaVista) and following the top *n* returned URLs. Larkey compared the performance of Acrophile to two manually developed online acronym databases: WWWAAS (World Wide Web Acronym and Abbreviation server)[5] at University College Cork in Ireland, and Acronym Finder[6], Mountain Data Systems' acronym database. The authors concluded that the recall and precision of Acrophile easily surpassed WWWAAS and rivaled the performance of Acronym Finder, with the added benefits of Acrophile being easily updatable and maintainable.

[4] Acrophile. http://ciir.cs.umass.edu/ciirdemo/acronym/
[5] World Wide Web Acronym and Abbreviation Server (WWWAAS). http://www.ucc.ie/cgi-bin/acronyms
[6] Acronym Finder. http://www.acronymfinder.com/

The Acronym Resolving General Heuristic (ARGH)[7] program uses a set of heuristic recognition and refinement rules for automatically identifying algorithms and associated definitions (Wren et al., 2005). ARGH assumes that acronyms are defined in standard format, using parentheses either for the short forms or for the long forms, and checks a window of text to the left and to the right of the parentheses, trying to match candidate acronyms with candidate definitions. In addition to returning all the associated long forms from the database, ARGH returns frequency statistics and examples for usage context to enhance sense disambiguation. ARGH contained approximately 257,000 unique biomedical acronyms and abbreviations extracted from MEDLINE as of August 2007, according to information posted on its home page.

The Stanford Biomedical Abbreviation Database[8] contained 2,074,367 abbreviations as of August 2007 according to information posted on its website. The database was dynamically constructed with an algorithm developed by Chang et al. (2002). The Chang algorithm uses a statistical learning algorithm and logistic regression to score candidate abbreviations and candidate definitions on their resemblance to a training set of expert-annotated abbreviations. The algorithm achieved 82% recall and 99% precision when run on MEDLINE abstracts published until the end of 2001 (Chang et al., 2002).

AcroMed[9], developed through collaboration between computational linguists at Brandeis University and biologists at Tufts University School of Medicine aims at improving access to relevant information in the biomedical field by automatically extracting and classifying acronyms/acronym definitions from MEDLINE[4]. Like

[7] ARGH: Biomedical Acronym Resolver. http://invention.swmed.edu/argh/
[8] The Stanford Biomedical Abbreviation Database. http://abbreviation.stanford.edu.
[9] Acromed. http://medstract.med.tufts.edu/acro1.1/index.htm

ARGH, AcroMed returns the frequency with which an acronym occurs in the database. AcroMed's automated classification is based on an ontology composed by UMLS (Unified Medical Language System) and GO (Gene Ontology) taxonomic terms (Wren et al., 2005). At the time of this research (August, 2007) AcroMed was available online at http://medstract.med.tufts.edu/acro1.1/index.htm but was not returning any results.

The Simple and Robust Abbreviation Dictionary (SaRAD)[10] developed by HP Laboratories in Palo Alto, California, uses a three step process to optimize its output (Adar, 2004). SaRAD starts by extracting candidate short and long forms using regular expressions (looking for a parentheses and at least one capital letter). In step 2, related abbreviation- definition pairs, such as plurals of simple abbreviations, are clustered to improve recall and precision. Step 3 refers to optimizing the output, making the data accessible for human or computer processing. SaRAD was tested on a data set containing 11 million abstracts from MEDLINE producing a precision of 95% and a recall of 85%. Adar mentioned that the recall of acronyms was also affected by the existence of undefined acronyms in MEDLINE.

The Abbreviation LIfter using Corpus-based Extraction (ALICE)[11] engine was developed in Japan as part of the PubMed Enhancer Toward Efficient Research (PETER) system. ALICE identifies candidate acronyms by looking for parentheses. Based on a set of heuristic pattern-matching rules, it tries to identify the associated long form within the same sentence. ALICE achieved 95% recall and 97% precision on randomly selected titles and abstracts from MEDLINE (Ao & Takagi, 2005).

---

[10] SaRAD. Demo page available at http://www.hpl.hp.com/research/idl/projects/abbrev.html
[11] ALICE. http://uvdb3.hgc.jp/ALICE/ALICE_index.html

Another Database of Abbreviations in MEDLINE (ADAM)[12], developed at the University of Illinois in Chicago, is another example of a database generated with an automated tool for acronym extraction (Zhou et al., 2006). ADAM's algorithm is a good illustration for the complexity of the task of automatically extracting acronyms from text, given the variety of practices involved in creating acronyms. ADAM's algorithm involves 5 steps, each step presenting its own set of challenges and difficulties. In step 1 candidate acronyms are extracted by looking for words in parentheses. In step 2 candidate long forms are identified within 3*N* words distance towards the left of the parentheses, where *N* is the number of letters in the candidate acronym. In step 3 short form and long form pairs are associated, and false positives are filtered out based on a complex scoring and cut-off algorithm. In step 4 the algorithm verifies that the short forms are used in the text separately from the long forms. In step 5 morphologically similar long forms, corresponding to the same acronym (or its lexical variants) are grouped together to improve recall and precision. Zhou et al. reported a precision of over 97% for the algorithm.

The support vector machine (SVM) model developed by Xu Jun and Yalou Huang identifies potential acronyms based on a set of pattern-matching rules. The algorithm incorporates rules such as acronym length should be between two and ten characters; acronyms contain at most one white space; an acronyms' first character is alphabetic or numeric; acronyms contain at least one capital letter; acronyms are not a known word in dictionary, are not person names, location names or words in a predefined list of stop words. Candidate acronyms and definitions are evaluated based on two

[12] ADAM: Another Database of Abbreviations in MEDLINE. http://128.248.65.210/arrowsmith_uic/adam.html

parameters, *maxoffset* (the maximum distance between acronyms and definitions) and *maxlength* (the maximum length of the definitions). The system is trained on a set of pre-selected acronym – definition pairs and the automatically extracted sets are compared to the training set. While SVM involves human input for the training of the system, making it potentially more costly to implement in the initial phase, it has the advantage that its efficacy can be improved with training and since its method is generic it can easily adapt to other domains (Xu & Yaluo, 2007).

In addition to the above mentioned resources for acronym/abbreviation disambiguation, a simple Google search on the Internet returned several other resources, some of them developed based on user contributions.

Acronym Finder[13], Mountain Data Systems' acronym database, mentioned by Larkey et al. (2000), contained more than 565,000 human-edited entries according to information posted on their website at the time of this research (August, 2007). The web site claimed that together with Acronym Attic[14], containing the unverified and unedited acronyms collected by Mountain Data Systems over the years, the site contained more than 4 million acronyms as of August, 2007.

STASA (Scientific and Technical Acronyms, Symbols and Abbreviations) Acronym Finder[15], provided by Wiley InterScience, contained over 200,000 of the acronyms most frequently used in science, according to information posted on their website at the time of this research (August, 2007).

---

[13] Acronym Finder. http://www.acronymfinder.com/
[14] Acronym Attic. http://www.acronymattic.com/
[15] STASA Acronym Finder. http://www3.interscience.wiley.com/cgi-bin/mrwhome/104554766/HOME

The Internet Acronym Server[16], mentioned by Larkey et al. (2000), has been collecting acronyms since 1988 according to information posted on the site. The site offers an interface for visitors to submit acronyms for review. It was impossible to determine from the site the number of acronyms contained in the database or if the Internet Acronym Server was specializing in any domain or not.

Abbreviations.com[17] offered search and browse by categories functionalities, 'abbreviation- word', 'word- abbreviation', and 'word in definition' search options. Main categories for browsing were Computing, Business, Internet, Governmental, Academic & Science, Community, Miscellaneous, Regional, Medical, International, with subcategories under the main categories, such as Hospitals, Laboratory, Physiology, Human Genome, Oncology and Veterinary subcategories under the Medical category. There were 6,494 entries listed under the Physiology subcategory and 1,849 entries listed under the Veterinary subcategory in August, 2007. Users could also browse abbreviations with an A-Z list. An additional piece of information to illustrate the number of abbreviations in Abbreviations.com - there were 17,209 entries under the letter A in August, 2007.

Answers.com[18], standalone at http://www.answers.com or incorporated into the Google search engine[19], can also be used to find common acronym definitions. The number of resources available online for acronym disambiguation indicates a general recognition of the acronym problem.

---

[16] Internet Acronym Server, http://silmaril.ie/cgi-bin/uncgi/acronyms
[17] Abbreviations.com. http://www.abbreviations.com/
[18] Answers.com. http://www.answers.com
[19] Google.com. http://www.google.com

In this section we have presented some of the efforts directed towards automatically identifying abbreviations and associated definitions in text. Most of the algorithms presume a somewhat standardized modality of defining abbreviations in text – abbreviations in all caps (AFP), abbreviations or associated long forms surrounded by parentheses (TLA, SaRAD, ARGH, ALICE, ADAM), abbreviations and associated definitions being adjacent in the text (TLA, ARGH, ADAM, SVM), abbreviations containing letters taken from the definitions (TLA, SaRAD, ALICE), etc. Some of the presented acronyms obtained precision and recall of above 90% when tested (ADAM, ALICE, SaRAD) but none of the tools could guarantee 100% efficiency. None of the algorithms developed so far can counter the problems created by acronym proliferation and acronym polynymy in literature.

Reading Speed and Content Comprehension

Information presentation format affects the cognitive load associated with processing information (McCrudden et al., 2004; DeStefano & LeFevre, 2007). Factors such as display media characteristics (monitor size, resolution, luminescence, contrast) text characteristics (length), text layout characteristics (margins, alignment, line length, column length), font typography, presentational format (vertically versus horizontally, one word at a time, one line at a time, one page at a time, continuous text, scrolling text, etc.) are all extraneous factors whose individual effects add up and contribute to the overall cognitive load. As the cognitive load influences reading speed and content comprehension, we turned to the literature for guidelines on how to minimize the effects of these extraneous factors on cognitive load.

The effect of various factors on reading speed and content comprehension has been in the purview of the Information Science literature for the last 20 years, while also comparing reading performance with different display media (paper versus computer monitor).

Reading from Paper versus Reading from a Computer Screen.

Although most of the studies in the field suggest that reading from a computer screen is significantly (10%-30%) slower than reading from paper (Gould, Alfaro, Barnes, Finn, Grischkowsky & Minuto, 1987; Muter & Maurutto, 1991; Ziefle, 1998; Kurniawan & Zaphiris, 2001), there is also evidence in the literature to suggest that

reading speed in not affected by presentation medium (Askwal, 1985; Cushman, 1986; Gould, Alfaro, Finn, Haupt & Minuto, 1987). Mills & Weldon (1987), in their review of empirical studies concerning the readability of text from computer screens acknowledge the performance difference between the two media but attribute some of the findings to the lower quality of text displayed on computer monitors due to low performance. In his review, Dillon (1992) attributes the earlier findings of differences in performance to a number of interfering factors not taken into consideration (such as luminescence in the room, size and performance of the visual display unit, etc) as well as to artificial tasks that had nothing to do with the activity of reading (such as identifying missing characters, spelling errors, missing and misfit characters in text). Both Mills & Weldon (1987) and Dillon (1992) emphasize the importance of taking all interfering factors into consideration when attempting to compare reading speed on paper and on computer screen.

Kurniawan & Zaphiris (2001) investigated the effects of text layout and reading patterns on reading performance. Their findings were consistent with earlier findings that reading from screen is slower than reading from paper. The authors noted that volunteers sometimes used their finger, a pen or a pencil to keep track of their place in the text and these aiding mechanisms might account for a certain percentage of the differences in reading performance.

Garland & Noyes (2004) reported that they found no significant difference in reading speed when reading from paper versus reading from monitors. The authors attributed the improved performance to improvements in technology and close matching in quality of printed and electronic materials. They suggested that although the data failed

to show a significant difference in reading performance when expressed as reading speed and content comprehension, reading from paper was still superior to reading from monitor as it favored the process of learning.

While there seems to be disagreement in the literature if reading from computer monitor is consistently slower than reading from paper, the studies suggest that the quality of the reading material, the tasks and the available hardware all play an influence on reading performance when reading from computer monitors.

Extraneous Factors Influencing the Cognitive Load

Text Formatting and Design

Identifying the optimal text presentation format to enhance reading performance and learning from text has been heavily researched for the last 30 years. Factors such as typography, character size, character and line distancing, line length, number of lines per page, number of characters per line, page alignment, text alignment, method of presentation, color, luminosity and contrast were all investigated and their effects on reading performance and content comprehension evaluated.

. Shieh, Chen & Chuang (1997) investigated if color combination and font type affect visual performance and fatigue when working on a visual display terminal (VDT). Participants were flashed a series of Chinese characters for 250 milliseconds and asked to write down as many characters as they could identify. The study showed a significant relationship between font typography and performance.

The effects of five different font types on reading speed were investigated by Boyarski, Neuwirth, Forlizzi & Regli (1998). While reading times and comprehension scores were not affected by font typography, expressed subjective preferences favored

anti-aliased fonts versus bitmapped fonts when reading text from computer monitor. The authors emphasized that the relatively short paragraphs used in the experiment might have contributed to the insignificant differences in performance and that longer chunks of text might have resulted in significant differences in the performance metrics.

Chan & Lee (2005) studied whether font type, character size, line spacing and display polarity affect reading speed and content comprehension when reading traditional Chinese Ming characters. User performances were calculated and subjective preferences were also collected. The study confirmed that although all the factors affected reading speed, character size also had a significant effect on content comprehension.

Early studies conducted in the 1980s aiming to investigate the effects of line length on reading performance were severely limited by the available technology, therefore only a limited spectrum of possibilities were explored, such as comparing line lengths of 40 characters versus 70 characters (Mills & Weldon, 1987). These early studies suggested a direct relationship between line length and reading performance: the longer the line, the better the performance. Dyson (2004) warned against unconditionally accepting these findings, as due to technical limitations, computer monitors used in the 1980s did not enable individually modifying variables such as character size, character width or line length (i.e. modifying one of these variables necessarily affected the rest of the variables). Dyson (2004), when analyzing these early findings, concluded that the measured differences in performance were expressing the combined effects of changes in all three variables rather than the effects of line length per se.

In a series of studies, Dyson (1997) and Dyson & Haselgrove (2000, 2001) investigated the effects of line length on reading speed and content comprehension.

While Dyson (1997) failed to show a significant relationship between line length and reading speed, Dyson & Haselgrove (2000, 2001) suggested that a medium line length of about 55 characters per line enables optimal performance both for reading speed and content comprehension. Dyson & Haselgrove also observed the existence of a reading speed-accuracy trade-off, the faster the volunteers read or skimmed the text the less able they were to answer detail questions about the text.

The effects of line length on reading speed and comprehension when displaying dynamic content on various monitor sizes were investigated by Laarni (2002). Five dynamic presentation methods (*rapid serial visual presentation [RSVP]* – one word presented at a time in the center of the screen; *vertical scrolling mode* – text continuously scrolling a raster at a time from the bottom of the page; *leading mode* – text moving from right to the left along a single line; *teletype mode* – controlled rate text where one character is added at a time to the line; and *window mode* – one line of text displayed at a time with a frame passing across the text and highlighting every word) were compared for data displayed on laptop screen, palm-type, PDAs, communicators and mobile phones. The findings suggest that a line length of about 50 – 60 characters correlate with best performance for dynamic text presentation.

The effects of text width and margin width on reaction time upon reading text from computer monitors were studied by Youngman & Scharff (1998). Participants had to scan the text for target words and then click on the corresponding icon on the bottom of the page. Results indicated that text width did not influence readability but there was a significant interaction between text width, margin width and performance. Small text

width (4-inch) and large margins, or the largest text width (8-inch) and no border correlated with best performance.

The effects of text justification on reading speed were also studied in a series of studies. Mills & Weldon (1987), in their review of empirical studies, suggest that unjustified text allows for faster reading while center or right justification negatively affects reading speed.

Dillon (1990) studied the effects of display size (20 lines a page versus 60 lines a page) and text splitting on content comprehension. The participants read a 3500-word text, particularly chosen for being unfamiliar in content for all of the subjects. Although display size did not seem to significantly affect measured content comprehension, the participants clearly favored the 60 line display, no text splitting presentation format. The author mentions that the text splitting was associated with high rates of paging back and forth, signaling an increased cognitive load.

Dyson & Kipping (1997) studied if the columnar presentation of text affects reading performance. Reading speed was measured and content comprehension calculated for two different display conditions – one page-wide column versus three columns. While the single wide column produced better reading speed, subjective preference favored the three-column layout. Interestingly, the data showed that the age-group of 18 to 24 was solely responsible for the significant difference in reading time and the authors suggested that reading patterns and better familiarity with the display format might account for the difference in performance.

Piolat, Roussey & Thunin (1997) studied reading performance and cognitive load when varying page dynamics – scroll mechanism versus paging. Paging enabled better

recall of the main ideas and details of the text than scrolling. In addition, the data suggested that providing a paged layout enabled readers to build a better *sense of text,* while the scrolling mechanism seemed to increase the cognitive load associated with content comprehension.

Lam, Lam, Liu & Shin (2000) investigated the same hypothesis that the columnar display of a text affects reading speed and content comprehension. Their results indicated that reading the one- column text was about 28% slower than reading the three-column text. The study found no significant relationship between content comprehension and reading speed.

Kurniawan & Zaphiris (2001) investigated reading patterns and reading speed on paper and on computer display while changing the columnar format of the display. Using reading material and comprehension questions from a standardized test they measured reading speed and content comprehension in one-, two-, and three-column formats. The study failed to find any significant difference in reading performance in the various columnar displays but suggested that the higher the number of columns, the less significant the difference between reading performance from screen and from paper.

Laarni et al. (2004) measured reading speed when text was presented vertically versus text presented horizontally. They found that user familiarity with the display format affected reading speed (as in Western languages text is traditionally presented vertically participants read horizontally displayed text faster than vertically displayed text).

McCrudden et al. (2004) studied the cumulative effects of text presentation, organization and example context on reading performance. The study's findings provided

support for cognitive load theory according to which the effects of extraneous variables are additive in nature. The text presentation format (blocks of text versus presenting text one line at a time) had the strongest effect on content comprehension and the perceived degree of difficulty of the text. The ability to see blocks of text at a time enabled participants to backtrack in reading if needed, while reading one line at the time required readers to rely on their short term memory when trying to process the content, increasing the cognitive load.

CRT Display Characteristics

The effects of the computer hardware, especially of the computer monitor characteristics on reading speed and content comprehension was investigated in a series of studies, in an effort to optimize reading performance. Ziefle (1998) investigated if manipulating display resolution affects proofreading speed and accuracy. While her findings failed to show any significant relationship between monitor resolution and reading speed, resolution and task performance, lower resolution was significantly tied to visual discomfort and low performance when trying to locate targets on a screen. Ziefle's findings were in line with previous findings that reading from paper is faster than reading from a computer monitor.

The effects of monitor size on reading speed and comprehension was investigated by Laarni (2002). Five dynamic presentation methods (RSVP, vertical scrolling mode, leading mode, teletype mode and window mode) were compared for data displayed on monitors with various resolutions (laptop screen, palm-type, PDAs, communicators and mobile phones). The findings suggest that different presentational methods are optimal for different monitor sizes.

Sheedy, Subbaram & Hayes (2003) investigated if monitor luminosity, contrast and disturbing reflections on a computer screen affect reading speed, reading performance and user comfort. Computer display characteristics were manipulated using various filters applied to the front of monitors. Applying filters had a significant effect on legibility and character legibility was shown to clearly affect reading speed.

Garland & Noyes (2004) compared reading speed and content comprehension when reading from paper versus reading from monitor. They concluded that while there were no significant differences between the two media format, monitor refresh rates, fluctuating luminance and contrast level influenced the process of learning when reading. The authors suggested that CRT monitor characteristics affect human cognitive processes and increase the cognitive load on readers.

Buchner & Baumgartner (2007) evaluated proofreading performance when changing polarity and color contrast. User physiological effort measures (breathing rate, heart rate, heart rate variability, skin conductance level) and physical and emotional discomfort (muscle strain, back pain, fatigue, eyestrain, self-reported mood) were assessed. Participants had to read 15 separate short stories, each story of about 875 words long. User performance was shown to be affected by display polarity, dark text on light background being tied to superior performance. User effort and discomfort was not influenced by the dependent variables.

Intrinsic Factors Influencing the Cognitive Load: Text Quality and Informational Content

The effects of the intrinsic factors such as text quality and informational content on reading performance was evaluated together with the effects of the extraneous factors

in some of the aforementioned studies. Thornton, Maryellen & Arnold (2000) studied the effects of phrase length on modification ambiguities. They devised an experiment where they attached prenominal adjectives to the noun phrase in a sentence and participants had to determine if the resulting phrase conveyed the same meaning as the original phrase. The words were presented sequentially, only one word being visible at the time. The study confirmed the hypothesis that phrase length influences content comprehension.

Braten & Stromso (2006) investigated if text understanding was influenced by presentation format, implicit text quality, and informational content. One group of students was presented a series of partially conflicting texts about AIDS while the other group read a textbook-like article on the topic. The study suggested that comprehension was mediated by personal epistemology- students holding naïve epistemological beliefs performed better after reading the textbook-like material, while students with sophisticated epistemological beliefs gained a better understanding of the topic after reading the texts with partially conflicting informational content.

Jason & Just (2007) studied the effects of lexical ambiguities on reading performance while taking MRI brain scans. The MRI brain scans showed increased brain activity when encountering ambiguous terminology. The authors suggested that the increased activity signaled a complicated process of sense disambiguation and inadvertedly the associated increased cognitive load: when multiple meanings were attributed to a word, the brain had to semantically analyze all the available variants; it had to evaluate the variants in the given context; and upon settling for the best alternative, the brain had to actively suppress all of the incorrect interpretations. The authors suggest that this lengthy process, involving intensive cognitive resources, is executed for every

instance of an ambiguous term in a text. The negative effects of ambiguity can be further accentuated by prior biasing context, the brain first evaluating the most recently used candidate meaning.

Personal Characteristics Influencing the Effects of Intrinsic and Extraneous Factors

The effects of the intrinsic and extraneous factors enumerated in the previous sections are mediated by the personal characteristics of readers, both physical and psychological. First of all, people with visual or physical disabilities are more affected by factors such as text layout, typography, luminosity, navigation options, than the rest of the population. Secondly, personal differences in reading rate, reading skill, analytic skills, attention spam, verbal skills and prior knowledge influence reading performance and the cognitive load.

Gillström & Rönnberg's (1995) findings when examining the reading performance and content comprehension of 111 Swedish high school students suggested that verbal skill levels and reading skill are factors influencing both reading speed and content comprehension. The authors suggested that high-level-verbal-skill students have developed reading skills enabling them to minimize the cognitive effort associated with the reading task, so they can use most of their resources for content comprehension. They also suggest that the comprehension process itself when reading, is different for high-verbal-skill students. These students form a mental understanding (*situation model* or *situational awareness* in cognitive psychology) of the material and integrate the content into their prior knowledge. In contrast, medium and low-verbal-skill students rely on memory and recall when processing the content of a text. While the situational model

enables high-level-verbal-skill students to quickly summarize the main ideas and to easily draw valid inferences on their own, medium and low-verbal-skill students find the same tasks more difficult to accomplish.

Dyson & Haselgrove (2000, 2001) conducted a series of experiments investigating if reading speed and scrolling patterns could be used as predictors of content comprehension. Participants' reading speed was assessed in a pre-test, and during training sessions they were asked to try to read twice as fast. Only people who were able to improve their reading performance at least 70% were allowed to continue the experiment. Their findings were that content comprehension was significantly better at normal reading speed than at artificially accelerated reading speed, suggesting a speed – accuracy trade off. Comprehension scores for the different types of questions (title, main idea, structure, main factual, incidental) were negatively affected by the increased reading speed.

Braten & Stromso (2006) investigated if prior knowledge and presentation format would affect content comprehension. Participants were classified as holding either naïve or sophisticated epistemological beliefs based on the Norwegian version of the Schommer Epistemological Questionnaire (SEQ), administered prior to the experiment, and then were divided into two groups. The groups were then divided again and sub-groups were presented with study material in two different formats, multiple texts versus textbook chapter. The experiment showed that there was a significant relationship between personal epistemological beliefs and content comprehension and between epistemological beliefs and the ability to deal with complex, partially contradicting pieces of information. The study suggested that while having prior knowledge in the field

enhanced content comprehension, having sophisticated epistemological beliefs was a better predictor of content comprehension than prior knowledge. Individuals who considered knowledge a personal, flexible construct based on available information, scored better on comprehension scores, even in the absence of prior knowledge, than people holding naïve epistemological beliefs who had prior knowledge in the field. Jason & Just (2007) performed MRI scans of the brain while people were reading text containing lexical ambiguities. The MRI brain scans showed increased brain activity in both brain hemispheres when encountering ambiguous terms. The lexical ambiguities increased the cognitive load and the time required for information processing. Mason & Just's experiment showed that people with short attention and limited short-term memory span were affected to a greater degree by lexical ambiguities than people with better attention and better working memory, providing support to the cognitive load theory.

Forming a complex structural and semantic mental map of the reading material and recoding it to integrate with previous knowledge, has been tied to improved comprehension (Gillstrom & Jerker, 1995; Piolat et al., 1997; Stone et al., 1999; McElree, 2000, DeStefano & LeFevre 2005; Braten & Stromso, 2006). Text quality (Thornton et al., 2000; Braten & Stromso, 2006), presentational format (Piolat et al., 1997) and short term memory span influence this process of internal mapping and developing situation models (DeStefano & LeFevre, 2007).

Reading skill and verbal skill levels affect content comprehension and the quality of knowledge that readers are able to gain (Gillstrom & Jerker, 1995; Piolat et al., 1997; Stone, Fisher & Eliot, 1999; Dyson & Haselgrove, 2001; Laarni et al., 2004). People with lower verbal skills employ more cognitive resources (DeStefano & LeFevre, 2007) and

show more cerebral activation during a listening span task than people with better skills (Mason & Just, 2007).

Summary

Overall, the literature suggests that reading speed and content comprehension are significantly affected by various intrinsic and extraneous factors influencing the cognitive load associated with reading a text. Extraneous factors such as text and phrase length, typography, monitor resolution and luminosity, text layout and presentation format all influence the experienced cognitive load. According to the cognitive load theory, the effect of the extraneous factors is additive and increasing the cognitive load will negatively affect reading performance (Gillström & Rönnberg, 1995; Garland & Noyes, 2004; McCruden et al., 2004; Braten & Stromso, 2006; DeStefano & LeFevre, 2007).

Text quality and informational content are intrinsic factors influencing the experienced cognitive load. The presence of lexical ambiguities in the text increases cognitive load and it deters reading speed performance and the comprehension process (Mason & Just, 2007; Xu & Huang, 2007).

The effects of the intrinsic and extraneous factors on the cognitive load are mediated by personal characteristics, first of all by working memory capacity. Decreasing the performance of the short term memory will negatively affect reading speed (Miller, 1956; Baddeley, 1994; Schiffrin & Nosofsky, 1994; Fortin & Rousseau, 1997; Cowan, 2000; McElree, 2000; Baddeley, 2003; McCruden et al., 2004; DeStefano & LeFevre, 2007). Prior knowledge, familiarity with the topic, with the terminology or even the presentation format enhance the processes of information recoding and structure model building (Miller, 1956; Dillon, 1990; Baddeley, 1994; Dyson & Kipping, 1997; Stone et

al., 1999; McElree, 2000; Laarni et al., 2004; McCruden et al., 2004; Braten & Stromso, 2006; DeStefano & LeFevre, 2007).

Short texts are not adequate to measure the effect played by external factors on reading speed or content comprehension, as the measured differences in performance are not large enough to demonstrate significance (Dillon, 1990; Boyarski et al., 1998; DeStefano & LeFevre, 2007).

The recall of different types of information is influenced by reading speed and text quality so content comprehension questions need to target the different types of information (Dyson & Haselgrove, 2000, 2001; Braten & Stromso, 2006).

Research Questions

The literature suggests that heavy use of unfamiliar acronyms in a text can increase the degree of ambiguity of the text and therefore increase the cognitive load. When encountering unknown acronyms readers have to employ extra cognitive resources to process the information: they might have to scroll in the text to find acronym definitions, and they have to hold acronyms and their associated definitions in their working memory. In addition, they have to mentally substitute ambiguous acronyms in text with their long forms when processing the information (Mason & Just, 2007). A tool that would help people interpret or disambiguate the acronyms may decrease the cognitive load and increase ease of reading, comprehending. We devised an experiment to investigate if the increased cognitive load resulting from the presence of a large number of unknown acronyms in a text would indeed negatively influence reading speed, content comprehension and user experience.

When designing the experiment we took into account several factors that could contribute to increased cognitive load. The presence of unknown acronyms in the text may make it difficult for readers to follow the text and develop appropriate situation models, and it may demand that readers rely heavily on their working memory for acronym sense disambiguation.

A multitude of unknown acronyms may surpass the capacity of the short term memory. Readers most likely will be unable to remember so many acronyms and the associated expansions, forcing them to constantly scroll for acronym definitions.

The large expansions associated with the acronyms will contribute to the increased cognitive load. Most of the acronyms in the text used in the experiment have expansions of three words or more and there are several sentences in the text containing more than one acronym. As word and sentence length affect content comprehension, we expected that readers would have considerable difficulties following sentences containing multiple acronyms with lengthy expansions.

Readers' unfamiliarity with the topic will make the process of recoding the information and integrate it with prior knowledge more difficult. Cognitive load on working memory decreases when people quickly integrate new information with prior knowledge as there is no need to keep the newly conveyed information in short term memory; prior knowledge from the long term memory can be recalled when needed. We expected that most of the participants did not have prior knowledge about healthcare initiatives and programs in Ethiopia and Africa and therefore would have to employ additional cognitive resources for keeping the conveyed information in the working memory. We expected that they would find it hard to follow the text and to build adequate situation models that would enhance the comprehension process. In the absence of an adequate situation model, we expected that participants would have to rely on memory and recognition when answering the comprehension questions, resulting in increased cognitive load and poor performance.

In this experiment we also wanted to study the potential benefits of using a tool for instant acronym sense disambiguation. The current practice of only defining acronyms once, when first introduced, potentially enables the passage of a large amount of time and text between seeing the acronym expansion and seeing the unfamiliar acronym again. This increases the likelihood that the reader will have forgotten what the acronym stands for, and he or she will have to invest additional cognitive resources in an effort to disambiguate the acronym. He or she might try to guess its meaning from the context, or might have to stop reading and try to find the acronym's expansion in the text, store it the working memory and then find his, her place in the text again. He or she might skip the problematic phrase and try to understand the meaning of the text from the rest of the passage. Any of these strategies will make the text harder to follow and will increase the cognitive load. Given the large number of acronyms in our text we expected that participants would find extremely problematic to remember acronym expansions, especially when defined early in the text but used significantly later.

To investigate if reducing the participants' efforts to disambiguate the acronyms would decrease the cognitive load, we provided the experimental group with an extra tool for accessing acronym expansions when needed. Participants in the experimental group were able to mouse over any instance of the acronym to see its associated definition as hover text (for a screenshot of an acronym and its associated definition displayed as hover, please see Figure F1). By eliminating the factor of time passed between seeing the acronym and its definition, this tool could potentially eliminate the need to store acronyms and acronym definitions in the short term memory when reading the text, and could eliminate the need for scrolling up and down in the text. The expected outcome of

using the tool would be decreased cognitive load and/or increased reading speed and content comprehension. In addition, being able to see the acronym and its long form at every occurrence might even enhance the process of learning new acronyms and the associated definitions, especially for acronyms that are frequently mentioned in the text.

Considering the expected effects on cognitive load of a large number of unknown acronyms and the expected benefits of having a tool for instant acronym disambiguation, we formulated the following null hypotheses.

H1: The presence of a multitude of unknown acronyms in the text will not adversely affect content comprehension. This hypothesis would be supported if participants rate the text as “easy to understand” or “somewhat easy to understand” and rate the acronyms in the text as “easy to handle”. This hypothesis would be rejected if participants rate the text as “very hard to understand”, identify the acronyms as one of the factors to make it difficult by stating that the acronyms in the text were “somewhat problematic” or “very problematic”, and/or say that answering the questions would have been easier if there were fewer acronyms in the text.

H2: The availability of the hover functionality will not help content comprehension. This hypothesis would be supported if there were no significant differences between the average comprehension scores attained by the two groups. This hypothesis would be at least partially supported if there were no significant differences between average ratings of text difficulty by the two groups. The hypothesis would be rejected if participants in the experimental group scored significantly better in content comprehension questions or if they scored at least as well as participants in the control group but were significantly faster at answering the questions.

H3: The availability of the hover functionality will not enable faster reading. This hypothesis would be supported if the average reading time of participants in the control group were the same or better than the average reading time of the experimental group. It would be rejected if participants in the experimental group read the text significantly faster than participants in the control group.

H4: The availability of the hover functionality will not enhance answering the comprehensive questions. This hypothesis would be supported if participants in the control group scored at least as well when answering content comprehension questions as participants in the experimental group, and the overall average time taken to answer questions for the two groups did not differ significantly.

Methodology

The Experiment

We designed a post-test only, between-groups experiment. Subjects were randomly assigned to the control group or the experimental group. Individual experimental sessions started by having the participants sign an informed consent form, then completing a demographic questionnaire. Then subjects read a long text filled with acronyms. Acronym expansions in the text were provided only once, when the acronym was first introduced, conforming to normal practice as well as the APA style manual[5]. Subjects in both groups read the same text, but subjects in the experimental group had the added ability to mouse over acronyms in the text and see the long form as a hover text. Throughout this paper we will refer to this added functionality of moving the mouse pointer over an acronym to see its definition as the *hover functionality* or simply *hover*. (For a screenshot of the user interface with the hover functionality please see Figure F1.) After reading the text, subjects answered SAT-style comprehension questions that focused on acronym-related topics. Participants were able to view the text when answering the comprehension questions. (For a screenshot of the user interface when answering the comprehension questions please see Figure F2) Participants in the experimental group could use the hover functionality in the text section throughout the experiment. The hover functionality was not available for the acronyms in the comprehension questions themselves. The experimental sessions ended with a debriefing

questionnaire and a demonstration of the hover functionality for the participants in the control group. Participant comments after the session were noted on the debriefing questionnaire.

This research was approved by the University of North Carolina UNC at Chapel Hill (UNC-CH) Institutional Review Board (IRB).

Materials

The demographic information questionnaire (see Appendix A) contained 11 questions and gathered general information about the participants, including age, gender, undergraduate degree, highest degree completed and graduate program enrolled in or completed. In addition, the questionnaire collected information about the subjects' attitudes towards acronyms and their favorite ways of coping with unknown acronyms in text. Finally participants were asked to rate their own reading speed. The questionnaire was used to see if the control and experimental groups had similar demographic characteristics, mechanisms for coping with unknown acronyms, and attitudes towards acronym usage in text.

Following the demographic questionnaire, participants were given group-specific instructions about the experiment (see Appendix B). Participants in the control group were instructed to read the text, bearing in mind that they would have to answer 10 comprehension questions and that each question would have only one correct answer. Participants in the experimental group received the same instructions, but in addition were told about the availability of the hover functionality and were encouraged to use this functionality as needed. Participants were not told in advance that the reading material would be available when answering the comprehension questions as we wanted to be able

to measure reading performance in the two settings (with and without the hover functionality).

All participants read the same text containing a multitude of acronyms (see Appendix C). We compiled the text for reading from passages of the published final evaluation report on Africare's 'The Gambella Child Survival Project', available on the web at http://www.oecd.org/dataoecd/0/3/35804945.pdf. We chose this article for two reasons. The first reason was to ensure that subjects would have little, if any knowledge about the topic and the organizations mentioned in the article, so all participants would find it equally unfamiliar. The second reason was that the article demonstrated several of the issues frequently mentioned in the literature when talking about problems created by acronym usage in text. First, the large number of acronyms could present obstacles for the comprehension process, especially for people not familiar with the topic and with the acronyms -- the published report is 29 pages long and its acronym list contains 66 entries. The article also illustrates the problems created by acronym polynymy: in the acronym list accompanying the original report WHO is defined as an acronym for both the World Health Organization and Wareda Health Office. In the absence of an associated definition in the sentence or in the paragraph, it is really difficult to tell which entity the acronym WHO refers to. The text also contains an undefined acronym –the acronym CORE is not defined in the text, nor in the acronym list accompanying the document.

The passages were taken from various sections (Summary, Assessment of Results and Impact of the Program, Cross-cutting Approaches) of the 29-page study, making sure that the resulting text formed a coherent extract without adding to the inherent difficulty of the text. The wording of some of the passages was modified for the purposes of this

study, and charts containing report data and results summary evaluations were omitted, as we did not want the participants to try to memorize indicator numbers and percentages in anticipation of a comprehension question. In the original paper, the authors used short forms and long forms of the acronyms intermittently. We modified the text to be conform with the recommended American Psychological Association (APA) style of defining acronyms only once, when first introduced and using only the acronyms afterwards[6]. In order to reproduce the common phenomenon of having multiple acronyms in the same sentence we also defined a few new acronyms for existing names in the text. The resulting text was about 2500 words long and it contained 31 acronyms.

After reading the text, all participants answered the same 10 comprehension questions (see Appendix D). The questions measured the different types of knowledge as defined by Dyson & Haselgrove (2000, 2001). All 10 questions asked for information that appeared in the reading material so in the absence of other factors affecting content comprehension, participants who indeed read the text would have an advantage over participants who just skimmed the text or simply gave up reading and clicked the "Show questions" button.

The first question involved identifying the main idea of the text, measuring the ability of the participants to grasp the gist of the reading; their ability to build the appropriate situation model. People who did not read the text would have problems answering this question.

Questions 2 -10 were detail questions meant to explore the degree of understanding of participants. Two questions asked for identifying relationships between acronyms, probably requiring the deepest level of understanding, as participants had to

remember not only the meaning of the acronyms but also their scope and be able to identify the acronyms whose scope allowed the stated relationship. Four questions asked for the meaning/range of acronyms that were used with various frequencies in the text, trying to discover if frequent usage enhanced the process of learning of the acronyms.

Questions 2 and 9 were the same question presented in different format – question 2 used acronyms in the answer choices and question 9 used the long forms. By spelling out the options we offered the possibility to instantly recognize the difference in entity type of the various choices and we were curious to see if making this difference visible would indeed enable prompt responses from the participants. Question 10 asked for the definition of one of the most frequently used acronyms in the text and it was the question where the efficacy of the hover function might make the most difference.

The debriefing questionnaire (Appendix E) was administered after participants completed the comprehension questions, and asked participants to rate how difficult they found the text and the comprehension questions, and if they found the presence of the numerous acronyms an obstacle when reading the text and answering the questions. The questionnaire also asked about the coping strategies they used to answer the questions in an attempt to determine the degree to which readers were able to master the acronyms and thus make an informed choice and the degree they relied on guesswork for their answers. It also sought to determine if participating in this experiment had influenced the participants' attitudes towards acronym usage. Subjects in the experimental group were also asked to evaluate the hover functionality.

Subjects

Graduate students from the School of Information and Library Science (SILS) at UNC-CH were recruited via mass emailing and posted flyers. We recruited graduate students from UNC-CH SILS in an effort to ensure that the participant pool would be likely to find a maximum number of unknown acronyms in the text and would be equally unfamiliar with the topic of our reading material.

Participation in the experiment was voluntary. The participants were not paid but those who completed the experiment were entered into a drawing for three prizes (two $30 iTunes gift cards and one $40 gift certificate for the University Student Stores).

Location and Equipment

Experimental sessions were held in various rooms in Manning Hall, home of SILS, UNC Chapel Hill. Most of the sessions were held in the SILS Computer Classroom during summer school sessions. With the start of the fall semester it became increasingly difficult to schedule experimental sessions in the computer lab due to academic courses being held there so the last 10 sessions were held in a small seminar room instead.

Windows workstations in the Computer lab with a monitor resolution of 1024 X 768 and a refresh rate of 60Hz were used for the experiment. Efforts were made to minimize the effects of hardware, software or outside performance disturbing factors – individual sessions were held when there were no classes or any other kind of activities scheduled in the lab. For the sessions held in the seminar room, IBM laptop computers were used.

The reading material was stored on a Linux RedHat server and was accessed with a web browser (Internet Explorer 7.0). The text was displayed with the browser's default

font type and font size. The width of the browser window was controlled with JavaScript and was set to 1000 pixels.

Data Collected

Individual reading times, the answers to the comprehension questions and individual answering times for each of the content comprehension questions were automatically recorded in a MySQL database with a PHP script running on the Linux RedHat server. The time was incremented in seconds. The individual responses from the demographic and debriefing sections were encoded by the principal investigator and added to the database. The collected data was analyzed using SPSS 15.0 for Windows.

Results

For the Results and the Discussion sections the following notation conventions will apply: *M* denotes statistical mean, *SD* denotes standard deviation, *r* denotes the Pearson correlation and *p* denotes the probability value.

Demographic Survey Results

Twenty six graduate students (18 females and 8 males) from SILS UNC-CH were recruited for this study. All participants were 18 years or older, the oldest participant being over 61 years old. Nineteen participants (73.1%) identified the BA/BS as the highest degree completed, 6 participants (23.1%) had a MA/MS and 1 participant (3.8%) had a PHD/PostDoc. Fifteen participants (57.7%) were enrolled in the Master of Science in Library Science (MSLS) graduate program and 11 participants (42.3%) were enrolled in the Master of Science in Information Science (MSIS) graduate program.

Seven participants (26.9%) reported sometimes encountering unknown acronyms in their readings. Fifteen participants (57.7%) reported frequently encountering unknown acronyms in their readings and 4 participants (15.4%) said that they almost always encounter unknown acronyms in their readings.

Only two participants (7.7%) reported that they like to construct their own acronyms when writing. Five participants (19.2%) regularly use acronyms but they mostly restrict their acronym usage to already established acronyms in the files. Eleven

participants (42.3%) reported keeping their acronym usage moderate and eight participants (30.8%) declared that they tend to avoid acronyms in their writings.

Fourteen participants (53.8%) favored scrolling or paging to find a definition as first strategy for coping with unknown acronyms in their readings. Ten participants (38.5%) would first try to guess from the context and two participants (7.7%) would try to look up the acronyms online.

As their second strategy to deal with unknown acronyms, ten participants (38.5%) opted for guessing from context, six participants (23.1%) would scroll or page to find a definition, five participants (19.2%) would look up the definition online and two participants (7.7%) would skip the acronyms and keep reading.

Skipping the acronym and keeping reading was favored as the third strategy for dealing with unknown acronyms in the text by 13 participants (61.9%). Other results for the third strategy were that three participants (14.3%) would try to look up the acronyms online, three participants (14.3%) would scroll or page to find a definition and two participants (9.5%) would try to guess from the context.

Participants were randomly divided into experimental and control groups. There were no differences in the demographic characteristics (age, gender, educational level, graduate program enrolled in, expected graduation year) between the two groups. The resulting groups were also similar in other reported metrics as well (frequency of encountering unknown acronyms in readings and self-assessed reading speed).

There were slight differences in the personal usage of acronyms (see Table 1), as both participants who preferred to construct and use their own acronyms when writing were assigned to the control group while six participants out of the eight who tend to

avoid using acronyms in their writings were assigned to the experimental group. These differences were not statistically significant.

Table 1

Acronym Usage among Group Members

Group	Acronym usage			
	I like to construct my own acronyms and use them in my writings	I regularly use acronyms but I mostly use already established acronyms in the field	I mostly use established acronyms in the field but I keep my acronym usage moderate	I tend to avoid using acronyms in my writings
Control Group	2	1	4	6
Exp. Group	0	4	7	2

There were differences between groups in the coping strategies when encountering unknown acronyms in reading, Most (nine out of thirteen) participants in the experimental group favored scrolling or paging to find a definition as first strategy while participants in the control group seemed to favor guessing the definition from the context (seven participants out of thirteen). When taking into consideration the first two coping strategies together the two groups were similar (see Table 2).

Table 2

Aggregate Data – First Two Choices for Coping with Unknown Acronyms in Text

Group	When encountering an unknown acronym in the text I would			
	Scroll or page to find a definition	Try to guess from the context	Look it up online	Skip the acronym and keep reading
Control Group	10	11	4	0
Exp. Group	10	9	4	2

Reading Performance and Comprehension Questions Results

Mean reading time for participants in the control group was M = 833, SD = 280.61 seconds. Mean reading time for the experimental group was M = 844.31, SD = 367.41 seconds. The difference in reading times was not statistically significant.

Mean answering time to answer the comprehension questions was M = 48.36, SD = 12.36 seconds for the control group and M = 40.76, SD = 20.45 seconds for the experimental group. The difference in answering times was not statistically significant.

Participants in the control group provided M = 80.77%, SD = 20.19 correct answers to the content comprehension questions. The answers to the comprehension questions for the experimental group were better, M = 90.00%, SD = 10.00, but the difference was not statistically significant.

There were no significant differences in the average answering times to the individual comprehension questions between the two groups.

Debriefing Questionnaire Results

Thirteen participants (50%) rated the text as "very hard to understand", ten participants (38.5%) rated the text as "not too hard to understand", two participants (7.7%) rated the text as "somewhat easy to understand" and one participant (3.8%) rated the text as "easy to understand". The two groups rated the difficulty degree of the text similarly (see Table 3).

Table 3

Ratings of Degree of Difficulty of the Text by Groups

Group	The text was			
	Easy to understand	Somewhat easy to understand	Not too hard to understand	Very hard to understand
Control Group	0	1	6	6
Exp. Group	1	1	4	7

Thirteen participants (50%) rated the acronyms in the text as "very problematic", eleven participants (42.3%) rated the acronyms as "somewhat problematic" and two participants (7.7%) rated them as "easy to handle". When broken down by group there was a slight difference, as both participants who rated the acronyms as being easy to handle were in the experimental group (see Table 4).

Table 4

Ratings of Degree of Difficulty of Acronyms by Groups

Group	The acronyms were		
	Easy to handle	Somewhat problematic	Very problematic
Control Group	0	5	8
Exp. Group	2	6	5

Twenty-three participants (95.8% of responses) considered that the acronyms were problematic because there were too many acronyms. One participant (4.2%) considered that the acronyms were too complex. Two participants did not respond to this question as they considered the acronyms to be easy to handle. Breaking down the responses by group, all participants in the control group considered the acronyms problematic because of their sheer number. Two participants (15.4%) in the experimental group did not consider the acronyms problematic; one participant (7.6%) considered the acronyms problematic for being too complex and ten participants (76.9%) considered that problems were caused by having too many acronyms in the text.

Sixteen participants (61.5%) considered that the comprehension questions were easy to understand, seven participants (26.9%) considered that the comprehension questions were a little confusing, one participant (3.8%) considered that they were somewhat confusing and two participants (7.7%) rated the questions as very confusing. The two groups differed somewhat in their ratings of the comprehension questions (see Table 5), but the difference was not statistically significant.

Table 5

Ratings of the Comprehension Questions by Groups

Group	The comprehension questions were			
	Easy to understand	A little confusing	Somewhat confusing	Very confusing
Control Group	9	2	0	2
Exp. Group	7	5	1	0

The debriefing questionnaire asked participants to estimate the number of instances when they knew the correct answer for the comprehension questions, the number of instances when they had to look into the text for the correct answer and the number of instances when they guessed the answer. Based on the self-reported numbers there were 77 instances where participants knew the correct answer (29.6%), 151 instances of participants looking up answers in the text (57.8%) and 33 answers (12.6%) based on guessing (for the individual responses see Table G1). The two groups were quite similar along this parameter as well (see Table 6).

Table 6

Strategies for Answering the Comprehensive Questions by Groups

Group	When answering the comprehension questions I		
	Knew the answer	Looked up answer	Guessed the answer
Control Group	41 (31.3%)	74 (56.4%)	16 (12.2%)
Exp. Group	36 (27.7%)	77 (59.2%)	17 (13.0%)

There was a significant negative correlation between using the hover functionality when reading the text and guessing the answer ($r = -.691$, $p = 0.01$). There was also a significant negative correlation between using the hover function when answering the questions and guessing the answer ($r = -.619$, $p = 0.05$).

The debriefing questionnaire also asked the participants to approximate the number of cases when they knew the meaning of the acronyms in the questions, when they had to look them up in the text and when they guessed. According to the self-reported numbers there were 53 instances of participants knowing the acronyms in the text (20.38%), 180 instances of participants looking up definitions in the text (69.23%), and 27 instances (10.38%) of participants guessing what the acronyms stood for (see Table G2 for the individual responses). There were no significant differences in the self-reported strategies between the two groups (see Table 7).

Table 7

Strategies for Coping with Acronyms in Comprehension Questions by Group

Group	When answering the comprehension questions I		
	Knew the acronym	Looked up the acronym	Guessed the acronym
Control Group	24 (18.4%)	93 (71.5%)	13 (10.0%)
Exp. Group	29 (22.3%)	87 (66.9%)	14 (10.7%)

There was a significant negative correlation between using the hover functionality when reading the text and guessing the acronym ($r = -.826$, $p=0.01$). There was also a significant negative correlation between using the hover functionality when answering the questions and guessing the meaning of the acronym ($r=-.627$, $p=0.05$).

Twenty-two participants (84.6%) judged that the comprehension questions would have been easier to answer if there were fewer acronyms in the text. Only four participants (15.4%) judged that the task would not have been easier to handle with fewer acronyms. One of these four participants rated the text as “Somewhat easy to understand”; two rated the text as “Not too hard to understand” and one of them judged that the text was “Very hard to understand”. Broken down by group, twelve participants (92.3%) from the control group considered that fewer acronyms in the text would have made the comprehension questions easier to answer versus ten participants (76.9%) in the experimental group. One participant in the control group and three participants in the experimental group considered that answering the questions would not have been easier with fewer acronyms.

Nine participants out of thirteen (69.2%) in the experimental group reported frequently using the hover functionality when reading the text. One participant reported sometimes using the hover functionality when reading and two participants did not use the hover functionality. The two participants who did not use hover when reading, rated the acronyms in the text as being very problematic.

Seven participants out of thirteen (53%) in the experimental group reported frequently using the hover functionality when answering the comprehension question, one participant used the hover functionality about half the time, one participant used the hover sometimes (1 – 4 questions) and four participants (30%) did not use the hover when answering the questions. Two of the participants who did not use hover when answering the questions reported not using the hover when reading the text either. The four participants who did not use the hover functionality for answering the comprehension questions all rated the acronyms in the text as very problematic.

Eleven participants out of thirteen (84.6%) considered that the hover functionality helped them understand the text and answer the questions. Two participants considered that the hover functionality did not help – the same two participants who reported not using the hover functionality at all. One of the two reported that he forgot about the hover functionality. The other participant reported that he did not use the hover and that he found the text so frustrating that he decided to simply skim the text and skip over segments of text containing lots of acronyms. Both of these participants reported that they found the acronyms very frustrating.

All twenty six participants said that research on acronyms is worth doing. Eleven participants mentioned the hover functionality in their individual comments about the experience as a useful feature that is worth implementing.

Three participants out of thirteen in the experimental group said that it is most likely that implementing a new standard requiring to make acronym definitions available as hover in electronic text would serve the readers' best interests. The remaining ten participants considered that such standard would definitely serve the readers' best interests.

Discussion

Data analysis showed that two participants in the experimental group reported that they did not use the hover functionality at all throughout the experiment. As the only difference between the control group and the experimental group was the availability of the hover functionality, and because one of the purposes of the experiment was to investigate if the availability of the hover functionality helped reading performance and content comprehension, we decided to do an additional analysis in which we included the performance of these two participants into the control group. Unfortunately this resulted in the situation of having 15 measurements for the control group and only 11 for the experimental group. In the following sections we will provide the results obtained after this adjustment with the phrase *after the adjustments* or *after adjustments*.

Our first hypothesis was that the presence of a multitude of unknown acronyms in the text would not adversely affect content comprehension. Most of the participants (88.5%) rated the text as very hard to understand (13 participants) and not too hard to understand (10 participants). Two participants rated the text as somewhat easy to understand and one participant rated the text as easy to understand. Thirteen participants (50%) rated the acronyms in the text as very problematic, eleven participants considered that the acronyms were somewhat problematic and two participants considered that the acronyms were easy to handle.

Twenty-two participants (84.6%) judged that answering the comprehension questions would have been easier with fewer acronyms and only four participants (15.4%) judged that having fewer acronyms would not had made a difference. Out of these four one participant rated the text as very hard to understand.

All participants agreed that research on acronyms is worth doing, thereby acknowledging the problems that acronym usage created for the content comprehension process. In their individual comments in the debriefing questionnaire 24 participants (92.3%) touched on the theme of too many acronyms negatively influencing content comprehension. Participants commented that the presence of so many acronyms in the text made the text "very confusing", it "impeded the reading speed" and the "fact retention", it clouded "reading comprehension" and it made the article "the most painful I have ever read". Participants mentioned that having the ability "to take notes" or having available a "table of acronym definitions would have helped" to "understand the article better". One participant commented that "the author should learn to write!" while another one wrote that the "overuse of acronyms seems to decrease the readability of text". We had one participant who reported being so frustrated by the large number of unknown acronyms in the text that she had to stop reading and calm herself down in order to be able to complete the experiment.

Ten participants in the experimental group considered that implementing a new standard of making available acronym definitions as hover text in electronic environment will "definitely" serve readers' best interests. The other three participants considered that it is "most likely" that such standard will serve the readers' best interests. All members of the control group who were shown the hover functionality after the experiment declared

that the hover feature would have been extremely useful in negotiating this text and the comprehension questions.

The unanimous agreement that research on acronyms is worth doing, combined with the participant feed-back to the experiment, shows that the presence of a large number of unknown acronyms in text does indeed affect reading performance and content comprehension. In the light of all these factors we can reject the null hypothesis that the presence of a multitude of unknown acronyms did not adversely affect content comprehension.

Our second hypothesis was that the availability of the hover functionality will not help content comprehension. Participants in the experimental group scored higher in the comprehension questions than participants in the control group (M = 90.00%, SD = 10.00, versus M = 80.77%, SD =20.19). After adjusting for the two participants in the experimental group who did not use the hover functionality, the performance metrics changed: experimental group M = 92.73%, SD = .09 correct versus M = 80.00%, SD = 18.89 for the control group. The average answering time for the experimental group was also better than the average answering time for the control group (M = 40.76, SD = 20.45 seconds versus M = 48.36, SD = 12.36 seconds). After adjustments the difference in answering times decreased: M = 46.44, SD = 12.63 seconds for the control group versus M = 42.00, SD = 22.06 seconds for the experimental group, again not significant.

Better scores on comprehension questions paired with better answering times would suggest that the added functionality of the hover functionality helped the experimental group, but the difference in performance was not statistically significant.

Based on our data we cannot reject the hypothesis that the availability of the hover functionality did not help content comprehension.

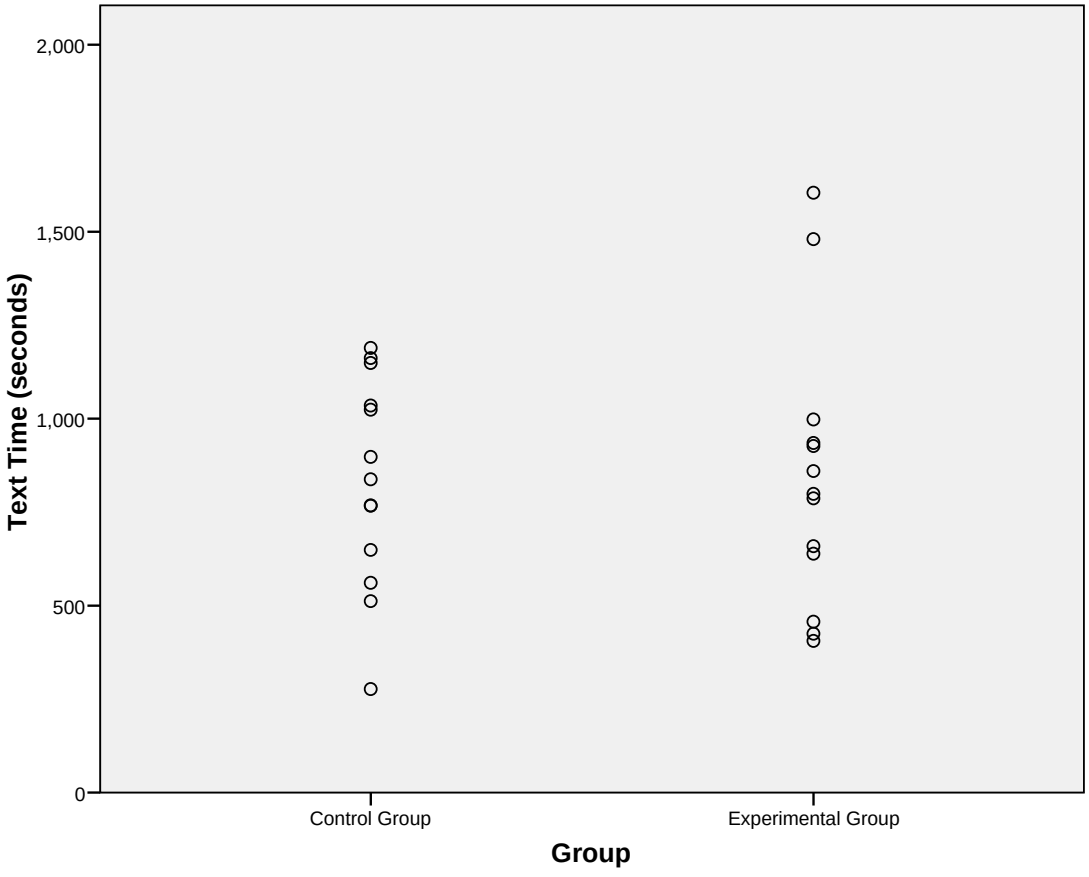
Our third hypothesis was that the availability of the hover functionality will not enable faster reading. The reading time for the control group (M = 833, SD = 280 seconds) was better than the reading time for the experimental group (M = 844, SD = 367.41 seconds), but the difference was not statistically significant. After adjusting for the two participants in the experimental group, the difference in performance increased: M = 814, SD = 279.01 seconds for the control group versus M = 872, SD = 381.39 seconds for the experimental group, but still was not significant.

When analyzing the reading performance we have to mention that there were two outliers in the experimental group whose performance accounted for the poorer performance of the experimental group (see Figure 1). Omitting the performance of either of the two outliers tips the balance towards the experimental group. One of the outliers was not a native English speaker which might account for a certain degree for the slower reading speed. The other outlier, who spent the most time reading, was the participant who mentioned after the session that she was so frustrated by the reading material that she wanted to quit and walk out after reading about half of the text. She said she had to pause and concentrate on calming herself down in order to be able to complete the experiment. She also reported severe headache and feelings of tension and deep frustration. If we exclude the performance of either of the outliers the experimental group outperformed the control group. If we exclude the reading performance of the participant who stopped reading, the experimental group outperformed the control group with almost a minute (52 seconds) in average (reading time M = 833.00, SD = 280.61 for the control

group versus *M* = 781.00, *SD* = 300.71 for the experimental group), but the difference was not statistically significant.

After adjusting for the two participants who did not use the hover, the difference in average reading time decreased to sixteen seconds (*M* = 798.8, *SD* = 310.06 seconds for the experimental group versus *M* = 814.2, *SD* = 279.01 seconds for the control group), the experimental group still performing better than the control group. The difference was not statistically significant.

Figure 1. Graph with Individual Reading Times by Groups



Based on 25 participants' performance it would seem that the availability of the hover function might have contributed to faster reading but the difference in performance

was not statistically significant. We cannot reject the null hypothesis that the hover functionality did not enable faster reading.

Our fourth hypothesis was that the availability of the hover functionality would not improve answers to the comprehension questions. Comprehension scores for the experimental group were higher than comprehension scores for the control group (M = 90.00%, SD = 10.00 versus M = 80.77%, SD = 20.19). The answer times for the experimental group (M = 40.76, SD = 20.45) were also better than the answer times for the control group (M = 48.36, SD = 12.36). After adjustments, the difference in comprehension scores increased (experimental group M = 92.73%, SD = 0.9, control group M = 80.00%, SD = 18.89) and the difference in answering times decreased (experimental group M = 42.00, SD = 22.06 seconds, control group M = 46.44, SD = 12.63 seconds) but the experimental group still outperformed the control group. This suggests that having the hover functionality available when answering the comprehension questions might have constituted an advantage. But the differences in performance (before and after the adjustments) are not statistically significant. We cannot reject the null hypothesis that the availability of the hover functionality did not enhance answering the comprehension questions.

Although we could not reject our null hypotheses that the availability of the hover function did not help content comprehension nor question answering, our data still suggest that on a subjective level, the availability of the hover function was welcomed by almost all participants. When introduced to the hover functionality after completing the experiment, all research subjects in the control group declared that the feature would have been helpful when reading the text and answering the questions. Among the participants

in the experimental group who had the hover feature available, only one participant did not consider that the hover functionality helped or would have helped in working with this text and it was one of the participants who reported not using the hover functionality at all. He did not report any feed-back on the hover functionality; it is unclear if he forgot about it or just simply ignored it as he was not interested. Eleven individual responses in the comment sections of the debriefing questionnaire mentioned the hover functionality as a useful feature that is worth implementing.

In addition, there was a significant negative correlation between using the hover functionality when reading the text and guessing the answer ($r = -.691$, $p = 0.01$). There was also a significant negative correlation between using the hover function when answering the questions and guessing the answer ($r = -.619$, $p = 0.05$). The more frequently participants used the hover functionality throughout the experiment the less inclined they were to simply guess the answer to the comprehension questions and move on. It would seem that having the additional functionality of the hover made the challenge of finding the correct answer in the text less intimidating. In addition, there were significant negative correlations between using the hover when reading and guessing the meaning of the acronym in the comprehension questions ($r = -.826$, $p = 0.01$) and using the hover when answering the comprehension questions and guessing the acronyms ($r = -.627$, $p = 0.05$). While the availability of the hover functionality did not significantly improve the percentage of correct answers or the time spent on answering the questions, it seems to have contributed to participants investing the effort on finding the correct answer, rather than simply guessing and moving on.

Conclusion

This study was designed as a first step toward assessing how the presence of multiple unknown acronyms in text affects reading performance and user satisfaction. In order to assess the role that acronyms play in content comprehension and reading performance we modified a published article to ensure that participants would experience increased cognitive load, as is likely to happen when reading acronym heavy text in real life. The independent variable in our study was the ability to access acronym definitions -- we wanted to see if providing instant access to acronym definitions via mouse hover would enhance term disambiguation therefore decreasing the cognitive load and hence improving reading performance and user satisfaction.

One limitation in the design of out study that we became aware of only afterwards was that by closely following the wording from the text in the comprehension questions we offered an unwanted method for quickly and correctly finding the right answer for some of the questions. The vulnerable questions include the second most difficult questions where participants were asked to identify relationships between concepts and some of the fill–in-the-blank questions. Users could simply copy key words from the question, run a "find" command on the text and quickly find the section of the text that would have enabled them to correctly answer the question without exploring the available answer variants. We did not notice the participants relying on this technique, although most of the participants used the find functionality of the software to find acronyms and

acronym definitions in the text. A more careful wording of the research questions and explicitly instructing the participants not to try to locate entire phrases in the body of the text with the find feature of the browser could have prevented this potential for the quick work-around.

With our recruitment process, we tried to ensure that the participant pool would be likely to find a maximum number of unknown acronyms in the text. Because of the artificial settings (modified document and controlled population) our findings cannot be extrapolated as being relevant to the practice of using acronyms in a written text in general. For the same reasons our findings cannot be instrumented to infer conclusions about the usefulness and efficacy of the practice of defining acronyms only once, when introduced. We have only looked at the acronym problematic from a narrow angle (having multiple unknown acronyms and only defining them once) in an effort to bring this not intensely researched problematic into the focus of the Information Science community. We are hoping that with our small scale research we will spark interest for further enquiries into the complex problem that acronym proliferation and acronym volatility present for user understanding and reading.

Our data shows that the current practice of only defining acronyms once, when introduced, can create serious degrees of user frustration if the text contains a large number of unknown acronyms. Our participants reported that the large number of unknown acronyms in the reading material negatively affected their reading performance and content comprehension. Our study suggests that, as expected, there is a threshold for unknown acronym frequency in a text, and if this threshold is surpassed, it negatively affects reading performance, content comprehension, and user satisfaction. Further

studies should investigate where this threshold is and the factors affecting this threshold.

Although we could not support our hypotheses that the hover functionality enhances reading performance and content comprehension when dealing with an acronym heavy text, the better performance of the experimental group when answering the comprehensive questions both in response accuracy and response time seems to suggest that the availability of the hover functionality might have constituted an advantage. Several factors could account for our inability to observe significant differences in the performance.

First of all, this was a small study. We only had twenty six participants resulting in two small groups of thirteen participants.

Secondly, not all participants relied on the hover functionality equally. Some of the participants in the experimental group reported using the hover frequently throughout the experiment, some participants' usage was more sporadic, and two participants reported not using the hover functionality at all. Given the small size of the groups this variety in use could have also influenced the lack of significance in our results.

The reading material might have influenced our inability to observe significant differences as well. We might have employed too many intrinsic factors to increase the cognitive load. We chose a topic and a text that was most likely to be unfamiliar for most of the participants and indeed most of the acronyms proved to be unknown to all of the participants according to their responses to the debriefing questionnaire. In addition, there was a high frequency of acronyms in the text, making the text highly ambiguous. The effects of all these factors combined might have made the text so difficult to follow that the availability of hover could not counter. Further studies should investigate the effects

of the hover functionality on reading performance with only one of the aforementioned factors influencing the cognitive load as the independent variable.

We also wanted to explore if the implementation of the hover functionality would be positively received by readers of electronic text. The almost unanimous positive feedback on the usability of the hover seems to suggest that it is worth looking into making the hover feature a new standard to govern acronym usage in electronic environment. Further studies should investigate if scholars reading published articles in the field would welcome the availability of this feature. Given that we are talking about a minor addition that would only be visible when needed, and all the potential benefits that could accompany the introduction of this feature, we believe that it is unlikely that implementing such a standard would face resistance from the scholarly community. By enhancing lexical disambiguation and hence reducing the cognitive load, implementing the standard of making acronym definitions available as hover could make texts containing a large number of unknown acronyms less challenging and less intimidating. Such a standard could also act as an incentive for the non-expert readers to engage in reading texts that currently might get abandoned and labeled as "this is way over my head". Our experiment showed a significant negative relationship between using the hover functionality and tendency to guess an answer or an acronym.

In addition to improving user satisfaction, marking up the acronyms in the documents could also improve the performance of automated acronym extraction tools, which could incorporate this mark-up into their acronym identification algorithms, as noted by Larkey et al. (2000).

This study did not investigate if the availability of the hover function helps the process of learning new acronyms. Although it seems likely that the hover functionality will reduce the learning curve required for mastering new terminology, further studies are required to conform or refute this hypothesis.

In addition we think it is worth investigating if the availability of the hover functionality could help avoid the confusion that can result from acronym ambiguity and the existence of undefined acronyms in text. Participants' responses suggested that it is often the case that in the absence of an acronym definition, readers do not invest extra effort in finding a definition. They might try to guess from the context what the acronyms stands for, and if that is hard to do, they might simply skip the acronym and keep reading, in a hope that sooner or later the context will reveal what the acronym stood for. Based on these findings it is very likely that in case of acronym polynymy, when readers are familiar with one definition of an acronym but the acronym is used with a different meaning in the text (for example in our article the acronym WHO was associated both with World Health Organization and Wareda Health Office), users will attribute the wrong meaning to the acronym and keep on reading. In extreme cases this may lead to severe confusion or even comprehension difficulties. The benefits of the hover functionality in this case are obvious: it could easily eliminate user confusion or misunderstandings. Future studies should investigate if users will take their time to "check" known acronym definitions with the hover functionality, especially if the definition does not integrate well with the context or will simply skip the confusing part and keep on reading.

Notes

[1] Oxford English Dictionary, Second Edition 1989, accessed through the UNC – Chapel Hill Libraries Digital Resources on November 2, 2007, from http://dictionary.oed.com.libproxy.lib.unc.edu/cgi/entry/50000218.

[2] Oxford English Dictionary, Second Edition 1989, accessed through the UNC – Chapel Hill Libraries Digital Resources on November 2, 2007, from http://dictionary.oed.com.libproxy.lib.unc.edu/cgi/entry/50000218.

[3] Oxford English Dictionary, Second Edition 1989, accessed through the UNC – Chapel Hill Libraries Digital Resources on November 2, 2007, from http://dictionary.oed.com.libproxy.lib.unc.edu/cgi/entry/50000218.

[4] What is Medstract? – News release at Medstract.org, retrieved on August 10, 2007, from http://medstract.org/

[5] Publication Manual of the American Psychological Association – Fifth Edition. American Psychological Association. Washington, DC. p 104.

[6] Publication Manual of the American Psychological Association – Fifth Edition. American Psychological Association. Washington, DC. p 104.

Bibliography

Adar, E. (2004). SaRAD: a simple and robust abbreviation dictionary. *Bioinformatics, 20*(4), 527-533.

Ao, H., & Takagi, T. (2005). ALICE: an algorithm to extract abbreviations from MEDLINE. *Journal of the American Medical Informatics Association, 12*, 576-586.

Askwall, S. (1985). Computer supported reading vs reading text on paper – a comparison of two reading situations. *International Journal of Man-Machine Studies, 22*, 425-439.

Baddeley, A. (1994). The magical number seven: still magical after all these years? *Psychological Review, 101*(2), 353-357.

Baddeley, A. (2003). Working memory and language: an overview. *Journal of Communication Disorders, 36*(3), 189-208.

Boyarski, D., Neuwirth, C., Forlizzi, J., & Regli, S. H. (1998). A study of fonts designed for screen display. *Proceedings of the SIGCHI 98*, 87-94.

Braten, I., & Stromso, H. I. (2006). Effects of personal epistemology on the understanding of multiple texts. *Reading Psychology, 27*(5), 457-484.

Buchner, A., & Baumgartner, N. (2007). Text-background polarity affects performance irrespective of ambient of ambient illumination and colour contrast. *Ergonomics, 50*(7), 1036-1063.

Chan, A. H. S., & Lee, P. S. K. (2005). Effect of display factors on Chinese reading times, comprehension scores and preferences. *Behaviour & Information Technology, 24*(2), 81-91.

Chang, J. T., Schutze, H., & Althman, R. (2002). Creating an online dictionary of abbreviations from MEDLINE. *Journal of the American Medical Informatics Association, 9,* 612-620.

Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral & Brain Sciences, 24(1),* 154-187.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87-114.

Cushman, W. H. (1986). Reading from microfiche, VDT and the printed page: subjective fatigue and performance. *Human Factors, 28(1),* 63-73.

DeStefano, D., & LeFevre, Jo-Anne. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behaviour, 23,* 1616-1641.

Dillon, A. (1992): Reading from paper versus screens: a critical review of the empirical literature. Retrieved April 01, 2007, from http://hdl.handle.net/2081/1156.

Dillon, A., Richardson, J,. & McKnight, C. (1990). The effects of display size and text splitting on reading lengthy text from screen. *Behaviour & Information Technology, 9(3),* 215-227.

Dyson, M. C. (2004). How physical text layout affects reading from screen. *Behaviour & Information Technology, 23(6),* 377-393.

Dyson, M. C., & Haselgrove, M. (2000). The effects of reading speed and reading patterns on our understanding of text read from screen. *Journal of Research in Reading, 23,* 210-223.

Dyson, M. C., & Haselgrove, M. (2001). The influence of reading speed and line length on the effectiveness of reading from screen. *International Journal of Human-Computer Studies, 54,* 585-612.

Dyson, M. C., & Kipping, G. J. (1997). The legibility of screen formats: are three columns better than one? *Computers & Graphics, 21(6),* 703-712.

Fortin, C., & Rousseau, R. (1998). Interference from short-term memory processing on encoding and reproducing brief durations. *Psychological Research 61(4),* 269-276.

Garland, K. J., & Noyes, J. M. (2004). CRT monitors: Do they interfere with learning? *Behaviour & Information Technology, 23(1),* 43-52.

Gillström, A., & Rönnberg, J. (1995). Comprehension calibration and recall prediction accuracy of texts: Reading skill, reading strategies, and effort. *Journal of Educational Psychology, 87(4),* 545-558.

Gould, J. D., Alfaro, L., Barnes, V., Finn, R., Grischkowsky, N., & Minuto, A. (1987). Reading is slower from CRT displays than from paper: Attempts to isolate a single variable explanation. *Human Factors, 29,* 269-299.

Gould, J. D., Alfaro, L., Finn, R., Haupt, B., & Minuto, A. (1987). Reading from CRT displays can be as fast as reading from paper. *Human Factors, 29,* 323-337.

Kurniawan, S. H., & Zaphiris, P. (2001). Reading online or on paper: Which is faster? Retrieved April 1, 2007, from http://personalpages.manchester.ac.uk/staff/S.Kurniawan/files/HCII_reading.pdf.

Laarni, J. (2002). Searching for optimal methods of presenting dynamic text on different types of screens. Proceedings of the second Nordic conference on Human-computer interaction, Aarhus, Denmark, Session: Short papers, 219-222.

Laarni, J., Simola, J., Kojo, I., & Risto, N. (2004). Reading vertical text from a computer screen. *Behaviour & Information Technology, 23(2),* 75-82.

Larkey, L. S., Ogilvie, P., Price, A. M., & Tamilio, B. (2000). Acrophile: an automated acronym extractor and server. Proceedings of the Fifth ACM International Conference on Digital Libraries, ACM Press, New York, NY, 205–214.

Lam, K., Lam, Y., Liu, J., & Shin, U. (2000). Reading comprehension and rate: one column vs. three columns. Retrieved April 2, 2007, from http://www.otal.umd.edu/SHORE2000/columns/index.html.

Lewis, R. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research, 25,* 93-115.

Liu, H., Aronson, A. R., & Friedman, C. (2002). A study of abbreviations in MEDLINE abstracts. *Proceedings of the American Informatics Association,* 464-468.

Oxford English Dictionary. Second Edition 1989. Accessed through the UNC – Chapel Hill Libraries Digital Resources on November 7, 2007, from http://dictionary.oed.com.libproxy.lib.unc.edu/cgi/entry/50000218.

Mason, R. A, & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brainresearch, 1146,* 115-127.

McCrudden, M., Schraw, G., Hartley, K., & Kiewra, K. (2004). The influence of presentation, organization, and example context on text learning. *The Journal of Experimental Education, 72(4),* 289-306.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29(2),* 111-123.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Mills, C. B., & Weldon, L. J. (1987). Reading text from computer screens. *ACM Computing Surveys, 19(4),* 329-358.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87(2),* 319-334.

Muter, P., & Maurutto, P. (1991). Reading and skimming from computer screens and books: the paperless office revisited?. *Behaviour & Information Technology, 10(4),* 257-266.

Piolat, A., Roussey, J.Y., & Thunin, O. (1997). Effects of screen presentation on text reading and revising. *International Journal of Human Computer Studies, 47,* 565-589.

Publication Manual of the American Psychological Association. Fifth Edition. American Psychological Association, Washington, DC.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology, 82,* 498–504.

Schwartz, A. S. & Hearst, M. A. (2003). A simple algorithm for identifying abbreviation definitions in biomedical text. Retrieved on June 7, 2007 from http://biotext.berkeley.edu/papers/psb03.pdf.

Sheedy, J. E., Subbaram, M. V., & Hayes, J. R. (2003). Filters on computer displays - effects on legibility, performance and comfort. *Behaviour & Information Technology, 22 (6),* 427-433.

Shieh, K. K., Chen, M. T., & Chuang, J. H. (1997). Effects of color combination and typography on identification of characters briefly presented on VDTs. *International Journal of Human-Computer Interaction, 9(2),* 169-181.

Shiffrin, R. M., & Nosofsky, R. M. (1994). Seven plus or minus two: A commentary on capacity limitations. *Psychological review, 101(2),* 357-361.

Stead, W. W., Kelly, B. J., & Kolodner, R. M. (2005). Achievable steps toward building a National Health Information infrastructure in the United States. *Journal of the American Medical Informatics Association, 12,* 113–120.

Stone, D. B, Fisher, S. K, & Eliot, J. (1999). Adult's prior exposure to print as a predictor of the legibility of text on paper and laptop computer. *Reading and Writing: An Interdisciplinary Journal, 11(1),* 1-28.

Taghva, K., Borsack, J., Bullard, B., & Condit, A. (1993). Post-Editing through approximation and global correction. Retrieved August 28, 2007, from http://www.isri.unlv.edu/publications/isripub/Taghva94d.pdf.

Taghva, K., & Gilbreth, J. (1999). Recognizing acronyms and their definitions. *International Journal on Document Analysis and Recognition, 1,* 191-198.

Thornton, R., Maryellen, C. M., & Arnold, J. E. (2000). The concomitant effects of phrase length and informational content in sentence comprehension. *Journal of Psycholinguistic Research, 29(2).*

Wiley, J., & Griffin, T. D. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology, 132(4),* 408-428.

Wilkinson, R. T., & Robinshaw, H. M. (1987). Proof-reading: VDU and paper text compared for speed, accuracy and fatigue. *Behaviour & Information Technology, 6(2),* 125-133.

Wren, J. D., Chang, J. T., Pustejoski, J., Adar, E., Garner, H. R., & Altman, R. B. (2005). Biomedical term mapping databases. *Nucleic Acid Research, 33,* D289-D293.

Wren, J. D., & Garner, H. R. (2002). Heuristics for identification of acronym-definition patterns within text: towards an automated construction of comprehensive Acronym-Definition Dictionaries. *Methods of Information in Medicine, 41(5),* 426-434.

Xu, J., & Huang, Y. (2007). Using SVM to extract acronyms from text. *Soft Computing, 11(4),* 369-373.

Yeates, S. (1999). Automatic extraction of acronyms from text. Retrieved April 4, 2007 from http://www.cs.waikato.ac.nz/~nzdl/publications/1999/Yeates 1999-Auto-Extract.pdf.

Youngman, M., & Scharff, L. (1998). Text width and margin width influences. (Presented at South Western Psychological Association). Retrieved August 10, 2007, from http://hubel.sfasu.edu/research/textmargin.html.

Zahariev, M. (2004). Automatic sense disambiguation for acronyms. SIGIR'04, July 25-29.

Zhou, W., Torvik, V. I., & Smalheiser, N. R. (2006). ADAM: another database of abbreviations in MEDLINE. *Bioinformatics, 22,* 2813-2818.

Ziefle, M. (1998). Effects of display resolution on visual performance. *Human Factors, 40(4),* 554-569.

Appendices

Appendix A – Demographic Questionnaire

Please provide us with some basic information about yourself. This information is solely collected for analysis of our target population.

Q1: Your age range:

- □ 18-25
- □ 26-30
- □ 31-35
- □ 36-40
- □ 41-45
- □ 46-50
- □ 51-55
- □ 56-60
- □ 61+

Q2: Gender:

- □ M
- □ F

Q3: Is English you first language?

- □ Yes
- □ No

Q4: Your undergraduate major: …….

Q5: Your highest degree completed:

- □ BA/BS
- □ MS/MS
- □ PhD/PostDoc
- □ Other: ……

Q6: Graduate program you are currently enrolled in/just graduated:

……..

Q7: Expected year of graduation

- □ 2007
- □ 2008
- □ 2009
- □ 2010
- □ Other: ……..

Q8: I encounter unknown acronyms in readings:

- □ Never
- □ Sometimes
- □ Frequently
- □ Almost always

Q9: When it comes to acronyms: (please select the sentence that most closely describes your usage)

- □ I like to construct my own acronyms and use them in my writings
- □ I regularly use acronyms but I mostly use already established acronyms in the field
- □ I mostly use established acronyms in the field but I keep my acronym usage moderate
- □ I tend to avoid using acronyms in my writings
- □ I never use acronyms in my writings unless I feel that I am required to use them

Q10: When I encounter an unknown acronym in my readings, I most likely:
(please select and rank all that apply
1 – first thing you would do, 2 – next thing you would do, etc.)

- □ Scroll or page to find a definition …….
- □ Try to guess from the context …….
- □ Look it up online …….
- □ Look in the acronym glossary that I own …….
- □ Skip the acronym and keep reading …….
- □ Other: …… …….

Q11: How would you rate your reading speed?

- □ slow
- □ below average
- □ average
- □ above average
- □ fast

Appendix B – Instructions

For the control group

You will be presented a text passage containing a multitude of acronyms. Most likely, some of the acronyms will be unfamiliar to you. The text was selected from a published article and adapted for the purposes of this study. Acronym definitions will be provided when acronyms are first introduced.

Disclaimer: the authors did not provide a definition for every single acronym in the original article. The section you are about to read might contain a limited number of acronyms that are not defined!

Try to strive for the best performance between reading speed and content comprehension.

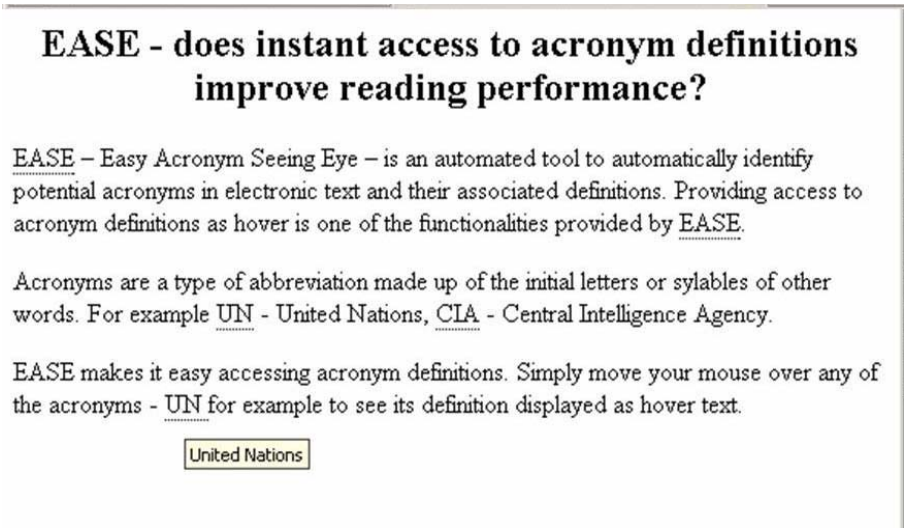
When you are finished reading, please click the Show Questions button. You will be presented a series of questions measuring content comprehension. The questions will be presented one at a time. Every question will have only one correct answer. Select the best answer of the answer choices given on the basis of what was stated or implied in the text passage. When in doubt, pick the most likely answer to be correct.

You will have to provide an answer in order to proceed to the next question. You will not be able to revert to the previous questions, so you need to settle on your final answer before clicking the Next Question button.

Click the Start Experiment button whenever you are ready.

For the experimental group

You will be presented a text passage containing a multitude of acronyms. Most likely some of the acronyms will be unfamiliar to you. Acronym definitions will be provided when acronyms are first introduced. In addition, acronym definitions are available as hover text when mousing over any acronym in the text. See illustration below.



We encourage you to use this functionality of mouse-over whenever you are in doubt about the meaning of an acronym. Try to strive for the best performance between reading speed and content comprehension.

Disclaimer: the authors did not provide a definition for every single acronym in the original article. The section you are about to read might contain a limited number of acronyms that are not defined!

When you are finished reading, please click the Show Questions button. You will be presented a series of questions measuring content comprehension. The questions will be presented one at a time. Every question will have only one correct answer. Select the best answer of the answer choices given on the basis of what was stated or implied in the text passage. When in doubt, pick the most likely answer to be correct.

You will have to provide an answer in order to proceed to the next question. You will not be able to revert to the previous questions, so you need to settle on your final answer before clicking the Next Question button.

Click the Start Experiment button whenever you are ready.

Appendix C – Text

Program Description and Objectives

Africare's Gambella Child Survival Project (GCSP) was implemented from October 1, 2000 to September 30, 2004 in 7 woredas (districts) of southwestern Ethiopia's Gambella Region: Gambella, Itang, Gog, Jor, Abobo, Godere and Dima. Two woredas were not targeted due to security concerns. The goals of the project were: 1) to reduce infant mortality caused by diarrheal and vaccine-preventable diseases, and 2) to reduce the spread of HIV/AIDS/STI s in the region. The objectives were: 1) to reduce vaccine-preventable diseases and vitamin A deficiency through full immunization and vitamin A supplementation; 2) to reduce diarrheal diseases (DD) through improved hygiene and sanitation and improved diarrhea treatment practices; and 3) to increase HIV/AIDS/STI prevention and control knowledge among women and secondary school children (aged 14-18). The project targeted 147,000 people, of which 31,200 were women of reproductive age (WRA) and 26,800 were children under five years. The project's HIV/AIDS intervention especially targeted 12,500 youth in public schools.

Program Accomplishments

- Provision of full immunization and vitamin A supplementation to children through support to the Regional Health Bureau's (RHB's) 21 static and 83 outreach sites. Support included training, cold chain equipment and transportation/logistics.
- Construction of 29 water systems (23 springs, 2 covered hand-dug wells, equipped with hand pumps, one public water distribution point, 2 piped water supply systems for health centers and one water supply system for a public slaughter house).
- Establishment of oral re-hydration therapy (ORT) corners in public health facilities (PHFs).
- Creation and support of HIV /AIDS/STI high school clubs and women's groups (commercial sex workers) with 6,272 benefiting from information/education and condom distribution.
- Strengthened capacity of public health workers (PHWs) in training, case management and supervision skills. Continuous education through courses and workshops provided.
- Strengthened grassroots community capacity through training of community health agents (CHAs), village health committees (VHC) and water user committees (WUCs).
- Introduction of Integrated Management of Childhood Illness (IMCI) to the region.
- Design, production and distribution of IEC/ BCC materials including posters and flip charts with trainer guides in three local languages.

Project Overview

In seven of southwestern Ethiopia's Gambella Region's nine districts, the project sought to: increase immunization coverage, control DDs (including promotion of exclusive breast feeding and appropriate weaning practices), and prevent HIV/AIDS/STI.

The project objectives were: 1) to reduce vaccine-preventable diseases and vitamin A deficiency through full immunization and vitamin A supplementation (40%); 2) to reduce DDs through improved hygiene and sanitation and improved diarrhea treatment at home and in PHFs (40%) and 3) to increase HIV/AIDS/STI prevention and control knowledge among men, women and secondary school children (20%). The approaches adopted to achieve these objectives were: 1) system-level capacity strengthening, 2) community organization, education and mobilization and 3) inter-sectoral collaboration.

Progress by Intervention Area:

a) Expanded Program on Immunization (EPI)

EPI was one of the project's main interventions. The objective was to reduce vaccine-preventable diseases and vitamin A deficiency through full immunization and vitamin A supplementation. To increase vaccination activities, Africare supported the maintenance and repair of refrigerators in PHFs and secured a supply of replacement batteries for solar refrigerators through UNICEF and the Federal Ministry of Health (FMOH).

The EPI program cost for the first two years was 100% covered by GCSP. During year three, the Gambella Regional Health Bureau (GRHB) covered 66% of the cost of EPI program. Africare participated in the Polio Eradication Initiative (PEI) in the region with financial support from the CORE group. Two EPI campaigns were successfully carried out in priority districts with the exception of Dima, where program activities seemed to lag behind schedule due to logistical problems.

b) Control of Diarrhea Diseases (CDD)

The major causes of diarrhea in the project area are a) lack of potable water, b) poor hygiene and sanitation, and c) lack of awareness of how to protect food and water from contamination. The project sought to increase access to potable water, improve case management of childhood diarrhea at household and community levels and to promote proper hand washing among mothers to reduce DDs in children. Activities included: 1) purchase and distribution of ORS (Oral Rehydration Solution) sachets to PHFs ; 2) training of mothers and caretakers on recognition of danger signs in children under five, improved case management of diarrhea at home using ORS , including correct mixing and utilization of ORS using potable water for the re-hydration of children; 3) purchase and distribution of utensils such as measuring jugs to PHFs and 4) construction of wells and springs to increase access to potable water.

ORT corners were established in 20 of the 30 PHFs of the region. Ten thousand sachets were purchased locally and distributed by the Africare project. In support of the ORT corners, Africare purchased supplies and utensils for 20 PHFs. The project educated 9,947 people on diarrhea control.

The Government of Ethiopia (GOE) has adopted IMCI as a strategy throughout the country. However, this strategy was not introduced in Gambella Region until Africare and its partner GRHB advocated for facility-based IMCI to complement the control of DDs and ARI (acute respiratory infection). The FMOH offered its support in April 2002. UNICEF provided support for the implementation of IMCI and provided training for PHWs on IMCI protocols, provision of equipment to PHFs and supervision related activities. To facilitate the introduction of IMCI, Africare and the GRHB, with the support of the Woreda Health Office (WHO), organized a three-day IMCI introductory workshop on IMCI. The facilitators for this workshop were from the FMOH and Addis Ababa University. Twenty nine PHWs from government health facilities and 3 HWs from Nongovernmental Organizations (NGOs) in the region participated. A twoweek IMCI case management training for higher-level PHWs was also organized by the FMOH in Hovan Region. Two PHWs drawn from GRHB and the Africare Country Office participated in this training of trainers on IMCI case management at Jima University. The goal was to prepare PHWs to manage diarrhea in PHFs. The PHWs who acquired knowledge of IMCI case management then taught mothers on proper home care and feeding of sick children

Various activities have been performed in the control and prevention of DDs and its case management at household and health institutions jointly by field program coordinators and PHWs from district health offices and zonal departments. Real involvement of the community, community leaders, heads of zonal health departments and district health offices in planning, implementation, monitoring and decision-making was evident to the evaluator. This participation is key to ownership and crucial to sustainability. It has created awareness in the community on the differences between contaminated water sources and safe/potable water. A slaughterhouse in Metti Town, Godere District, which formerly obtained water from an unprotected source, was provided with a piped water supply system by Africare. This has enabled the slaughterhouse to wash the floor, to clean all equipment such as knives and hooks and to keep the persons who slaughter animals clean.

WUC were established in all villages where springs were developed. Male and female WUC members were officially registered with local government. They were trained in spring maintenance and keeping the area around the springs clean and free of standing water.

The skills and organizational experience gained by local government in these cost-effective solutions to clean water provision are likely to result in the implementation of similar projects the district in the future. Africare was able to demonstrate how communities, local government and NGOs can pool their resources to realize clean water at very low cost compared to similar government sponsored projects in the past.

c) HIV/AIDS/STI Prevention and Control

The program objective was to increase knowledge and awareness regarding the spread and prevention of HIV/AIDS/STI among mothers and secondary school students. The project staff conducted a baseline assessment of HIV/AIDS/STI knowledge among

secondary school students in 2002 with the data analyzed and the report finalized in 2003. High school HIV/AIDS clubs were established in seven high schools. In 2003, Africare purchased and distributed mini media supplies including tape recorders, video decks and megaphones to all high schools in the project area. An estimated 6,272 students, of which 1,062 females, were involved in club activities. The project trained high school teachers in HIV/AIDS prevention. Students took anti- HIV/AIDS messages into the general population through drama and music.

Africare initiated an activity named "Love Life" under its Africa-wide HIV/AIDS Voluntary Service Corps. Love Life focused on women commercial sex workers with multiple sexual partners. In 2003, the number of participant "bar ladies" reached 176. They were organized into 17 groups of ten women per group. Each group has its own leader, elected by the members. The leaders were trained in group facilitation skills and HIV/AIDS prevention and control themes.

Each of the groups met bi-weekly and discussed issues pertinent to the transmission of HIV/AIDS/STI. They were trained in HIV/AIDS/STI prevention, testing and control, negotiation skills and confidence building. All women received free condoms and the group leaders distributed condoms to bars, hotels and restaurants. The knowledge of commercial sex workers on HIV/AIDS/STI has greatly improved. At least one leader ceased being a commercial sex worker and with the project's assistance obtained employment as a mechanic in a local garage. The GCSP baseline KPC survey also informed the project as to priority HIV/AIDS intervention areas and set program objectives. Regarding HIV/AIDS prevention and control, women were asked a series of questions on knowledge and practice. There were significant increases in knowledge of the modes of transmission and means of prevention of HIV/AIDS and where to access condoms from baseline to final.

Cross-cutting Approaches

1. Capacity Building at Community Level
Training of VHC and CHAs helped to bridge the gap between the community and PHFs for health information dissemination to increase communication awareness on health service utilization for immunization and other services and early care seeking at PHFs.

Forty CHAs who were originally trained by UNICEF in Abobo and Godere Districts received a one-week refresher training on the three project intervention areas as well as on malaria control, water/sanitation and harmful traditional practices. It was financially supported by Africare and technically supported by project staff in collaboration with senior PHWs from the GRHB. To use them as assistants in project implementation in areas were there are no CHWs or enable them to play a role in social mobilization, 48 members of VHC were trained for five days in all districts except Jor. The members of the VHC s contributed greatly to the success of the National Immunization Days (NIDS), HIV/AIDS/STI prevention and control and defaulter tracing. Peripheral EPI modular training improved the quality of immunization services and enhanced the cold chain system. This oneweek training organized by GRHB and Africare in Gambella Town

involved twenty-five staff from selected PHFs and one project field coordinator. Participatory project design and health management training was jointly conducted at Jima University in collaboration with the GRHB. Eleven persons, two of whom were Africare staff, participated in the training. A public health expert from Jima University provided the training in a very participatory way. The training was supported by a field visit which enhanced the participants' understanding of project monitoring and evaluation. The participants designed two projects and presented them in class. The trainers reportedly provided constructive feedback to the trainees on their proposals. Training strengthened the region's capacity in project design, monitoring and following up.

Training for water resource management committees at village level involved WUC members from all developed springs and taps. They learned how to handle developed water sources in a proper way and use them safely for consumption. They were trained how to take care of protected springs and water points, and how to repair them whenever necessary.

2. Capacity Building of Local Partner Organizations

Africare emphasized improved organizational capacity, information management systems and health service management as a whole. Health planning exercises were conducted at the district level for the first time so as to decentralize micro health planning. District health managers, members of woreda social service sectors, planners and heads of PHFs were involved in the training. Trainees were provided with basic concepts of planning, identification of health problems, priority setting, objective setting, monitoring, evaluation and feedback. This improved the knowledge of PHWs in recording and data processing for planning and decision making.

Training was conducted in six of the seven project districts with 13-20 health officials from each district participating in the 2-3 day workshops. The project trained PHWs were in IMCI, immunization, and water/sanitation. The GOE adopted IMCI throughout the country but it had not been introduced in Gambella until Africare convinced the FMOH to do so. To launch the IMCI strategy in the region, the project provided refresher training on IMCI case management to enable them to train others. The FMOH organized the twoweek IMCI case management for high-level PHWs in Harar Region. In 2003, FMOH senior staff conducted a three-day IMCI orientation workshop for health managers and other local government officials. Africare, in collaboration with the regional health bureau, sponsored a 15-day IMCI case management training for 25 PHWs at Jima University's Pediatric and Child Health Department. Trainees visited Gambella Hospital and Itang, Pugnido and Metti Health Centers to observe the application to IMCI techniques.

The extensive project intervention in capacity building changed partner organizations. Improvements were reported in management (utilization and reporting on donor funds such as UNICEF's), better cold chain implementation and work habits.

3. Capacity Building of Health System
Africare was effective in strengthening the health management information system and providing technical and logistic support (per diems PHWs participating in EPI outreach, fuel, ORS and ORT material purchase, cold chain maintenance and motor bike maintenance) The baseline KPC survey and health facility assessment helped the GRHB to identify problems in the health institutions (i.e. low EPI coverage, high incidence of diarrhea, low awareness on HIV/AIDS prevention, and low clinical performance of PHWs). Key points are noted below:

- The plans/programs that were supported by the project are parts of the FMOH activities built in already existing structures. These plans will be continued by the FMOH. The capacity building conducted at community and institution level is significant for project sustainability after the project phase out.
- The sustainability plans are realistic as they are within the FMOH plans and as they are built with active participation of the community.
- The capacity building activities at the community level will help to link the facility and the community and vice versa.

4. Strengthening HW Performance
Key results are noted below:

- HW performance was strengthened through training (seminars, workshops) practical work, supervision and review meetings.
- Supportive supervision activities motivated PHWs to keep up their ethics and morale, resulting in better performance.

5. Training
Key points are noted below:

- Training objectives were met as reflected by PHWs in PHFs and VHC members' performances during post training and supportive supervision.
- While a final health facility assessment was not done, improved practices at health institutions and at EPI outreach sites were noted during supportive supervision.
- The best practices and lessons learned were those things observed as PHWs showed good morale in doing their work more effectively than before. The PHWs have started to monitor the cold chain, check vaccine stocks and to request supplies in a timely manner.

Appendix D – Comprehension Questions

Q1: Identify the main topic of the passage you just read
(select one answer that best reflects the content of the text).

The article emphasizes the need to improve the health care system in Ethiopia in order to stop child killer diseases such as HIV/AIDS/STI and diseases caused by the lack of potable water.

- □ The article describes actions taken jointly by UNICEF and WHO for improving the health care system in Ethiopia.
- □ The article describes how activities sponsored by Africare contributed to improving health education and general health in Ethiopia.
- □ The article urges the government of Ethiopia to fund and maintain activities sponsored by Africare.
- □ The article identifies the benefits of implementing the principles of IMCI, as developed by WHO, in health care.

Q2: Which of the options listed below was NOT mentioned as one of the activities/programs sponsored by Africare?

- □ CDD
- □ EPI
- □ NIDS
- □ GCSP
- □ GOE

Q3: Which acronym pair is best suited to fill in the blank?

_____ efforts included: training mothers to wash hands, to recognize early danger signs in children and to properly mix _____ and drinking water.

- □ WRA - IEC
- □ CDD - ORS
- □ IMCI - BCC
- □ ORT - PEI
- □ NIDS – CDD

Q4: Fill in the blank:

_____'s main objective was to reduce vaccine-preventable diseases and vitamin A deficiency through full immunization and vitamin A supplementation.

- □ ARI
- □ EPI
- □ KPC
- □ CDD
- □ WUC

Q5: Fill in the blank:
The UNICEF and WHO sponsored training at Addis Ababa University is mentioned as one of the programs aimed at training _____s.

- □ FMOH
- □ GRHB
- □ PHF
- □ NGO
- □ PHW

Q6: Fill in the blank

_____'s were organized and its members trained in spring maintenance and keeping the area around the springs clean and free of standing water in as effort to provide low cost potable water for the communities.

- □ PHF
- □ PEI
- □ CHA
- □ WRA
- □ WUC

Q7: Fill in the blank:
The text mentions that the training of VHCs and CHAs helped to bridge the gap between the community and _____s.

- □ ORT
- □ WRA
- □ VHC
- □ CDD
- □ PHF

Q8: Which pair of terms is best suited to fill in the blank?
UNICEF provided training for _____ s on IMCI protocols and provision of equipment to _____s.

- □ WRA - ORS
- □ CHW - KPC
- □ ARI - FMOH
- □ GOE - NIDS
- □ PHW - PHF

Q9: Which of the options listed below was NOT mentioned as one of the activities/programs sponsored by Africare?

- □ Control of Diarrhea Diseases
- □ Expanded Program on Immunization
- □ National Immunization Days
- □ Gambella Child Survival Project
- □ Government of Ethiopia

Q10: In this article IMCI stands for _____.

- □ Integrated Management of Community Illness
- □ Integrated Management of Childhood Illness
- □ Integrated Management of Community Immunization
- □ Integrated Management of Childhood Immunization
- □ International Medical Care Initiative

Appendix E – Participant Debriefing Questionnaire

For the Control Group

Q1: I think the text was:
- □ Easy to understand
- □ Somewhat easy to understand
- □ Not too hard to understand
- □ Very hard to understand

Q2: The acronyms in the text were:
- □ Easy to handle
- □ Somewhat problematic
- □ Very problematic

Q3: Answer this questions only if you have answered Somewhat problematic or Very problematic to the previous question
The acronyms were problematic because:
- □ There were too many acronyms in the text
- □ The acronyms were too complex, I did not understand what they stood for
- □ They were not clearly defined
- □ They were not defined
- □ Other…..

Q4: How well did you understand what the comprehension questions were asking?
- □ They were easy to understand
- □ They were a little confusing
- □ They were somewhat confusing
- □ They were very confusing

Q5: What strategy did you use to answer the comprehension questions? Approximate the number of questions you answered using each strategy (eg. **5** of the 10 questions):
- □ knew the correct answer - ____ of the 10 questions
- □ had to look in the text for the answer - ____ of the 10 questions
- □ guessed the answer - ____ of the 10 questions

Q6: What strategy did you use when working with the acronyms in the comprehension questions? Approximate the number of questions you answered using each strategy (eg **5** of the 10 questions):

- □ knew all the acronyms in the question - ____ of the 10 questions
- □ had to look up acronym definitions in the text - ____ of the 10 questions
- □ guessed the acronym definitions - ____ of the 10 questions

Q7: Do you think that you would have found the questions easier to answer if there would have been fewer acronyms in this text:

- □ Yes
- □ No

Q8: Do you think you would have found it easier if acronym definitions were provided in the comprehension questions:

- □ Yes
- □ No

Q9: Please provide your comments, observations, suggestions about the experiment, the text, the acronyms, the comprehension questions, etc.

Q10: Please include whether you think research on use of acronyms is worth doing?

Q11: Did participating in this study influence in any way your attitude towards acronym usage in text?

For the Experimental Group

Q1: I think the text was:
- □ Easy to understand
- □ Somewhat easy to understand
- □ Not too hard to understand
- □ Very hard to understand

Q2: The acronyms in the text were:
- □ Easy to handle
- □ Somewhat problematic
- □ Very problematic

Q3: Answer this questions only if you have answered Somewhat problematic or Very problematic to the previous question
The acronyms were problematic because:
- □ There were too many acronyms in the text
- □ The acronyms were too complex, I did not understand what they stood for
- □ They were not clearly defined
- □ They were not defined
- □ Other…..

Q4: How well did you understand what the comprehension questions were asking?
- □ They were easy to understand
- □ They were a little confusing
- □ They were somewhat confusing
- □ They were very confusing

Q5: What strategy did you use to answer the comprehension questions? Approximate the number of questions you answered using each strategy (eg. **5** of the 10 questions):
- □ knew the correct answer - ____ of the 10 questions
- □ had to look in the text for the answer - ____ of the 10 questions
- □ guessed the answer - ____ of the 10 questions

Q6: What strategy did you use when working with the acronyms in the comprehension questions? Approximate the number of questions you answered using each strategy (eg **5** of the 10 questions):

- □ knew all the acronyms in the question - ____ of the 10 questions
- □ had to look up acronym definitions in the text - ____ of the 10 questions
- □ guessed the acronym definitions - ____ of the 10 questions

Q7: Do you think that you would have found the questions easier to answer if there would have been fewer acronyms in this text:

- □ Yes
- □ No

Q8: Do you think you would have found it easier if acronym definitions were provided in the comprehension questions:

- □ Yes
- □ No

Q9: I used the hover functionality for acronym definitions when reading:

- □ Never
- □ Sometimes
- □ Frequently

Q10: I used the hover functionality for acronym definitions when answering the comprehension questions:

- □ Never
- □ Sometimes (1 -4 questions)
- □ About half the time
- □ Frequently (6 – 10 questions)

Q11: The hover functionality helped me to understand the text and answer the questions:

- □ Yes
- □ No

Q12: Do you think that implementing a new universal standard of requiring provision of acronym definitions as hover in electronic environments would serve the readers' best interests?

- □ Definitely no
- □ Maybe
- □ Most likely
- □ Definitely yes

Q13: Please provide your comments, observations, suggestions about the experiment, the text, the acronyms, the comprehension questions, etc.

Q14: Please include whether you think research on use of acronyms is worth doing?

Q15: Did participating in this study influence in any way your attitude towards acronym usage in text?

Appendix F – User Interface

Figure F1. Reading the Text with the Hover Functionality

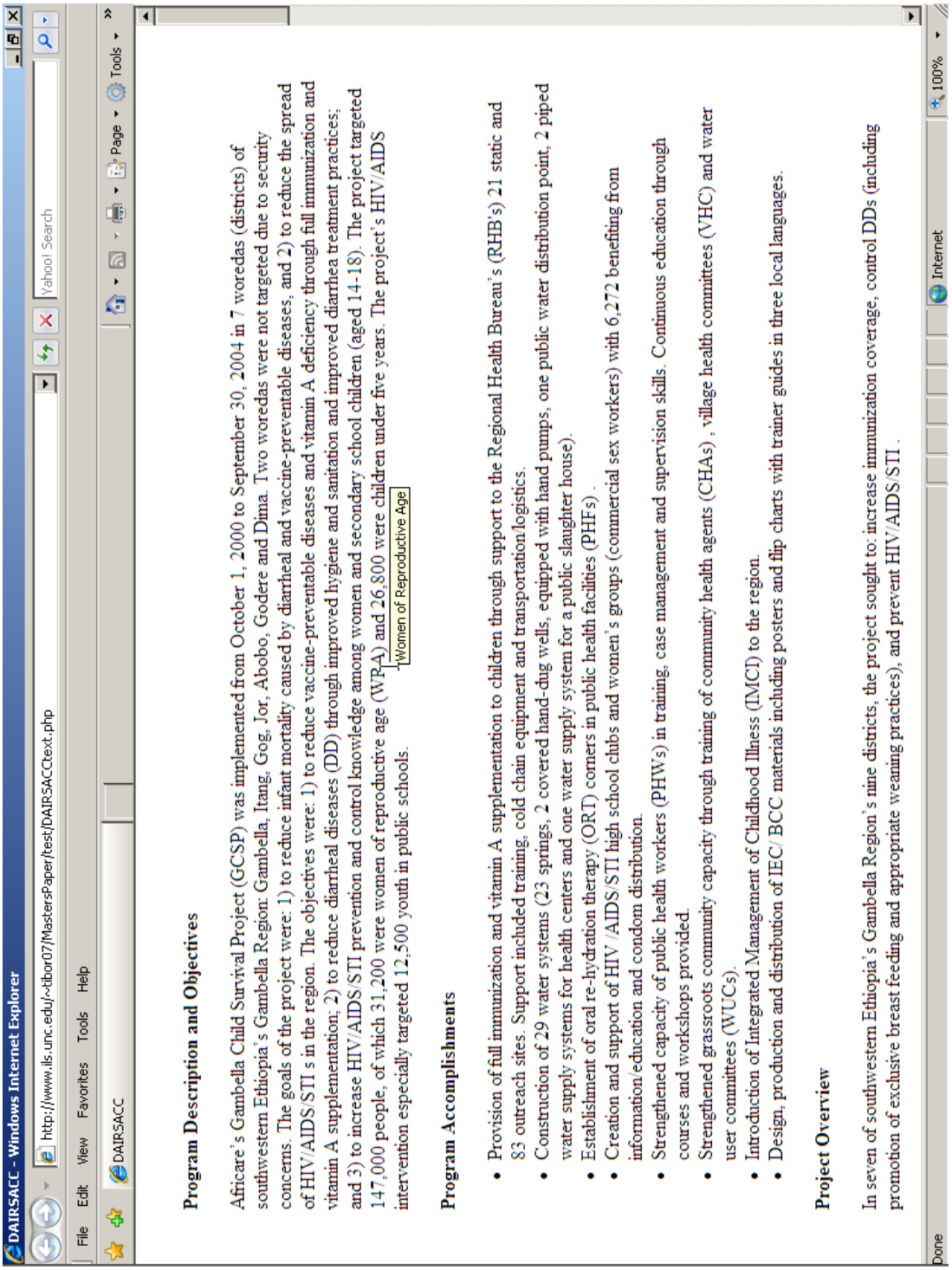
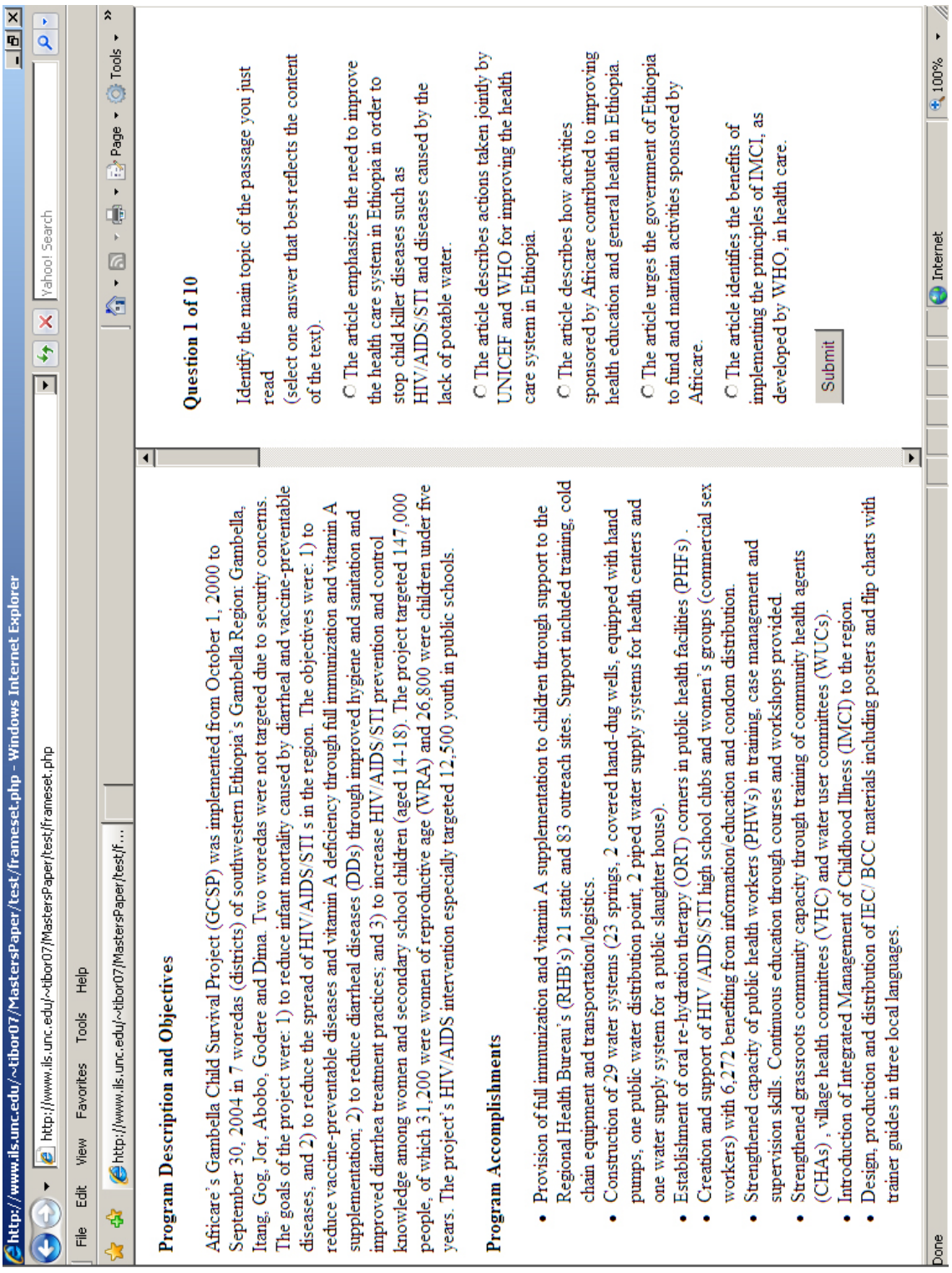


Figure F2. User Interface - Answering the Comprehension Questions



Appendix G – Additional Tables

Table G1

Strategies for Answering the Comprehensive Questions – Individual Responses

	Participant ID	Knew Answer	Looked Up Answer	Guessed Answer
Control Group	1	2	8	0
	3	4	6	0
	5	2	6	2
	7	0	8	2
	9	6	2	2
	11	3	5	2
	13	3	5	2
	15	3	7	0
	17	3	6	1
	19	8	2	0
	21	4	4	2
	23	2	7	2
	25	1	8	1
Exp Group	2	2	8	0
	4	2	7	1
	6	2	8	0
	8	7	0	3
	10	0	10	0
	12	3	5	2
	14	2	7	1
	16	3	7	0
	18	2	5	3
	20	2	4	4
	22	0	8	2
	24	6	4	0
	26	5	4	1

Table G2.

Strategies for Coping with Acronyms in Comprehension Questions- Individual Responses

	Participant ID	Knew Acronym	Looked Up Acronym	Guessed Acronym
Control Group	1	4	6	0
	3	1	9	0
	5	2	6	2
	7	2	8	0
	9	4	5	1
	11	0	8	2
	13	2	6	2
	15	2	8	0
	17	4	5	1
	19	0	10	0
	21	3	6	1
	23	0	8	2
	25	0	8	2
Exp. Group	2	2	8	0
	4	0	10	0
	6	2	8	0
	8	8	0	2
	10	0	10	0
	12	3	7	0
	14	0	9	1
	16	2	8	0
	18	0	6	4
	20	2	4	4
	22	0	8	2
	24	6	4	0
	26	4	5	1