Summary Street: An Intelligent Tutoring System for Improving Student Writing through the use of Latent Semantic Analysis

by

David J. Steinhart

B.S., University of California, Santa Barbara, 1986

M.S., University of California, Santa Barbara, 1989

M.A., University of Colorado, Boulder, 1996

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Psychology

2001

This thesis entitled:

Summary Street: An Intelligent Tutoring System for Improving Student Writing through the use of Latent Semantic Analysis

written by David J. Steinhart

has been approved for the

Department of Psychology

Walter Kintsch

Thomas K Landauer

Gerhard Fischer

Mitchell Nathan

March 21, 2001

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

HRC protocol #1198.19

Steinhart, David J. (Ph.D., Psychology)

Summary Street: An Intelligent Tutoring System for Improving Student Writing through the use of Latent Semantic Analysis

Thesis directed by Professor Walter Kintsch

This dissertation describes the design, evolution, and testing of *Summary Street*, an intelligent tutoring system which uses Latent Semantic Analysis (LSA) to support writing and revision activities. *Summary Street* provides various kinds of automatic feedback, primarily whether a summary adequately covers important source content and fulfills other requirements, such as length. The feedback allows students to engage in extensive, independent practice in writing and revising without placing excessive demands on teachers for feedback. The efficacy of this system was examined in three classroom studies in a Boulder County school.

In the first study, students read texts about three Mesoamerican civilizations and then composed summaries of those texts. One summary was produced using Summary Street, while the other two were produced using a traditional word processor. The students who used Summary Street to summarize the most difficult text produced better summaries than those students who used a word processor.

In the second study, students learned about the human circulatory system and summarized two texts about the heart and lungs. The same pattern from the first study was observed—namely, the students who summarized the more difficult lung text using Summary Street produced better summaries than those students who used a word processor.

The third and final study produced the clearest results. Ten different texts were used in the study, and there was a correlation between text difficulty and the value of the feedback from Summary Street—the more difficult the text, the more Summary Street helped students write better summaries.

In addition to the aforementioned results, individual differences and issues regarding transfer are discussed. Finally, intelligent tutors and the role of technology in the classroom are examined, and Summary Street is compared to existing intelligent tutors.

DEDICATION

This dissertation is dedicated to all children who struggle in their journey to becoming proficient readers and writers.

Acknowledgements

Even though a dissertation has only one author, a dissertation is never the work of just one individual. This work was influenced by countless individuals whom I was fortunate enough to meet during my graduate career. While space (and poor memory) does not permit me to acknowledge them all, I would be remiss if I did not acknowledge the following individuals whose guidance, support, and wisdom so greatly influenced this body of work.

First and foremost, this dissertation would not have been what it is if it weren't for the unique background and perspective each of my committee members brought to the project. Walter Kintsch, my advisor, has been an unwavering source of support and a genuine inspiration. In addition to being *the* luminary in the field of text comprehension, perhaps more importantly, he believed in me, even when I didn't believe in myself. Tom Landauer *is* Latent Semantic Analysis—he brought LSA to the University of Colorado, and was the catalyst for many new and exciting LSA-based projects in addition to this one. He also has an uncanny knack for explaining the most difficult concepts in a way that anyone can understand. Dick Olson is a leading expert on reading disabilities, and his insight into the minds of both children and researchers was invaluable. Gerhard Fischer's expertise in collaborative design and educational technology contributed greatly, and he too possesses the ability to vastly simplify the complex for all to understand. Mitch Nathan's expertise in the field of education and educational technology was right at home in this project, and his work on intelligent tutors was instrumental. Finally, I acknowledge Gerry Stahl, for his expertise in collaborative design, and perhaps more importantly, he was one of the primary architects of State the Essence, the predecessor of Summary Street.

The teachers at Platt Middle School—Lisa Kihn, Ron Lamb, and Cindy Matthews—were enthusiastic about this project from the start, and they provided many wonderful suggestions and a great deal of insight into the teaching process. Somehow, they managed to remain upbeat through the difficult early trials of Summary Street and State the Essence.

Eileen Kintsch provided invaluable organizational and logistic help. I dropped the ball on numerous occasions while running studies and analyzing data, and she was always there to pick it it up! She also provided help with editing and proofreading, two of many tasks at which she greatly excels.

The CLIPR folks—Jon, Ernie, Linda, and Taki—all provided help with technical aspects of this dissertation, often at the 11th hour. They cheerfully provided solutions to many problems that had me in a panic.

Chick Judd and Bill Oliver provided much-needed statistical help, and Corder Lewis coined the name "Summary Street."

I must also gratefully acknowledge Missy Schreiner for help and support above and beyond the call, and for writing a dissertation that was similar enough to mine that I could cite it!

Most of all, I must acknowledge my wife, Eileen. She suffered through missed deadlines, rotten moods (mine, not hers), and many long hours of being apart, and still managed to love and support me in this endeavor. I would not have succeeded without her.

In closing, I would also like to acknowledge the following foods for their help in preparing this dissertation: the unbeatable Butter Paneer from Taj Mahal III in Louisville, York Peppermint Patties, Cinnamon Altoids, and the malted milk balls sold in bulk at the Alferd Packer Grill in the University Memorial Center.

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

Chapter 1:

Introduction

It has been noted that the primary usage of the world-wide-web, that of delivering information, “neglects the power of the Web as an environment through which programs may be accessed and controlled interactively” (Neilson, Thomas, Smeaton, Slater, and Chand, 1996). The ability to interact with the world-wide-web and receive immediate feedback offers exciting possibilities for computer systems that support educational activities such as automated tutoring, distance learning, and collaborative learning. One such system is *Summary Street*, a world-wide-web-based automatic tutoring system designed to help students improve their summarization skills. This dissertation describes the design and evolution of Summary Street, as well as the design and outcomes of school trials designed to test Summary Street in a real-world environment.

Summarization skills are important for students and need to be taught sensibly and efficiently. Brown and Smiley (1978) have shown that given additional study time, younger children do not tend to improve on their original brief recall of a text, as opposed to older children and adults, who do improve. Brown, Day, and Jones (1983) therefore argue that it is “necessary to ensure that one can distinguish between a product that is ‘all the child remembers’ and one that is the result of judgment and effort.” In addition, Brown, Day, and Jones (1983) noted that summaries of younger children who did not prepare a rough draft were dominated by ideas from the first half of the stories (that is, they ran out of room and were unable to include ideas from the second half)[1]. The younger children who prepared a rough draft did not run out of room and included equal amounts of information from both halves of the story. Their summaries were therefore more similar in content coverage to those of older students. The older students' summaries did not exhibit this lopsided content, even when they neglected to prepare a rough draft (Brown, Day, and Jones, 1983).

Summary Street (SS) provides an environment in which students can prepare multiple drafts of a summary, and can apply judgment and effort in their writing. As such, SS engages the student in a cycle of writing, obtaining feedback, and rewriting, that can be continued as long as that student is motivated to improve. In this dissertation, I first justify the choice of summarization as a skill that needs to be improved. Then, SS is described and compared with other interactive learning environments. Next, three experiments designed to assess the efficacy of Summary Street are presented, and finally the ramifications of the results obtained in these experiments are discussed in detail.

---

[1] This behavior of the younger students was informally observed by the experimenter, but was later borne out by dividing the text into two halves and calculating the number of idea units from each half contained in the summary. We witnessed this same behavior while supervising the studies described in the next chapter. In addition, we noticed what seems to be a metacognitive deficit in some children, in that they had difficulties judging the quality of their summaries. It was typical for students to produce what they believed to be a summary

THEORETICAL MOTIVATION: WHY SUMMARIZATION?

Summarization is a difficult skill for children to acquire, one which emerges gradually over years of experience. Brown and her colleagues (Brown and Day, 1983; Brown, Day, and Jones, 1983) found that while summarizing, fifth and seventh grade students relied primarily on the inefficient "copy-delete" strategy (whereby elements of the original text are simply copied verbatim, with some irrelevant portions deleted), whereas older high school and college students used more sophisticated condensation rules, such as construction and generalization. Additionally, Brown and Smiley (1977) had students rate the importance of story units (ideas) using a four-point scale and found that fifth graders were only able to distinguish the most important units as being more important to the theme than the others. Unlike older students, fifth graders were unable to make distinctions beyond that.

The educational benefits of summarization have been borne out in numerous studies. Not only does summarization training improve the quality of students' summaries (e.g., Brown, Campione, and Day, 1981; Cunningham, 1982; Hare and Borchardt, 1984), it has also produced transfer effects on reading comprehension measures (Baumann, 1984; McNeil and Donant, 1982; Bean and Steenwyk, 1984; Rinehart, Stahl, and Erickson, 1986). Furthermore, summarization training has also been demonstrated to benefit students with learning disabilities (e.g., Gajria and Salvia, 1992; Jitendra, Cole, Hoppes, and

---

of an entire text, when in fact their summary only contained information about the first section of a multi-section text.

Wilson, 1998; Malone and Mastropieri, 1992; Palincsar and Brown, 1984; Reis and Leone, 1987; Weisberg and Balajthy, 1990). In light of these results, Rinehart and Thomas (1993) argue that summarization training should begin early in a child's education and further that summarization helps students move appropriately from an "egocentric, 'what's-important-to-me' viewpoint of text to what is important according to others."

Summarization is also an important method for the acquisition of basic content knowledge in many school settings and instructional methods. For example, Palincsar and Brown (1984), espouse summarization as a strategy for fostering comprehension in their *reciprocal teaching* model, and Uttero (1988) developed a teaching model based on *cooperative learning* (Slavin, 1983) in which students use summarization to enhance their understanding of a text after first having paraphrased its essence.

In order to summarize a text, the student must read and comprehend the material, isolate the main ideas, and convey those ideas succinctly. These tasks clearly involve deeper processing than simply reading a text (van Dijk and Kintsch, 1983), and therefore teach critical study skills because the student is required to focus on the main ideas in the text and separate them from the details. Indeed, Baumann (1984) found that students who were instructed on how to identify main ideas (both explicitly and implicitly stated) in paragraphs significantly outperformed non-instructed students on several "near transfer" measures (identifying main ideas vs. supporting details in both explicit and implicit paragraphs and in passages) as well as two "far transfer" measures (outlining paragraphs and passages).

Summarization allows students to develop a deep understanding of complex material and, additionally, to articulate that understanding so that it can be shared with teachers and / or classmates. Teachers have noted marked differences in depth of understanding of topics that have been summarized as opposed to merely read by students (CM and RL, personal communication, 1999). Additionally, students' recall of a text is enhanced if they have summarized it as opposed to merely reading it (Taylor, 1982; Taylor and Beach, 1984; Anderson and Biddle, 1975; Ross and DiVesta, 1976), and in classroom discussions they appear to have a deeper understanding of the material, as evidenced by their detailed reasoning and thoughtful contributions (CM and RL, personal communication, 1999).

Furthermore, the task of summarization makes students aware of the need to learn writing strategies above and beyond simply adding or deleting single words, phrases, or sentences (e.g., Brown and Day, 1983; cf. van Dijk and Kintsch, 1983). Summarization can therefore serve as a starting point for the introduction of higher-level writing strategies in the classroom, such as generalization, synthesis, and maintaining coherence.

Finally, summarization provides practice in expository writing and requires active meaning construction much more so than choosing the best response from a set of choices, or even than writing short answers to isolated questions. Summarization is therefore a highly effective means for constructing and integrating new knowledge, and provides a more authentic evaluation of student knowledge than do traditional tests of comprehension.

Therefore, summarization clearly has a great deal of potential for improving students' learning and writing, but the amount of work required to review and grade multiple drafts of students' writing can be overwhelming. This is where the advantages of a computer-based system such as Summary Street can be found. The next section introduces Latent Semantic Analysis, upon which Summary Street is based, and the subsequent section discusses State the Essence, the precursor to Summary Street.

LATENT SEMANTIC ANALYSIS: AN INTRODUCTION

Latent Semantic Analysis (LSA; Landauer & Dumais, 1997) is a statistical method for representing the meaning of words and passages based on the analysis of a large amount of text. LSA can measure the semantic distance between two "documents" (where a document is simply defined as a collection of one or more words) in a "semantic space," which is created from the input text.

The space is generated by first constructing a matrix whose cell entries are the number of times a given word appears in a given document. As an example, consider the document "The sixth sheik's sixth sheep's sick." If row *i* corresponds to the aforementioned document and column *j* corresponds to the word "sixth", then the value in cell *ij* would be 2 because "sixth" appears in the document twice. The other words appear only once, so the values in the other cells in row *i* would be 1.

In practice, spaces are built from large document sets, where the "documents" consist of individual paragraphs[2]. As an example, the "general

[2] Documents can actually be as small as sentences, but experience has shown that, in general, somewhat large documents tend to work better.

knowledge" space *TASA-all* (provided by Touchstone Applied Science Associates, and consisting of actual text samples from books used in first through twelfth grade classrooms in the United States), consists of over 37,000 documents (approximately 300 words per document), and contains over 92,000 unique words. It should be clear that because a given document only contains a small fraction of the number of unique words that make up the space, most of the entries in the matrix will be zeroes, and it is therefore termed a sparse matrix.

Once the matrix has been filled in, the entries are adjusted using an information-theoretic weighting and the matrix is then subjected to a mathematical technique called singular value decomposition, which is similar in effect to training a neural network. At this point the semantic space is constructed and contains considerably fewer dimensions than the original matrix (typically 200-400). Each original word and each original document is represented as a vector in this space.

The space can be viewed in an instructive but oversimplified way: Documents with similar semantic content are located near one another in the space, and words that have occurred in similar documents are also located near one another in the space. This view uncovers one of the benefits of dimension reduction—words that did not co-occur, but occurred in similar *contexts* (e.g., "doctor" and "physician"), will be grouped together as well. Given that the words and documents in the space are arranged according to their semantic content, it is therefore termed a "semantic space."

Once the space has been created, it can be used to compare novel documents, or those contained within the input set. The documents to be compared are represented as vectors (either already existing within the space for documents that were in the input set, or generated “on the fly” for novel documents by simply summing the vectors of the words contained in the document), and the relation between these two vectors can be calculated. Typically, what is measured is the cosine between the vectors, which yields, in effect, the semantic distance between two words or documents. Another useful measure is the length of a vector, which indicates the amount of “semantic content” a given word or document contains[3].

The *TASA-all* space has been used successfully in numerous applications requiring general knowledge, but specialized spaces must be used when the desired depth of knowledge goes beyond that which could be gleaned from a simple survey of a topic. In these cases, the input set consists of a large number of texts about a specific topic, or the entirety of a textbook about a specific topic, e.g., Mesoamerican Civilizations, the Human Circulatory System, or Introductory Psychology.

---

[3] Here are two examples that may prove helpful in understanding these measures. In the TASA-all space, the cosine between (the vector representations of) the words *bird* and *duck* is 0.51, indicating the words to be highly related. Their respective vector lengths are 2.04 and 0.52, indicating that LSA “knows more” about birds than it does about ducks. Now consider the prepositions *of* and *for*, which, while serving an important syntactical function, contain no semantic content, per se. The cosine between them is 0.95 indicating they are almost indistinguishable to LSA, and their respective vector lengths are 0.06 and 0.17, confirming that as far as LSA is concerned, these words carry very little meaning.

Pre-history of Summary Street: State the Essence

Summary Street (SS) is based on an earlier summarization tool called *State the Essence* (StE)[4]. The goals of StE were the same as those of Summary Street: (1) to provide extended practice with expository writing and revision without overburdening teachers; (2) to integrate seamlessly into teachers' instruction; and (3) to make students responsible for improving their own writing. In addition, while not an explicit goal, it was hoped that StE (and by extension, SS) would act as a starting point for teaching higher level summary writing and revision skills.

StE was tested with sixth graders using texts that were approximately 1000-2000 words in length. These texts were divided into sections and each student was required to provide adequate content coverage of each section in order to write a complete summary. The student either composed a summary in the text window of StE, or copied and pasted a summary written in a text editor or word processor. Once the student covered each section of the text, the next step was to make it the right length. If the summary was too long, the student was then required to shorten the summary while ensuring that the content was still covered.

StE provided several kinds of feedback, including a spell-check. First, a 0-100 point score was determined. To provide a measure of the overall quality of the student's summary, the point score combined three elements: (a) the

---

[4] State the Essence was written and maintained by Gerry Stahl, Rogério Abreu de Paula, Darrell Laham, Michael Jones, Maureen Schreiner, and David Steinhart.

cosine[5] between the student's summary and the original text; (b) the cosines between the summary and each section of the original text; and (c) the summary length. Second, the length of the summary (in words and sentences) was displayed. Finally, for each section of the text that the student was summarizing, StE indicated how well the student had covered it by giving one of three response types: (1) The summary does a nice job of covering the section; (2) The summary covers some of the main ideas of the section; or (3) The summary is missing information about the section. This information constituted the first, general level of feedback. In addition, the student could ask for more detailed feedback: A redundancy check looked for sentences in the student summary that were highly redundant with one another, and a plagiarism check enabled StE to flag sentences that were too close to sentences in the original text.

Early trials with StE showed promise, but students who used StE to compose their summaries fared no better than those who did not, based on teacher evaluation of their finished summaries. Post-experimental interviews indicated that students were overwhelmed with the amount of feedback they received from StE and found it difficult to consolidate the feedback and act upon it. For example, a common situation was for students to be told that their summary was too long, but also was lacking coverage of one or more sections. These contradictory messages were difficult for the students to process, and they were unsure how to begin improving their summaries.

---

[5] Actually 100 times the cosine because cosines range from –1.0 to +1.0, and students prefer 87 to .87 when it comes to a point score.

Students also had difficulty with the coarse-grained section feedback—in many cases addition of content resulted in no change in the section coverage message when in fact the summary now covered more material than it did previously. Finally, some students were overly concerned with their point score and focused on trying to raise it by one or two points to meet or exceed the score of their neighbor.

Following our trials with StE, we felt that the tool held promise, and if we could overcome the problems identified by the students who had used StE, we could develop a tool that met the goals previously outlined and helped students write significantly better summaries. Because StE was maintained by several people, all of whom were also busy with other projects, the program was in a state of disarray by the time the experimental trials were completed, because it had been continuously modified during the trials without regard for its long-term maintainability. Therefore, it was decided that the tool should be completely rewritten[6], retaining the principles that guided the creation of StE, while incorporating what we had learned in our experimental trials. That new tool, named by one of the sixth graders in our trials, is called Summary Street. Several classroom trials of StE, along with the evolution of StE to Summary Street are described in Kintsch, Steinhart, Stahl, Matthews, Lamb, and the LSA Research Group (2001).

---

[6] Summary Street was completely rewritten by the present author with the exception of the database routines which were written and are maintained by Rogério Abreu de Paula.

How Does Summary Street Work?

As with StE, the student composes or copy-pastes a summary into the text window of Summary Street (SS), which operates via a web connection to software running on *lsa.colorado.edu*, a Unix machine maintained by the LSA research group. The student can click a button to get SS to perform a spell-check, and flag any words it does not know. Another button allows the student to request more detailed feedback. While this feedback isn't drastically different than that of StE, the context in which it is proffered is dramatically different (see Figure 1).

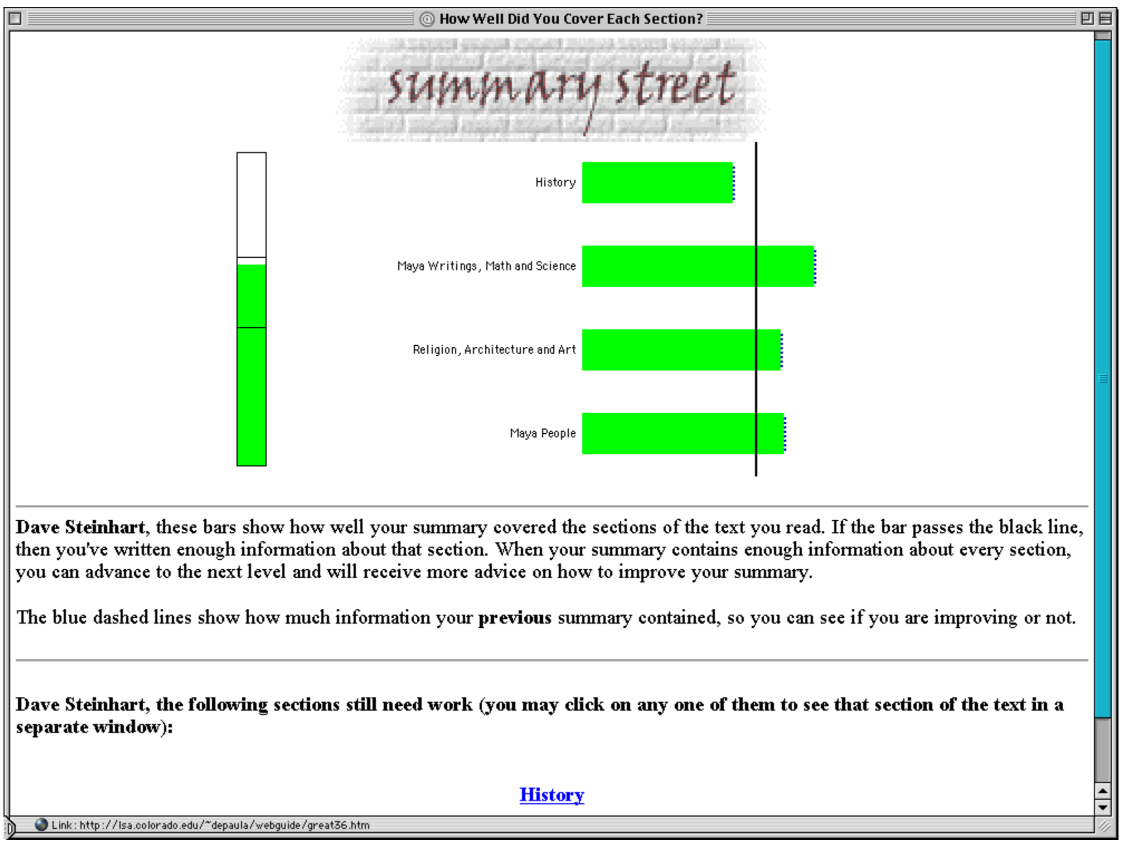


Figure 1. Summary Street feedback display

A Java applet graphically displays the length of the summary (as well as the upper and lower length limits) as a vertical bar. The bar is green if the length is appropriate or red if it is too short or too long. Content coverage for each section is demonstrated by horizontal green bars that either surpass or fall short of a vertical threshold line. In addition, SS provides indicators of how well the section was covered during the last round of feedback, allowing the student to discern at a glance whether or not progress has been made. Those sections of the source text that SS judges to be inadequately covered can be accessed on the screen via hyperlinks provided by SS. The graphical display addresses several problems inherent in StE. First, because the actual length of the summary is not displayed, students are less concerned with it, and therefore are less likely to be engaged in the process of adding or deleting a single word to change the length of their summary. Second, the section feedback is now fine-grained, showing the students at a glance the effect of their latest change. This helps alleviate what was a common source of frustration StE's users: After adding content StE would continue to indicate that the summary covered some, but not all, of the main ideas of a given section. In sum, the graphical display was a marked success. Indeed, many students reported their favorite part of the program as being the "green bars."

The overall point score has been discarded, primarily because it was computed rather arbitrarily, but also because, as noted earlier, it resulted in students obsessing about the score, rather than focusing on content coverage and appropriate length.

As with StE, section coverage is determined by examining the cosine between the entire summary and text of the section, and comparing it to a threshold for the section—if the threshold is met or exceeded, the section is considered adequately covered. One important difference has to do with how the thresholds are computed. During our trials with StE the thresholds were computed empirically, by examining previously written summaries, adjusting the thresholds to fit them, and likely adjusting them several times during the testing process. This method is clearly problematic, as it requires previously written summaries and a great deal of adjustment. With SS we strove to find an automated method for computing the thresholds. In early trials with SS our attempts were largely unsuccessful, but we did settle on an automated method, explained later, which results in satisfactory performance and does not require having access to previously written summaries on the topic.

After the threshold for each section has been met, the student is congratulated and given access to two additional kinds of feedback, a redundancy check, and a relevance check. These additional checks can be used to shorten a summary that is too long, or simply to improve the quality of a summary that is of appropriate length.

The redundancy check operates by computing the cosine between every pair of sentences in the summary, and flagging those pairs whose cosine exceeds a threshold. This method invariably results in the flagging of sentence pairs that are not actually redundant, but merely have a high inter-sentence cosine, such as when the pairs contain several words in common. Additionally, sentence pairs containing antonyms (which generally have a high cosine) can be improperly flagged as being redundant (e.g., “If you’re good you go to heaven,” and “If you’re bad you go to hell” have a cosine of 0.90 in the *TASA-all* space). This “over-flagging” has the positive effect that students must critically examine each flagged sentence pair to decide whether or not they agree with the computer's advice. If they agree that the sentences are redundant, they can either combine the redundant sentences into a single sentence, or delete one of the redundant sentences in its entirety.

The relevance check examines each sentence of the summary to ensure that the cosine between it and the original text as well as its vector length exceed a relevance threshold. Those sentences that do not meet or exceed the threshold are flagged as irrelevant. The relevance check sometimes flags sentences that are relevant, or at least somewhat so. However, because relevance is measured with a combination of cosine and vector length, the relevance check can find sentences that are completely off-topic (e.g., "Thank you for reading my summary"), as well as detail sentences that are relevant, but do not convey much information about the main topics. As with the redundancy check students must carefully evaluate the flagged sentences and judge their relevance.

Summary Street in Action

Having explained how SS works, this section demonstrates SS in action. In the scenario that follows, the writing and revision of one of the students from one of our studies that used texts on source of energy are shown. To protect the student's identity, she will be referred to simply as "B."

The text she is summarizing is about biomass, a renewable energy source. Biomass is any organic matter than can be converted into energy, such as wood, methane, and corn (the complete biomass text can be found in Appendix A). B's rough draft is shown below, spelling errors intact.

My renewable energy source is biomass. Biomass is any organic matter such as wood crops seaweed or animal waste that gets turned into energy. Biomass is the oldest source of enrgy we have. People burn wood to heat thier homes. Biomass gets it's energy from the sun. All plants get energy from the sun durring a prosses called photosyntheses. This prosses gives the plants the energy they need to make oxygen. There are many ways of using biomass energy. You can burn biomass and turn it into steam for electricity. You can preduce methane with biomass. You can make alcohal out of it. You can turn it into Methenol or other gases or liquid gases. There are four types of biomass we use today.

After her rough draft was entered, she saved changes and asked for feedback on her spelling, shown in Figure 2.

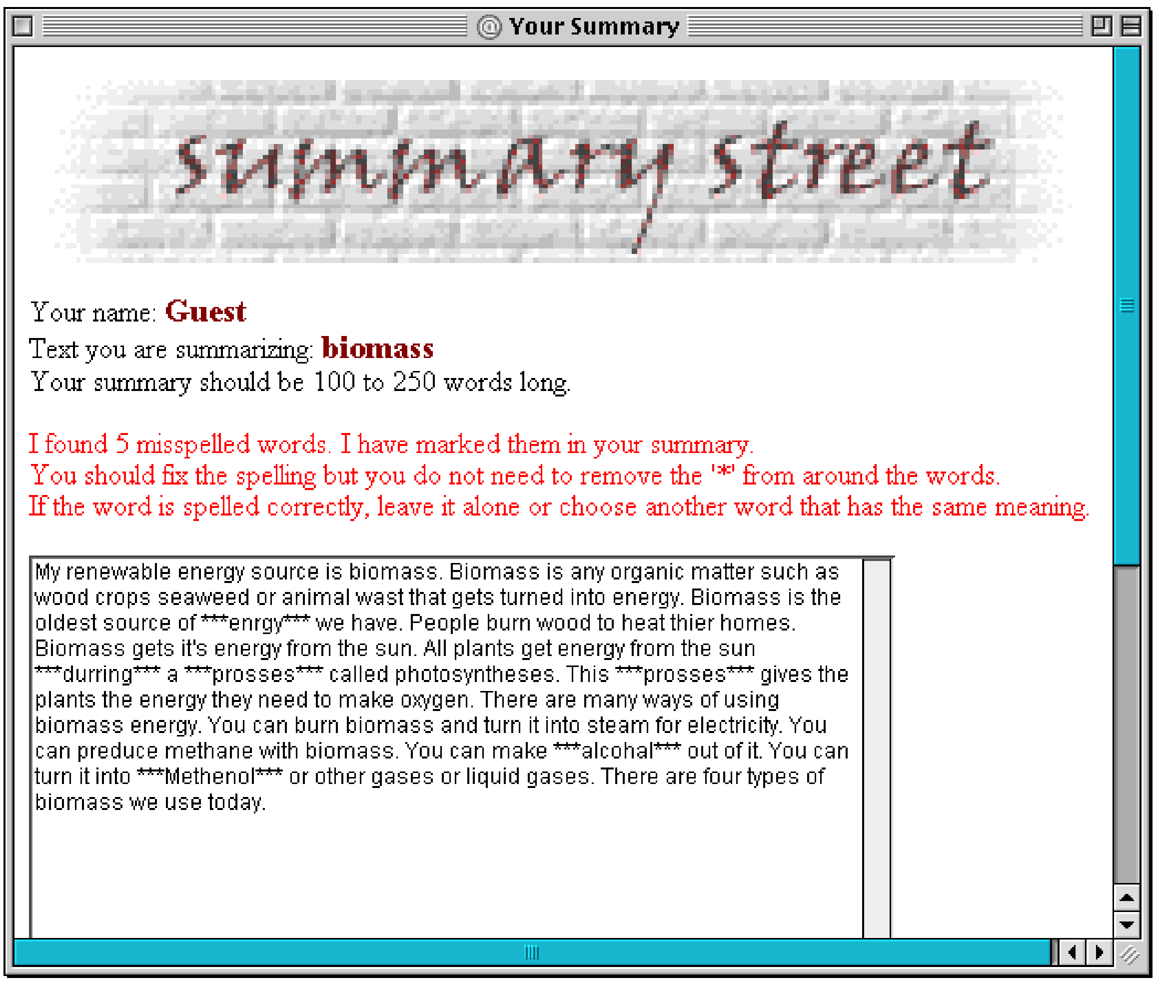


Figure 2. B’s rough draft, after the spelling check.

She fixed the spelling errors, and then submitted the same summary for feedback, which is displayed in Figure 3. According to SS, she has adequately covered sections 1 and 2 of the text, but needs to add more information from sections 3 and 4.

Figure 3. Feedback on B's rough draft.

At this point, she added sixty words to her summary and introduced four new spelling errors, which she corrected in a subsequent submission. Her revised summary is shown below:

> My renewable energy source is biomass. Biomass is any organic matter such as wood crops seaweed or animal waste that gets turned into energy. Biomass is the oldest source of energy we have. People burn wood to heat their homes. Biomass gets it's energy from the sun. All plants get energy from the sun during a process called photosyntheses. This process gives the plants the energy they need to make oxygen.

There are many ways of using biomass energy. You can burn biomass and turn it into steam for electricity. You can produce methane with biomass. You can make alcohol out of it. You can turn it into Methanol or other gases or liquid gases. There are four types of biomass we use today. Wood and agricultural products. Solid waste. Landfill gas and alcohol fuel. Until the mid 1800's wood gave Americans 90 percent of there energy now we only use 3.2 percent of biomass energy. Biomass is good for the environment better than fossil fuels because it reduces the amount of a lot of bad gasses. Thank you for listening to my report I really appreciate it.

She then resubmitted her summary for feedback, shown in Figure 4. At this point she has covered all four sections of the biomass text, and her summary is the proper length. She now has access to the redundancy and relevance checks. She chose the relevance check, the results of which are shown in Figure 5.

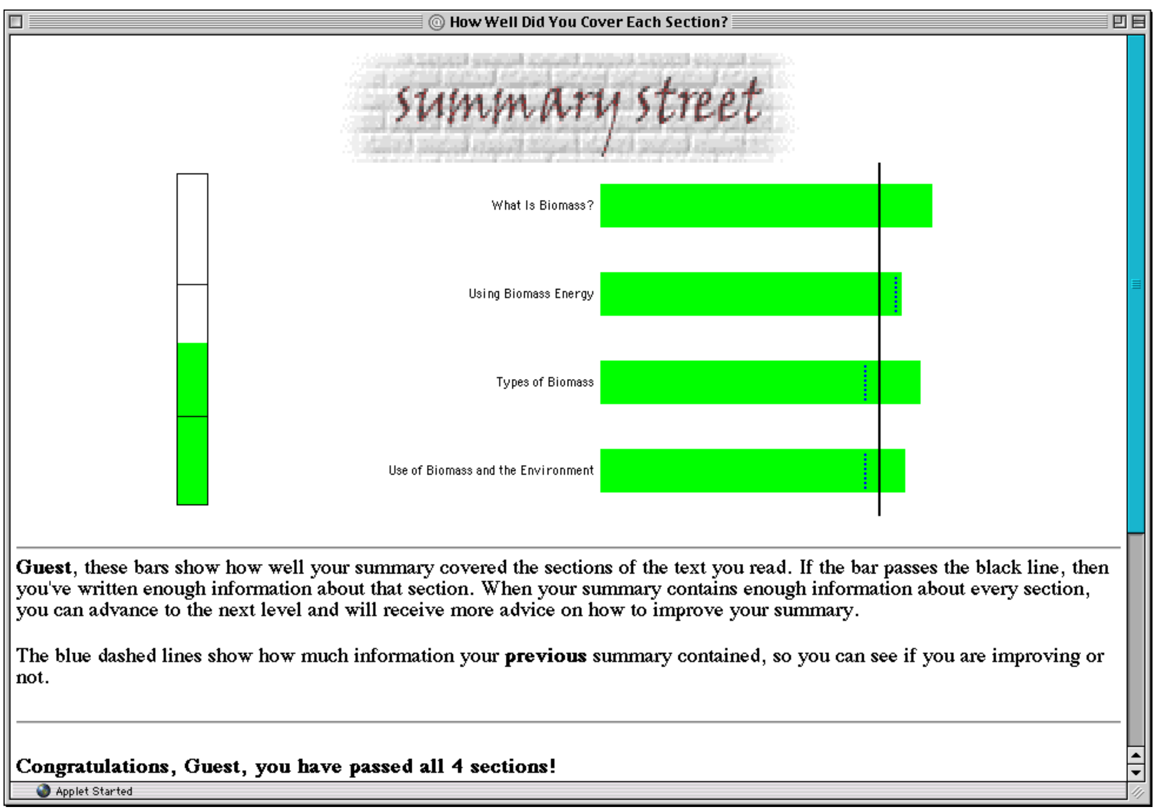


Figure 4. B’s second draft, submitted for feedback.

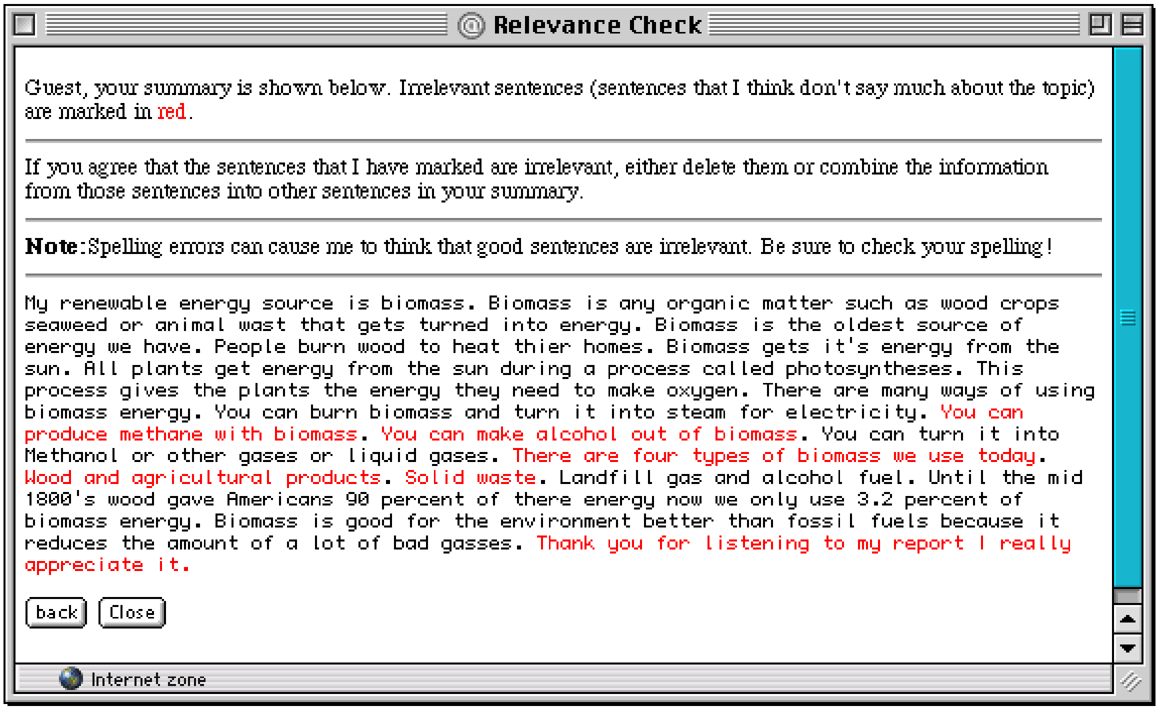


Figure 5. The relevance check performed on B’s near-final draft.

The relevance check did a pretty good job at flagging sentences in B’s summary that are irrelevant and could be removed. The first three flagged sentences are reasonable and were no doubt flagged because they are short and do not convey much meaning. The fourth and fifth sentences (“Wood and agricultural products” and “Solid waste”) are merely sentence fragments, and as such, should be removed or combined with other sentences.

Notice that SS failed to catch the subsequent fragment, “Landfill gas and alcohol fuel.” The final flagged sentence conveys no meaning about the topic, and while it would do fine in an oral presentation about the topic, we argue it does not belong in a summary (especially one where words are at a premium).

B proceeded to remove her final sentence, and with adult assistance, she combined the sentence fragments into a single sentence outlining the four types of biomass. At that point she turned in her summary, shown below:

> My renewable energy source is biomass. Biomass is any organic matter such as wood crops seaweed or animal waste that gets turned into energy. Biomass is the oldest source of energy we have. People burn wood to heat their homes. Biomass gets it's energy from the sun. All plants get energy from the sun during a process called photosyntheses. This process gives the plants the energy they need to make oxygen. There are many ways of using biomass energy. You can burn biomass and turn it into steam for electricity. You can produce methane with biomass. You can make alcohol out of it. You can turn it into Methanol or other gases or liquid gases. There are four types of biomass we use today: wood and agricultural products, solid waste, landfill gas and alcohol fuel. Until the mid 1800's wood gave Americans 90 percent of there energy now we only use 3.2 percent of biomass energy. Biomass is good for the environment better than fossil fuels because it reduces the amount of a lot of bad gasses.

B’s case was a bit atypical in that she made wholesale changes, and the total number of submissions was low. Most students tended to make smaller changes for each submission, resulting in a larger number of total submissions.

Having introduced SS and shown how it works, we now consider how an instructional system such as this compares with other computer-based tutors. While there are many similarities, SS proves to be unique in several areas.

THE EVOLUTION OF COMPUTER-BASED INSTRUCTION

Without attempting a comprehensive review of computer-based instruction (CBI) and interactive learning environments (ILE), it is nonetheless instructive to compare Summary Street (SS) to a handful of the wide variety of computer-based tutoring systems that have been developed in the last few decades. After a brief description of the evolution of CBI from drill-and-practice through intelligent tutoring systems (ITSs) and beyond, subsequent sections explore the architecture of ITSs, the types of problem they are designed to solve, and the importance of assessment of ITSs.

Koschmann (1996), building on the work of Kuhn (1972), views the evolution of CBI as a series of paradigmatic[7] shifts in the field of Instructional Technology (IT). According to Koschmann, the initial CBI paradigm, entitled Computer-Assisted Instruction (CAI)[8] emerged in 1960 with the release of Coursewriter I, a courseware authoring tool (described in Suppes and Macken, 1978). Coursewriter I made it possible for computer-based teaching materials to be developed by persons lacking training in computer programming. Koschmann further notes that applications developed under the CAI paradigm are primarily "straightforward and practical instructional tools designed around the identified needs of the classroom." Studies of CAI applications are designed to assess the instructional efficacy of a newly-introduced technology, as measured by a difference in the students' displayed proficiency.

---

[7] "A scientific achievement represents a paradigm for Kuhn if it raises a compelling set of researchable questions and attracts a following of workers intent on pursuing those questions." (Koschmann, 1996).

[8] Koschmann (1996) points out that the term CAI has been used in many different ways in the IT literature, but he uses it to mean a programmed instruction / instructional design paradigm.

The CAI paradigm spawned the first computer-based learning systems, which were non-intelligent drill-and-practice (DAP) systems such as PLATO IV (McCann, 1975) and Spelling Bee (Fox, 1982; see also Nakatani, Egan, Ruedsueli, Hawley and Lewart, 1986). DAP systems actually existed well before computers had been invented (e.g., Pressey, 1927), but with the advent and widespread adoption of computers, DAP became considerably easier to implement, modify, and maintain**.**

The idea behind DAP, as the name suggests is to drill repeatedly, enabling the student to practice solving a set of problems (or answering a set of multiple choice questions) until they are mastered. While DAP can clearly be used to stimulate rote learning of facts and figures, this method has serious limitations. Because DAP systems are devoid of "smarts," they cannot adapt to novel input on the part of the student; they can only deal (intelligently, that is) with input they expect. Furthermore, DAP systems cannot give feedback beyond indicating whether or not an answer is correct—they cannot indicate *why* an answer is incorrect.

Koschmann argues that the next paradigm shift—intelligent tutoring systems (ITS)—was founded by members of the artificial intelligence (AI) community who had migrated to the educational community. Their motivation was as follows:

> Research in AI is founded upon the conjecture that cognition is, in some sense, a computational process than can be studied through the construction of "intelligent" systems that serve as functional models of the otherwise inaccessible processes of the human mind (Pylyshyn, 1989). If machines can be programmed to display intelligent behavior, there is no reason, at least in principle, that systems could not be designed to assume the role of a skilled teacher. Since one-on-one tutoring is commonly considered the gold standard against which other methods of

> instruction are measured (Bloom, 1984), the paradigm is founded on the proposition that education could be globally improved by providing every student with a personal (albeit machine-based) tutor (Lepper, Woolverton, Mumme, and Gurtner, 1993).
>
> —Koschmann, 1996

The first ITS was called SCHOLAR (Carbonell, 1970). It operated by engaging the student in a dialogue on South American geography. The program and student communicated using natural language questions and answers. SCHOLAR and subsequent dialogue tutors (e.g., Stevens and Collins, 1977; Woolf and McDonald, 1984) were somewhat successful, but the challenge involved in natural language comprehension ultimately proved too great, and the overwhelming majority of current ITSs use a *coached practice* method of tutoring.

Coached practice is a technique whereby the student attempts to solve a problem, the tutor evaluates the solution and gives feedback, and the student can try again if the solution is incorrect. The interactional style of such tutors is certainly less natural than it would be with a human (or dialogue-based) tutor, but the obvious advantage is that the tutor need not understand arbitrary input from the student, and the range of responses the computer must be capable of can be restricted.

While SS relies on coached practice to guide the student through the revision process, it should be clear to the reader that SS performs natural language comprehension when assessing the degree to which the student's writing covers the topic being summarized. Therefore, while it is not a dialogue tutor in the sense of posing questions and evaluating student answers, its ability to comprehend natural language elevates it well above the typical coached practice ITS.

Having briefly covered the evolution from CAI to ITSs, the following section explores the architecture of a typical ITS, and how SS differs from it. These differences are crucial to understanding a further evolution of ITSs, which is discussed in the subsequent section.

The ITS Architecture and Summary Street

Corbett, Koedinger, and Anderson (1997) describe the typical ITS architecture as consisting of a task or problem-solving environment, a domain knowledge module, a student model, and a pedagogical module. The *problem-solving environment* defines the problem-solving activities that the student is trying to perform (e.g., solving a word problem, generating a proof, or writing a summary). The *domain knowledge (or expert) module* contains the knowledge that the student is trying to acquire. In a classical ITS, the domain knowledge module consists of an expert system that is able to solve the same problems the student is trying to solve. The *student model (*or *student modeling module*; cf. Littman and Soloway, 1988) models what the student knows so far. Typically, the student model consists of two parts—a duplicate of the domain knowledge, with each knowledge unit tagged as to whether or not the student has learned it, and a set of misconceptions (frequently called a *bug catalog*), with indications as to whether or not the student has "acquired" that misconception.

Unlike the typical ITS, SS does not contain an expert module, at least in the sense described above. Instead, SS relies on the information in the text being summarized and knowledge contained within LSA (or more correctly, within a semantic space) in order to perform assessment. At first blush, one might liken the semantic space to an expert module, but the analogy is tenuous at best, because a semantic space contains knowledge rather than possessing the ability to solve problems. The text being summarized is perhaps a better analogy to an expert module, but the analogy is still somewhat lacking.

The ability to rely on LSA's semantic knowledge greatly simplifies SS, and in addition simplifies the modifications necessary for SS to handle new texts. Given that the general knowledge space is currently used for two-thirds of the texts that SS knows about, it is possible that the general knowledge space can be used for many new texts. However, even in the case where adding a new text requires building a new space, the labor involved is certainly less than that required to build an expert system, and the resulting semantic space does not exist solely for SS—it can be used by other LSA applications as well.

In a further departure from the typical ITS, SS lacks a student model. In some sense the idea of a student model of summarization is moot—there is no need to model the student's knowledge because SS can determine the knowledge contained in a summary by evaluating the semantic distance between the summary and the text the student is attempting to summarize.

In spite of these architectural differences, we still consider SS to be an ITS. It is certainly intelligent, at least insofar as it is able to deal intelligently with novel input from the student (the summary) and to provide individualized feedback based on what the student has written. And it is certainly a tutoring system in that it provides feedback, allows the user to act on that feedback and then try again, ad infinitum. It could be said that SS is an ITS which revives the good feature of DAP—namely, the ability to interact with the system repetitively—while providing instant feedback that adapts to the content of the student's writing.

Comparison with other Intelligent Tutoring Systems

In a classical ITS, the student is attempting to solve a well-formed problem having a well-defined sequence of steps involved in its solution. The system can therefore identify each step as it is completed by the student and add that to the student model. In addition, it can identify whether the student has fallen prey to common misconceptions and help overcome them.

The majority of the ITSs have operated on formal domains such as mathematics, computer programming, or electronics. More generally, these and many other ITSs have been designed to tutor well-formed problems having narrow, well-defined solutions paths, e.g., how to solve a geometry proof (Anderson, Boyle, and Reiser, 1985), how to write a LISP program (Anderson, Farrell, and Sauers, 1984), how to solve algebra word problems (Nathan, Kintsch, and Young, 1992; Reusser, 1993), or troubleshoot a radar system (Kurland and Tenney, 1988).

Because the problems these systems are designed to tutor are well-defined, these systems tend to be highly structured and directed. Bonar and

Cunningham (1988) describe the LISP tutor as highly directive, "with no provision for informal ideas and intermediate components not directly contained in the top-down goal decomposition."

Summarization, on the other hand, is an ill-defined problem. There is no well-defined sequence of steps one must follow when writing a summary[9], and there are an infinite number of summaries one could write for a given text[10]. In addition, a summary can be written in any order, as opposed to well-formed problems, which are generally solved sequentially.

THE IMPORTANCE OF ASSESSMENT

"Building a tutor and not evaluating it is like building a boat and not taking it in the water" (Shute and Regian, 1993). And yet it is not uncommon for the designers of an ITS to do just that. This section discusses different types of assessment of CBI systems, and concludes with the argument that educational assessment is of paramount importance when attempting to determine the efficacy of an ITS.

Koschmann (1996) argues that concomitant with a paradigm shift, there is a shift in the research issues surrounding the paradigm. Whereas CAI researchers are concerned with how well a new tool works, ITS researchers are concerned with the degree to which the application mimics the behavior

[9] This is not to suggest that there aren't guidelines for summary writing—indeed there are, and some of these guidelines are accessible from the main SS page. However, the guidelines need not be followed in order. In fact, they need not be followed at all.

[10] If length constraints are enforced the number of unique summaries that can be written is finite, albeit VERY large.

of a skilled tutor, changing the assessment focus from the effect on student learning to how well the system performs. “To an ITS researcher, a completed program serves as an existence proof for a theory, whereas to a CAI researcher, no project is complete until the application’s value has been demonstrated in the classroom” (Koschmann, 1996).

Given that the ITS paradigm shift was effected by AI researchers who had migrated into the educational field, it is no surprise that the assessment focus shifted away from student learning. Corbett, Koedinger, and Anderson (1997) point out that “ITSs are generally evaluated according to AI criteria—the coverage of the systems in interpreting and responding to student behaviors—rather than with respect to a cost / benefit analysis of educational effectiveness.” It is clear that ITSs will not make inroads into the educational system until they are assessed from an educational, rather than an AI, standpoint. Indeed, in the words of Corbet et al. (1997), “for ITSs to seriously penetrate the education / training system, the evaluative focus must begin to shift to educational impact and away from AI sufficiency.”

Educational assessment brings into focus the conflict between *internal validity*, where there is a clear causal link between the manipulations of the experiment and the dependent measures, and *external validity*, where there is the ability to generalize from the sample used in the study to the population of interest. According to Shute and Regian (1993), internal validity is easier to accomplish when the study of an ITS is performed in a controlled laboratory environment, and external validity is easier to accomplish when the study of an ITS is performed in a classroom. Educators and policymakers are more likely to be swayed towards the adoption of an ITS that is externally valid, but the experimental design is crucial: “The outcome of the evaluation occasionally reflects the goodness (or poorness) of an experimental design, rather than the efficacy of the ITS” (Shute and Regian, 1993).

In order to assess the efficacy of SS, we decided to focus on its external validity; that is, we tested SS in a classroom environment with actual students. Our experiences are outlined in Chapter 3, which describes in detail the design, execution, and results of three classroom studies of SS. Prior to that, Chapter 2 describes the inner workings of SS—a look “under the hood,” so to speak.

CHAPTER 2:

THE DETAILED INNER WORKINGS OF SUMMARY STREET

Summary Street (SS) consists of an HTML front-end that is accessed through a Java-enabled web browser such as Internet Explorer™ or Netscape Navigator™. The computation and feedback generation is performed by a Perl CGI script, which invokes several LSA tools written in C.

Before delving into the details of how SS provides feedback, it is first necessary to outline what is required to get SS to "understand" a text well enough to provide feedback about (a summary of) that text. Thus far SS has been used to grade summaries of fifteen different texts on topics such as Mesoamerican Civilizations, Sources of Energy, and the Human Circulatory System. In theory, SS will work on any expository text (and probably some narratives as well).

First, the text must be divided into sections (the texts we've worked with so far have been divided into four to six sections, but SS has been used with texts having as few as two and as many as seven sections[11].) Ideally, in order to make it easier on the student, the sections should be roughly equal in

[11] SS works by analyzing the amount of content from each section that is contained in the user's summary, so it will therefore work on a text consisting of a single section. Then, however, the feedback would essentially be restricted to either telling the user whether or not the summary covers all the material in the text.

length and represent logical subtopic divisions of the text, but they need not be.

Next, an appropriate semantic space must be chosen or created. Unfortunately, there is currently no automatic way to determine whether an existing semantic space can be used, or whether a new space needs to be created. After the space has been determined, the section thresholds are computed (the threshold computation is discussed in detail later in this chapter).

Next, a group of files must be created in a text-specific directory on the server machine. The individual sections are each stored in a separate file, while the entire text is stored in an additional file. The LSA vectors (computed using the LSA application **tplus**) for the individual sections and the complete text are contained in their own files. These LSA vector files are used by SS to determine how close the user's summary is to the original text.

In addition, it is recommended that the text be made available in HTML format in order that students may refer to it online while writing their summaries. For each section that is not up to par, SS presents the student with a hyperlink, which, when clicked on, will bring up that section of the text in a new window. If the text is not stored online, the hyperlink will not work; it could be set up to go to a page which says that the text is not available online and that the student should refer to a paper version of the text.

Finally, SS uses a configuration file that must be updated to reflect the existence of the new text. For each text that SS is intended to "understand," the configuration file contains: (1) the name of the text; (2) the name of the semantic space used to grade summaries of this text; (3) the lower and upper word limit for acceptable summaries; (4) the section thresholds; (5) the title of each section of the text; (6) the redundancy and relevance thresholds for the text; (7) and the web address of the online version of the text.

Introductory Page

The main page contains a brief introduction to SS, along with a set of guidelines for summary writing[12], each of which contains a hyperlink to an example of that guideline in use (see Figure 6). The main page also contains hyperlinks to HTML versions of each of the texts that may be summarized using SS. Finally, the page contains a "login" section. In this section, the user is required to enter her initials, select the topic to be summarized, and decide whether to continue working on a previously-written summary, or to enter a brand new summary (see Figure 7).

[12] These guidelines are adapted from Brown and Day (1983).

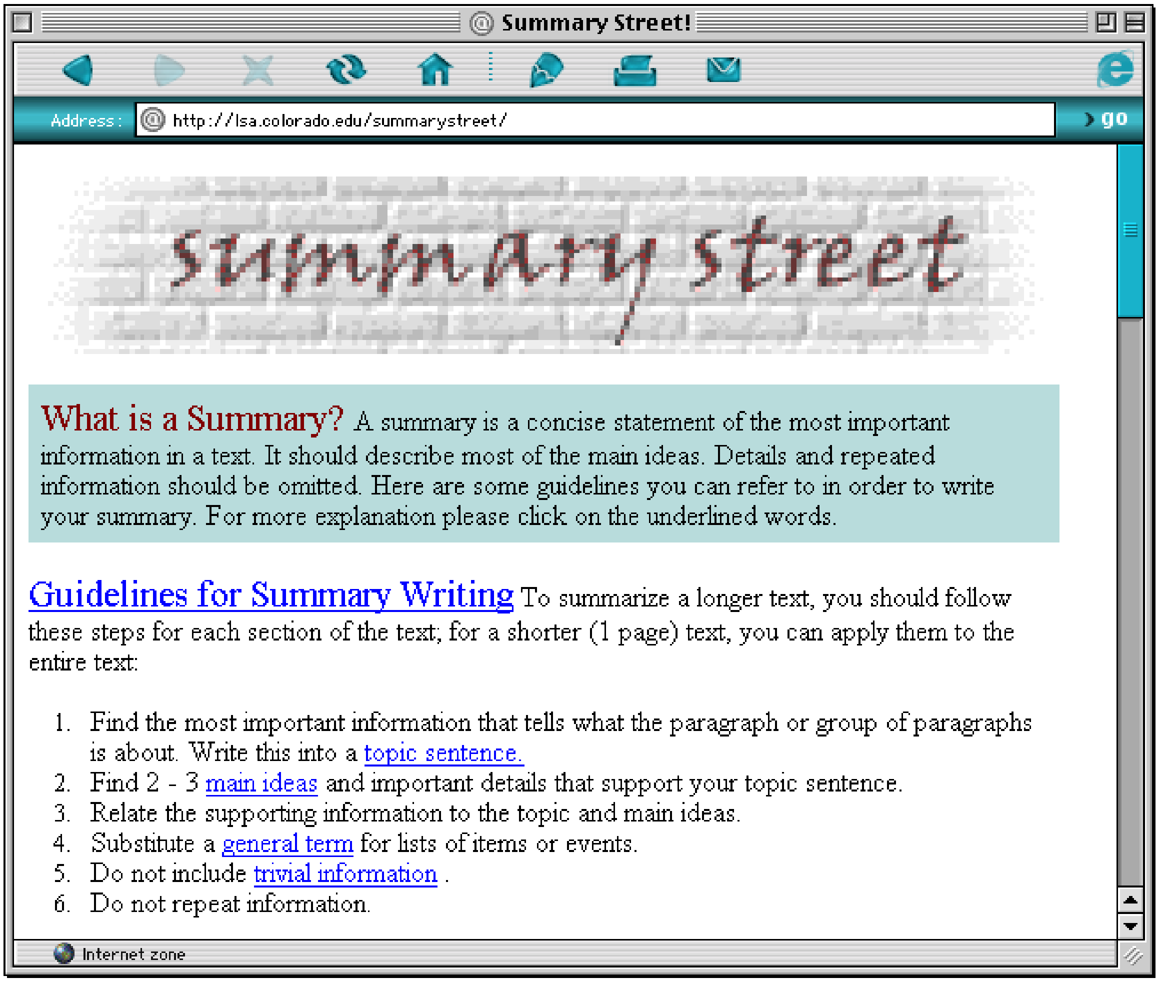


Figure 6. The Summary Street main page, containing guidelines for summary writing adapted from Brown and Day, 1983

Figure 7. The Summary Street login section, requiring users to enter their initials, select a topic, and choose whether they will be working on a previous summary or starting a new one.

SS stores user information and all versions of the users' summaries in a MySQL database. When the user logs in to the system, a new window entitled *Your Summary* is displayed. If the user chooses to continue working on a previous summary, the latest summary is displayed in the window (see Figure 8). Otherwise the user is presented with an empty text box in which to enter a summary. At any time while entering the summary, the user has three choices: (1) save changes / check spelling; (2) receive content feedback; or (3) format the summary for printing.

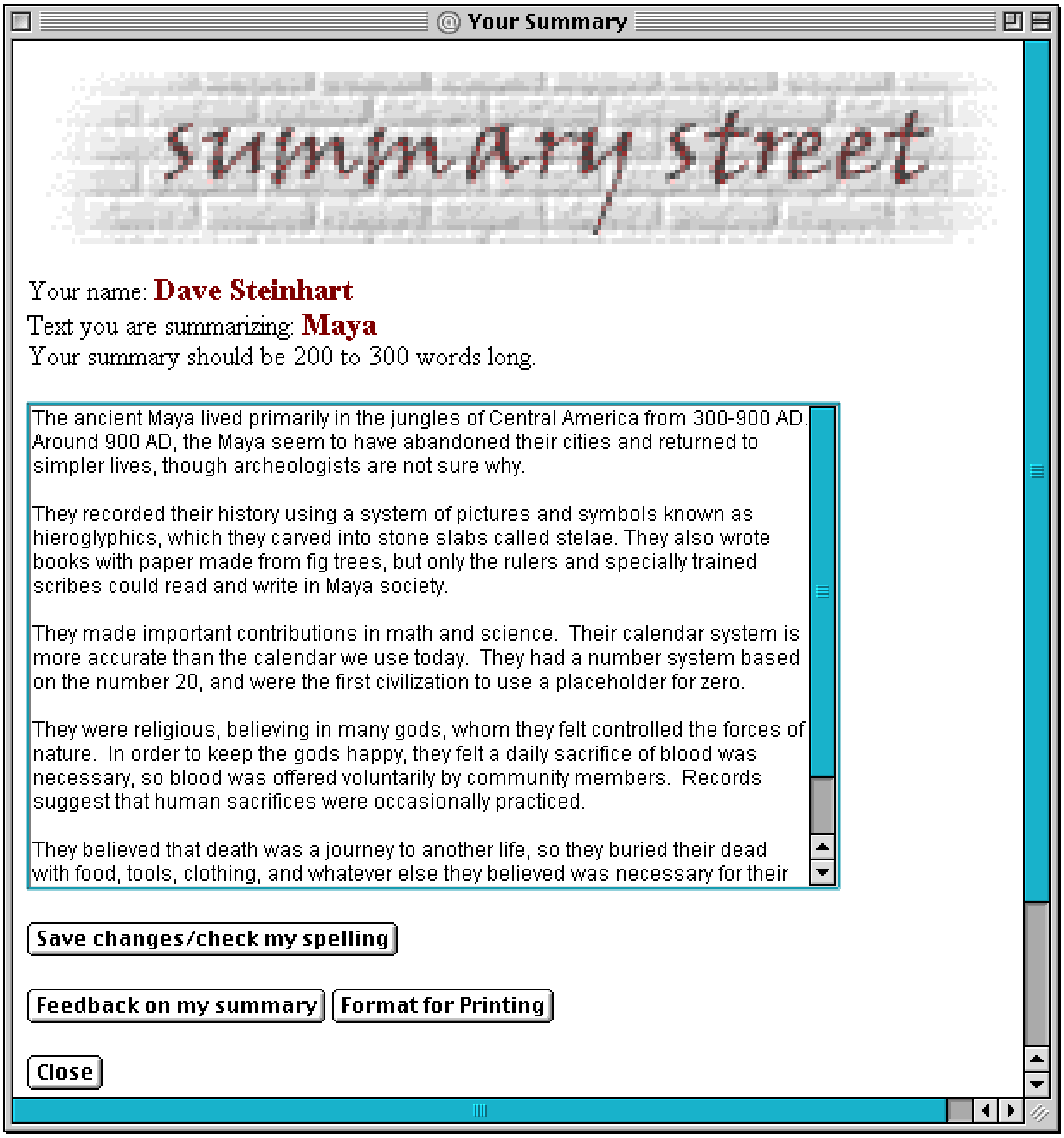


Figure 8. *Your Summary* window showing a saved summary written about the Maya and the various option buttons

When the user chooses to save changes, the entire summary, as well as the date and time of the submission are stored in the database[13]. In addition, SS checks the user's spelling. The spell check, which at first glance seems a rather simple procedure, is actually performed in three passes. On the first pass, the Unix **spell** command is invoked on the entire summary. Due to the shortcomings of the online Unix dictionary (and indeed of any dictionary), this first pass will flag too many words as being misspelled. Many domain-specific words such as *chinampas*[14] and *yellowcake*[15], as well as proper names such as *Greeley* [16]and *Mojave*[17] are not in the Unix online dictionary, but can be found in one or more of the semantic spaces.

Therefore, the second pass takes the misspelled words from the first pass and uses the LSA tool **getkeydb** in an attempt to find some or all of those words in the semantic space being used to grade the summaries. Those that are found in the semantic space are removed from the list of misspelled words. Finally, in some cases there are correctly spelled words that are

[13] Currently, users are only able to retrieve their latest summary, but SS could be easily modified to allow retrieval of earlier versions as well.

[14] Floating gardens built by the Aztecs.

[15] A yellow powder consisting predominantly of uranium, produced when crushed uranium ore is placed into an acid solution and then allowed to dry after the acid is drained off.

[16] Charles Greeley Abbott, an American Astrophysicist, invented an extremely efficient solar boiler in 1936.

[17] California desert where LUZ, the world's first solar electric plant was built.

neither in the Unix online dictionary nor in the semantic space itself [18]. Therefore the final pass involves consulting a wordlist contained in an auxiliary Perl configuration file, effectively allowing the teacher or administrator to create a custom dictionary of such words "on the fly."[19]

FEEDBACK ON MY SUMMARY

When the user requests feedback, a new window is generated, and a Java applet is invoked to display the feedback in graphical form (see Figure 9). The window contains two types of feedback. First, a vertical bar indicates the length of the user's summary relative to the upper and lower length bounds[20]. The user can determine at a glance whether the summary is the appropriate length because if so, the bar is green; otherwise, the bar is red.

---

[18] The word *yellowcake* is such an example. It is in the nuclear energy text that students read and summarized, but it does not exist in the TASA-all semantic space that was used to grade their summaries. (Which means that *yellowcake,* while not flagged as misspelled by Summary Street, does not actually add any semantic content to a summary in which it is contained, because it is not in the semantic space. If many of the words of a text are not contained in the semantic space being used to grade summaries of that text, it would be advisable to "fold in" the text to the existing space, or create a new space containing that text and others on the topic.)

[19] Of course this auxiliary dictionary can be created in advance, but if a student should use a correctly-spelled word that the system believes is misspelled, the error can be fixed immediately.

[20] The length bounds for each topic are held in a Perl configuration file, which may be modified by the administrator. Given that it is more difficult to write a short summary than a long one, the length bounds can be reduced in order to create a more challenging task for the student.

Figure 9. Feedback window showing length and section coverage feedback. At the bottom is a hyperlink to the first section of the Maya text, which is not adequately covered in my summary.

The second type of feedback is content feedback, which is presented as a series of stacked horizontal bars. One bar is displayed for each section of the text, representing the amount of content from that section that the user has included in the summary. The length of the bar is a reflection of the cosine between the user’s entire summary and the corresponding section of the original text. On the right side of the screen is a vertical threshold (criterion) line[21]. Finally, blue dashed lines indicate where the bars were positioned during the user's *previous* submission, so that the user can see whether or not progress is being made[22].

In order for the user's summary to be considered complete, each section must be covered adequately; that is, each section bar must meet or exceed the threshold line.[23]

THRESHOLD COMPUTATION

Determining the thresholds for each section of the text is crucial. Several methods were tried, with the best method (and its shortcomings) described below. The other methods are discussed in the next section.

First, the section is divided into sentences. Next, the cosine between each sentence and every other sentence in the section is computed. That sentence with the highest average cosine to all the other sentences in the section is considered to be the "typical" sentence for that section. This process is repeated for each section of the text, and each of these typical sentences is then combined into a "typical" summary. Finally, the cosines between the

---

[21] While it appears that the threshold is the same for each section, this is actually not the case. The bars represent the percentage of the threshold that has been covered by the student's summary (i.e., the cosine between the student's summary and the corresponding section of the text, divided by the threshold).

[22] In the figure shown, the blue dashed lines are located directly at the end of the green bars, indicating that there was no change in content from the previous submission.

[23] While it appears that each section has the same cosine threshold, in actual fact, that is not the case. The bars are scaled so that they represent a proportion of the threshold, rather than an absolute value. The threshold computation is explained in the next section.

typical summary and each section of the text are computed, and those cosines become the thresholds for each section of the text.

As an example, consider the text below, which is the first section of a text on geothermal energy that was used in our Sources of Energy study (see the next chapter). The average cosine between that sentence and all of the others in the section is shown in parentheses following each sentence of the text. The bold-faced sentence is the one with the highest average cosine, or the “typical” sentence.

> **Geothermal energy comes from the heat within the earth**[24] (0.40). The word geothermal comes from the Greek words *geo*, meaning earth, and *therme*, meaning heat (0.26). People around the world use geothermal energy to produce electricity, to heat buildings and greenhouses, and for other purposes (0.27). The earth's core lies almost 4,000 miles beneath the earth's surface (0.36). The double-layered core is made up of very hot molten iron surrounding a solid iron center (0.23). Estimates of the temperature of the core range from 5,000 to 11,000 degrees Fahrenheit (0.17). Heat is continuously produced within the earth by the slow decay of radioactive particles that is natural in all rocks (0.31). Surrounding the earth's core is the mantle, thought to be partly rock and partly magma (0.380). The mantle is about 1,800 miles thick (0.24). The outermost layer of the earth, the insulating crust, is not one continuous sheet of rock, like the shell of an egg, but is broken into pieces called plates (0.34). These slabs of continents and ocean floor drift apart and push against each other at the rate of about one inch per year in a process called continental drift (0.13). Magma (molten rock) may come quite close to the surface where the crust has been thinned, faulted, or fractured by plate tectonics (0.32). When this near-surface heat is transferred to water, a usable form of geothermal energy is created (0.33). Geothermal energy is called a renewable energy source because the water is replenished by rainfall, and the heat is continuously produced by the earth (0.37).

---

[24] In this case the topic sentence turned out to be the one with the highest average cosine, but that is not always the case.

The process is repeated with the other five sections, and the six "typical" sentences are combined to form a "typical" summary, shown below.

> Geothermal energy comes from the heat within the earth. Electricity is now produced from geothermal energy in 21 countries, including the United States. Usually geothermal energy is deep underground. Two main types of hydrothermal resources are used to generate electricity: dry steam vapor-dominated reservoirs, and hot water liquid-dominated reservoirs. Geothermal power plants can produce electricity as cheaply as some conventional power plants. Newer geothermal power plants can even inject these gases back into the geothermal wells.

This "typical" summary actual reads quite well and is even a reasonable synopsis of the text. However, it often reads much worse; this matters little given that it is not read by anyone. Instead, this summary is used to generate the thresholds, which are simply the cosines between it and each section of the text. After computing these cosines, the thresholds for sections one through six turn out to be 0.51, 0.78, 0.48, 0.72, 0.76, and 0.65, respectively. In effect, the summary above is deemed to be the minimum that a student could produce and have it be accepted by SS as covering enough content for all of the sections. Of course we hope the student will do much better at choosing the important information from the text. Furthermore, in addition, the student's summary should be somewhat more elaborated depending on the minimum length constraint; often it is considerably longer[25].

One problem with this method has to do with using only the cosine to determine whether or not the student has written enough about each section of the text. Considering only the cosine is problematic because a high cosine can be obtained with a single word. For example, the word “solar” has a cosine of 0.42, 0.41, 0.37, 0.37, 0.37, and 0.40, respectively, with the six sections of the Solar Energy text used in our study. This does not usually present a problem in practice, because there is also a minimum length requirement that must be met in order for the summary to be considered complete. However, it is not an acceptable long-term solution to use the minimum length requirement to mediate a shortcoming in the threshold computation[26]. One alternative will be discussed in the next section.

FORMAT FOR PRINTING

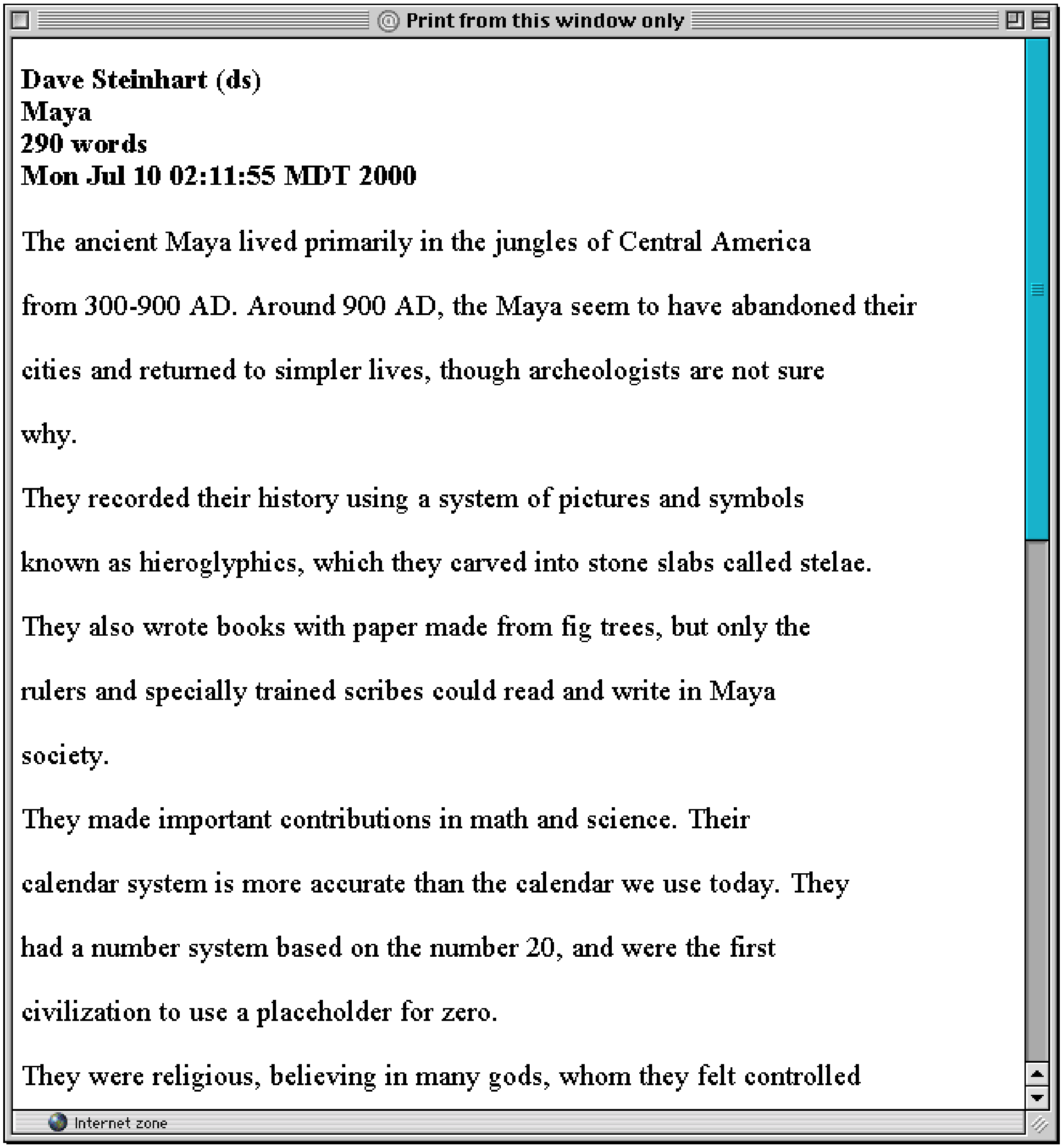
This function was added so that the student’s summary could be nicely printed on paper for proofreading or turning in to the teacher. When this function is invoked, a new window is generated, and the user’s summary is displayed, double-spaced, along with the user’s name, the topic being summarized, and the date/time (see Figure 10). In addition, the length of the summary is printed at the top of this window. This was done as a courtesy for

---

[25] The “typical” summary above is 93 words long, whereas the students in the study were required to write summaries between 150 and 250 words long.

[26] I am of the belief that a minimum length requirement is artificial and should be avoided. Instead, the threshold computation should be improved. Surely we don’t want to accept a two-sentence summary of a text when a summary of approximately 275 words is the assignment. But if the length requirement is 250-300 words, should we penalize a student who covers all the relevant material in 187 words? I hope it would be possible to devise a threshold-setting technique that would enable SS to accept a well-written summary that is shorter than the norm without requiring the student to add needless information. This is likely to be a rare occurrence, but it should be allowed.

the teachers, so that they could determine at a glance whether or not the student's summary was the appropriate length. However, as soon as the students figured out that they could access the actual length in this window, they often used the format function simply to check on the exact word count of their summaries. This function should be modified to prevent it from printing out the summary length. Instead, the length should be stored in the database and accessible to the teacher.



Print from this window only

Dave Steinhart (ds)
Maya
290 words
Mon Jul 10 02:11:55 MDT 2000

The ancient Maya lived primarily in the jungles of Central America from 300-900 AD. Around 900 AD, the Maya seem to have abandoned their cities and returned to simpler lives, though archeologists are not sure why.

They recorded their history using a system of pictures and symbols known as hieroglyphics, which they carved into stone slabs called stelae. They also wrote books with paper made from fig trees, but only the rulers and specially trained scribes could read and write in Maya society.

They made important contributions in math and science. Their calendar system is more accurate than the calendar we use today. They had a number system based on the number 20, and were the first civilization to use a placeholder for zero.

They were religious, believing in many gods, whom they felt controlled

Internet zone

Figure 10. Format for printing window showing the summary double-spaced, ready to be printed.

Redundancy Check

When the student has adequately covered each of the sections of the text being summarized (i.e., all of the section bars meet or exceed the criterion line), access is granted to two additional functions which can help reduce the length of the student's summary. Given that our past experience showed that students are unable to process a great deal of feedback at once, we decided to restrict access to these additional functions in order to avoid overwhelming the students. The first of those functions, the redundancy check, is described in this section. The second function, the relevance check, is described in the next section.

The redundancy check opens up a new window, and then computes the cosine between each pair of sentences in the user's summary. Those pairs whose cosine exceeds a redundancy threshold are flagged in red (see Figure 11).

The redundancy threshold was defined to be two standard deviations above the average sentence-to-sentence cosine in the text being summarized This method means that each text may have its own redundancy threshold (stored in the SS configuration file), but in practice one threshold has been used so far for all of the texts. This threshold is adequate, but certainly not perfect, with the result being that too many sentence pairs are flagged as being redundant. Some of the erroneous flagging cannot be helped—as mentioned earlier, many pairs of words have high cosines with one another and are not redundant (e.g., antonyms). If one or more of these pairs appear in a summary, the sentences that contain them can have a cosine of 0.90 or higher, and will certainly appear to be redundant.

An example that came up in the Mesoamerican Civilizations trial was a student who wrote "If you're good, you go to heaven," and followed it with

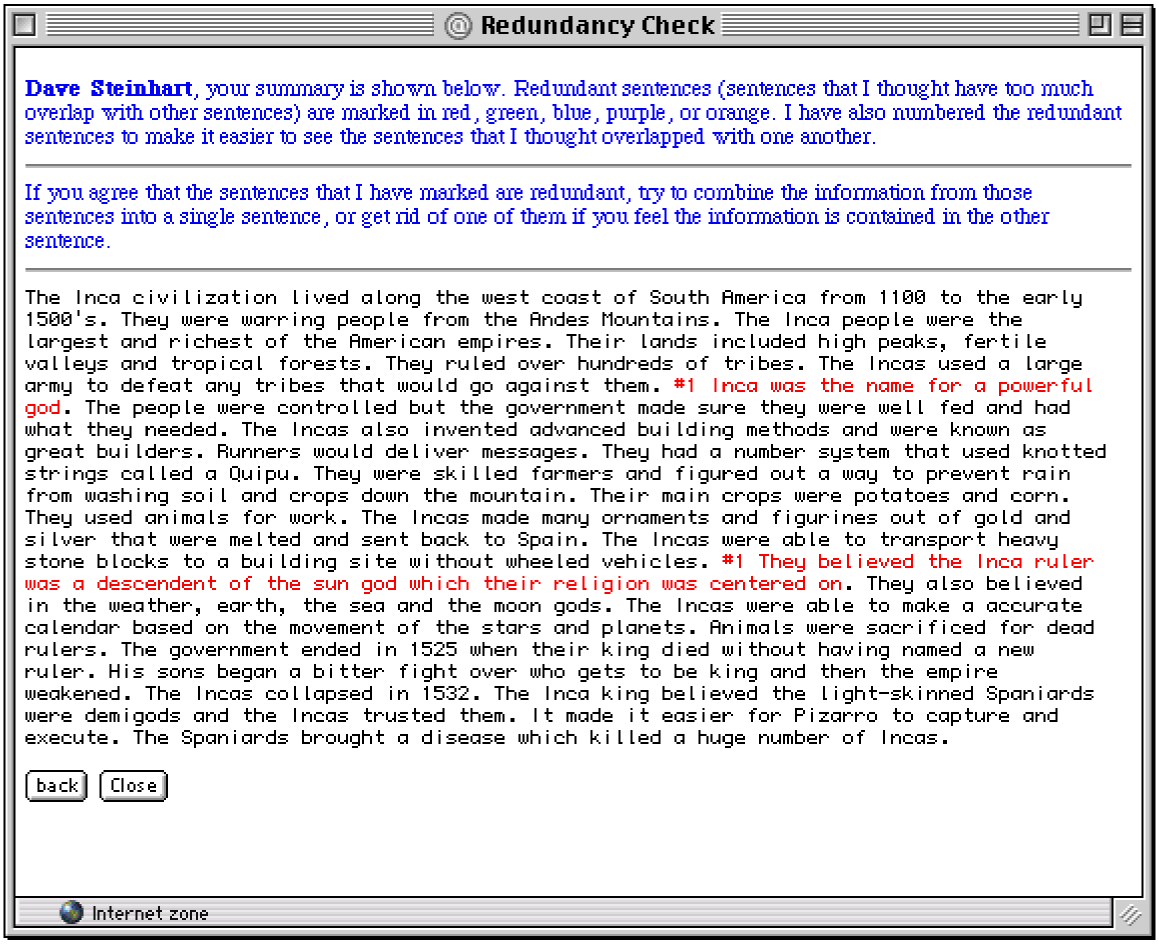


Figure 11. The redundancy check window.

"If you're bad you go to hell." With the exception of good / bad and heaven / hell, all of the words in the two sentences overlap. Further, the cosine between "good" and "bad" is 0.65, and the cosine between "heaven" and "hell" is 0.57. Therefore it is not surprising that the cosine between the two sentences is 0.80 (these cosines were computed in the Mesoamerican space, rather than in the *TASA-all* space), and yet they are not redundant.

As with the redundancy check, the relevance check is only available after the student's summary has covered all of the sections of the text. The relevance check operates on a sentence by sentence basis, computing a "relevance index," and flagging any sentence whose index falls below the threshold (see Figure 12).

The relevance index consists of the sum of the maximum cosine between the sentence and each of the sections of the text and the vector length of the sentence. When trying to assess relevance, both the cosine and the vector length are important. Considering only the vector length is equally

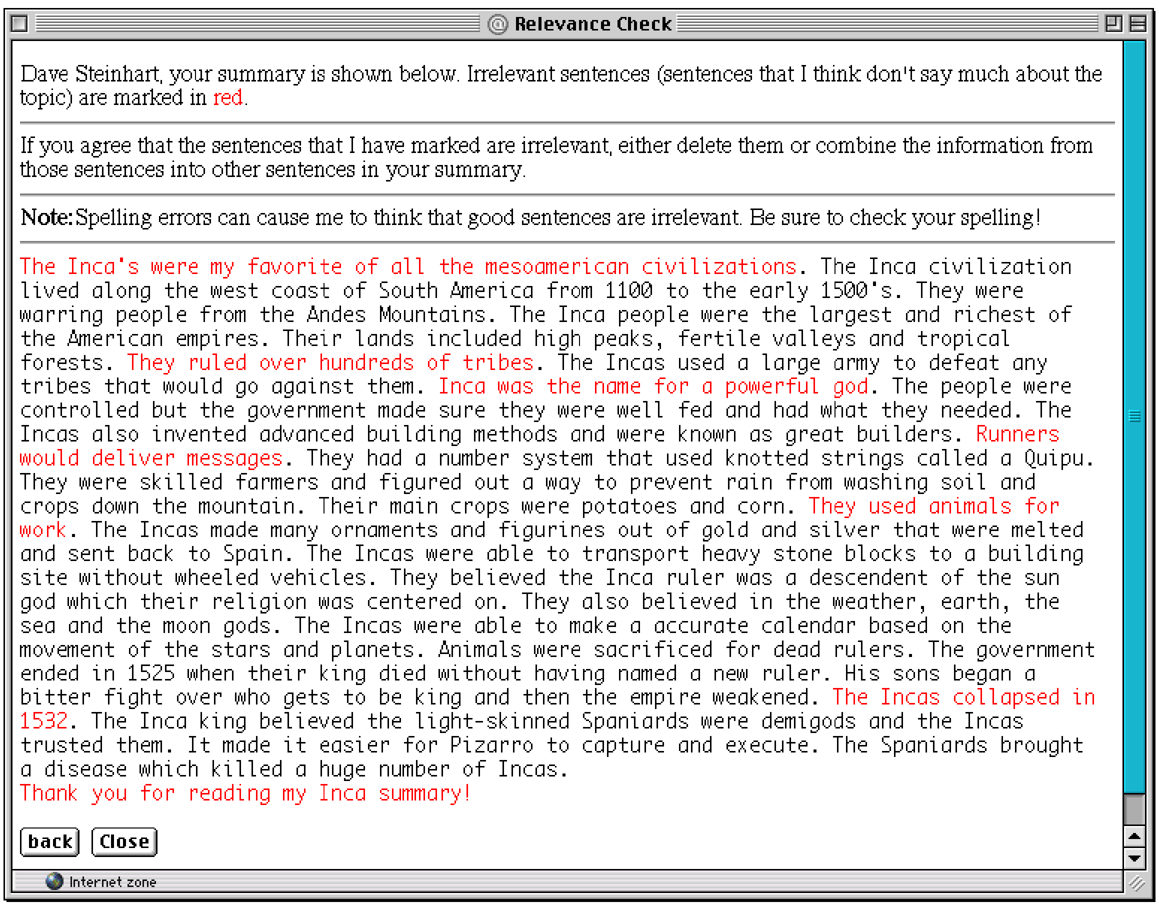


Figure 12. The relevance check window.

problematic (or perhaps worse) because the vector length is effectively a measure of the amount of information contained in the passage, but does not at all measure the semantic content of the passage. Therefore, the vector for a long off-topic sentence will be longer than that of a short on-topic sentence[27].

The relevance threshold that is currently used in the system was determined in an ad hoc fashion, and a better method is needed. After several failed attempts using other methods, the threshold was determined by examining summaries from previous studies and choosing a threshold that resulted in a majority of irrelevant sentences being flagged. The threshold computation involves the sum of the vector length and the cosine, but should be implemented using the dot product, which is the numerator of the cosine computation: $COS_{a,b} = \frac{a \bullet b}{\|a\| \times \|b\|}$.

[27] As an example, the sentence "Solar energy is good for the environment" has a vector length of 0.29, while the sentence "As a practical method for the characterization of word meaning, we know that LSA produces measures of word-word, word-passage and passage-passage relations that are well correlated with several human cognitive phenomena involving association or semantic similarity" (Landauer, Foltz, and Laham, 1998) has a vector length of 0.46.

CHAPTER 3:

METHODS, RESULTS, AND POST HOC ANALYSES

INTRODUCTION

In order to obtain formative evaluation with external validity, Summary Street (SS) was incorporated into the curriculum of two team-taught 6th grade classes at Platt Middle School (Boulder, Colorado) and its efficacy was ascertained in a series of three classroom trials. The teachers involved with these trials had previously used State the Essence, and they, as well as their former students, had provided invaluable feedback that guided the design of SS. These trials and the results obtained from them are described in detail in the sections that follow. In addition, this chapter discusses individual differences between texts as well as students, and concludes with a discussion of new analysis techniques that improve the quality of feedback returned by SS.

STUDY 1: ANCIENT CIVILIZATIONS

As part of their learning activities for a unit on Ancient Civilizations, fifty-two students were required to write summaries on the Maya, Aztec, and Inca cultures. The students wrote one of the summaries using SS, and the other two by hand or using a traditional word processor. Rough drafts were prepared as homework and then the summaries were revised in class.

The teachers scored the summaries on a 10-point scale, but were not blind as to experimental condition. The results, shown in Figure 13, show our first significant advantage for SS—students who used SS to compose their summaries on the Inca civilization fared significantly better than those who used traditional means to compose their summaries ($t_{50} = 2.47$, $p = .02$).

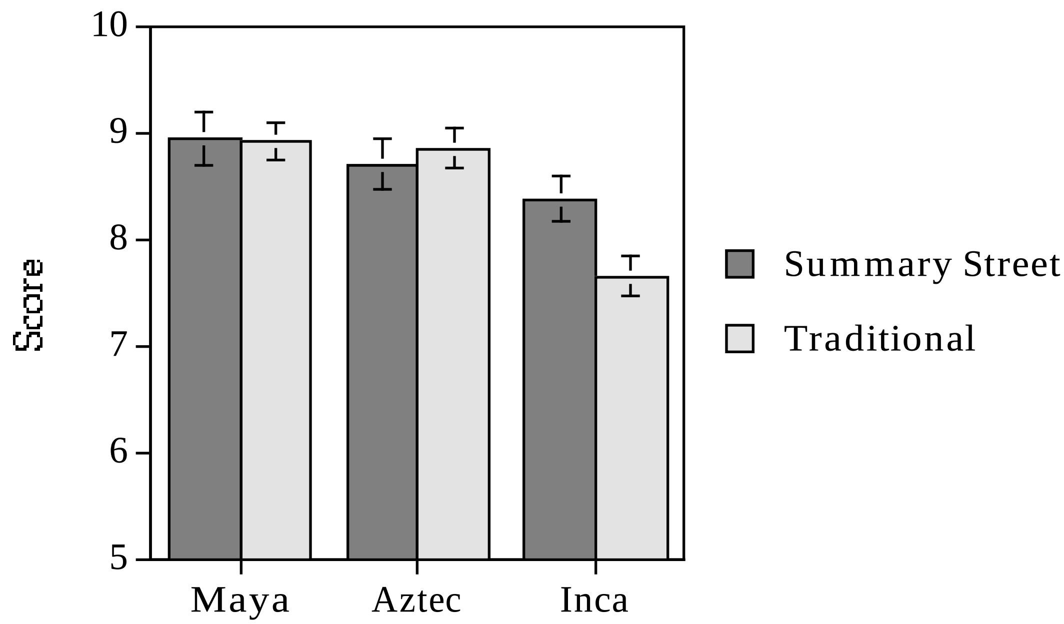


Figure 13. Results of Ancient Civilizations Trial

For the Maya and Aztec summaries, there were no significant mean differences in quality related to how the summaries were written. However, both the teachers and the students considered the Inca text to be the most difficult of the three (confirmed by the lower mean grades these summaries received—8.02 vs. 8.94 and 8.79 for the Aztec and Maya, respectively), and we conjectured that SS may be more beneficial when the text being summarized is difficult. That is, for easy or moderately easy texts, perhaps the students are able to write a reasonable summary on their own, but when the difficulty increases, they need assistance.

STUDY 2: THE HUMAN CIRCULATORY SYSTEM

The second trial took place in April of 1999 with the same students. This time the curriculum dealt with the human circulatory system, and the students each read and summarized two texts—one on the operation of the lungs, and one on the operation of the heart. At the end of the unit, a post-test was administered to gauge the students' knowledge of the circulatory system. Teachers scored the summaries on a 5-point scale, and this time they were unaware of the identity of the students and whether or not SS had been used to compose the summary. The results of this trial, shown in Figure 14, follow a similar pattern to those of the Ancient Civilizations trial. Overall, lung summaries received significantly lower grades than heart summaries (t_{100} = 3.61, $p < .001$), confirming our belief that the lung text was the more difficult of the two. Because the lung text was considerably shorter than the heart text, the length constraint for the lung summary was shorter, which made it more difficult for the students to summarize. In spite of this, lung summaries composed with SS received significantly higher grades than those composed by traditional means (t_{50}= 2.33, $p = .02$). There were no statistically significant differences in the quality of the heart summaries (t_{48} = 1.26, $p = .21$), regardless of how they were written, and no statistically significant

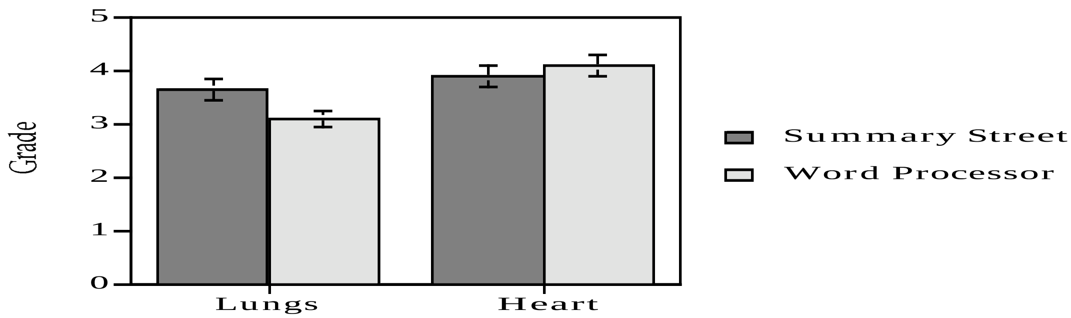


Figure 14. Results of the Human Circulatory System trial.

differences in post-test grades across text (t_{100} = 0.21, p = .83) or condition (t_{100} = .05, p = .96).

These two trials seem to suggest that SS is more helpful when students are required to summarize more difficult texts or when some other aspect of the assignment, such as length constraint, is especially difficult.

STUDY 3: SOURCES OF ENERGY

The third and final trial, completed in October of 1999, took place during an instructional unit about various sources of energy. Fifty-two sixth-grade students learned about ten energy sources (five renewable and five non-renewable) and the teachers devised five pairs of energy sources (one renewable and one non-renewable); the students then chose one of the pairs of energy sources to summarize. The students were therefore required to write two summaries about two different energy sources. The summaries were produced one week apart, and in order to control the experimental environment, the rough drafts were composed on paper during a 45-minute class period (a rough draft is the sort of task that would generally be carried out at home, but it was important to ensure that it was the student, rather than a parent, who was doing the writing). The rough drafts were revised using SS the following day[28]. In order to examine the efficacy of the section-based and redundancy / relevance feedback, a restricted version of SS was created which offered length feedback via the vertical length indicator, but did not give section-based or redundancy / relevance feedback During each session, half the students used the standard SS, while the other half used the restricted SS. Having two versions of SS eliminated variability due to the method of composition of the summary and also allowed us to collect time-on-task data for both groups.

For simplicity, the students who used the restricted version of SS will be referred to as having "received no feedback" or having been in the "No-feedback" condition. Likewise, those students who used the standard SS will be referred to as having "received feedback," or having been in the "Feedback condition. The groups were counterbalanced so that if a students did not receive feedback during the first week, then that student received feedback during the second week, and vice versa.

---

[28] After the rough drafts were written, they were entered into SS by Eileen Kintsch and the present author. This was done due to large variance in typing speed on the part of the students and to maximize the amount of time the students were allowed to revise. No spelling or other corrections were performed during the transcription.

The summaries were once again blind-scored by the teachers, but this time the teachers assigned two scores. The first, a quality score on a scale of 0-5, was intended to judge the summary's overall quality, considering all factors—writing style, mechanics, punctuation, grammar, etc., in addition to content. After that score had been assigned, the teachers were instructed to grade the content of each section, irrespective of other factors. Content was scored on a 0-2 scale, where 2 = adequate section coverage, 1 = some section coverage, but not quite adequate, and 0 = no section coverage. In addition to those data, the time that students spent revising their summaries was measured.

TIME ON TASK

The first result, shown in Figure 15, was that students who received feedback spent more than twice as much time revising their summaries than those who did not receive feedback (M = 72 vs. 33 minutes, t_{43}= 5.88, $p < .0001$[29]). The marked difference in time spent revising is not surprising given the observations that were made during this trial. In general, students who did not receive feedback tended to work only until their summaries were of the appropriate length. Without external motivation, they did not revise further. On the other hand, students who received feedback were forced to revise because they knew they were not done until SS told them they had covered all of the sections of the source text. The value of the increased time on task is discussed in the next chapter.

---

[29] Some students failed to write one or both summaries due to illness, and time on task data was not available for several students, which is why df=43.

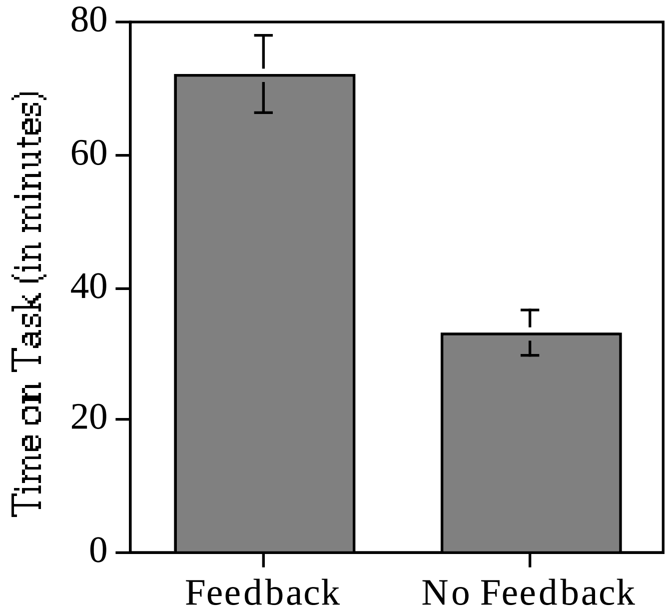


Figure 15. Time on task by condition.

Content versus quality scores

Given that SS provides feedback about a summary's content rather than its quality (that is, it does not provide feedback about grammar, mechanics, or other aspects which would fall under the heading of "quality"), our primary concern was whether or not SS feedback helped students improve the content of their summaries. However, during previous trials, the teachers informed us that judging a summary's content irrespective of other quality factors was a difficult task. This is why the teachers were asked to provide a quality score in addition to the content score—it was hoped that their content scoring would not be influenced by other factors in the students' summaries.

SS feedback did indeed have a significant effect on the content of student summaries as shown in Figure 16. Those summaries composed with feedback had significantly higher content scores than summaries composed without feedback, (M=1.29 vs. 1.01, t_{45} = 3.91, p = .0003)[30].

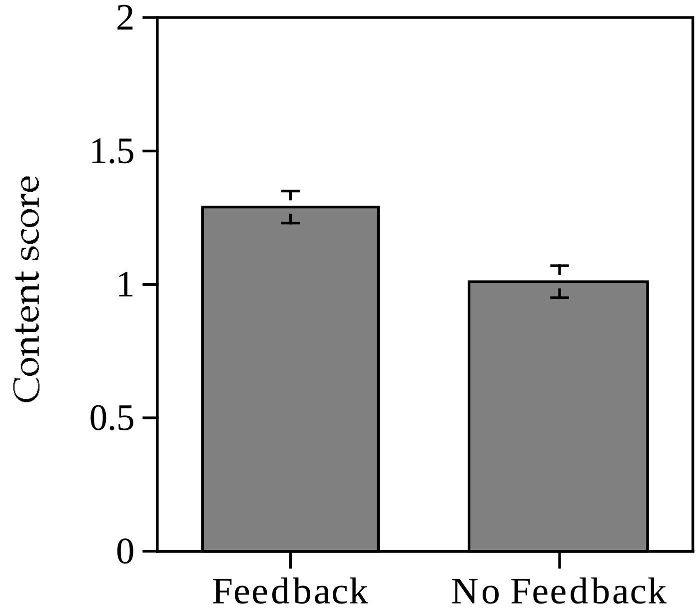


Figure 16. Mean Content Score by Condition.

In addition to the content difference, there was a small but significant quality difference—summaries produced with feedback had a higher average quality score than those produced without (M = 3.19 vs. 2.87, t_{45} = 2.16, p = .036).

In order to determine whether this quality difference was distinct from the content difference, a repeated-measures ANCOVA was performed on the quality scores, using the content scores as the covariate[31]. After controlling for

---

[30] Even though the teachers provided content scores for *each* section of the text, the scores reported herein represent the average content score for all sections of the text.

[31] Actually, in accordance with Judd, Kenny, & McClelland (2001), the ANCOVA was performed on the quality and content difference scores between the feedback and no-feedback conditions.

content, the quality difference was no longer significant ($t_{44} = 0.36$, $p = .72$), and therefore the quality difference will be ignored. Some possible reasons for the high correlation between quality and content scores will be discussed in the next chapter.

Text Difficulty

The interaction of feedback and text difficulty was examined in two ways. First, the improvement in content scores as a result of receiving feedback was computed and termed the *feedback effect*. (That is, how much more content was included in students' summaries as a result of receiving feedback from SS?) Second, the degree to which the similarity between student summaries and the texts (i.e., the cosine between the summary and the text) improved from first to final draft was examined, and termed the *feedback effect on revision*. (That is, how much did SS feedback help the students' summaries become more like the text?)

For the first analysis, the average content scores of summaries produced without feedback was used as the measure of the difficulty of each text. This measure of text difficulty was plotted against the feedback effect, and the result is shown in Figure 17. There is a significant negative correlation ($r = -0.83$, $p = .003$) between the difficulty of the text and the improvement shown by students who received feedback. That is, for easy topics (topics for which students perform well without feedback), SS feedback does not help much (for one topic, biomass, students actually performed worse with feedback, but not significantly so), whereas for difficult topics, SS feedback greatly helps content scores. These results, as well as those from the

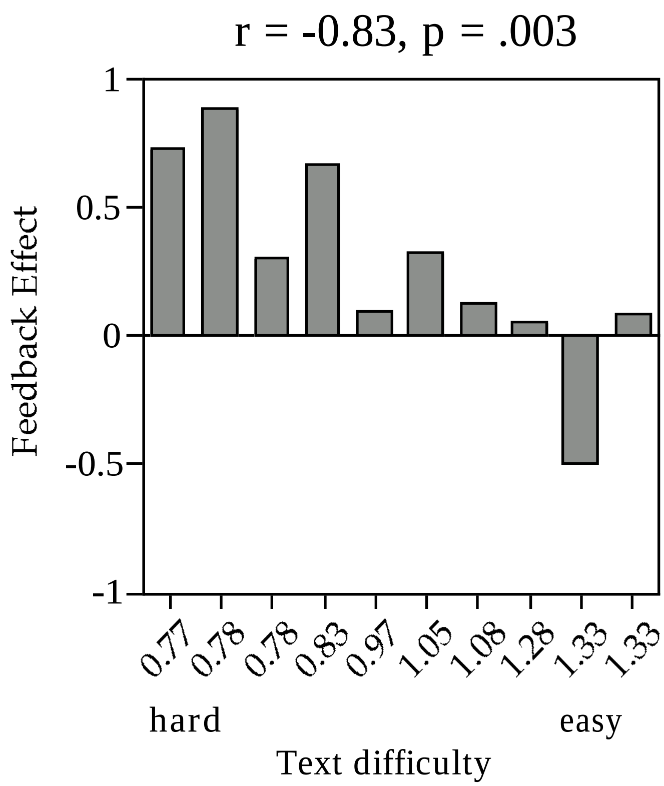


Figure 17. Feedback effect (improvement in content score for students receiving feedback) vs. text difficulty (average content score of students summarizing this topic without feedback).

preceding studies thus support the hypothesis that SS is more helpful for difficult texts.

Having identified that the feedback from SS has a positive effect on summary content, the next topic of interest was whether or not SS had an effect on revision. That is, does feedback from SS help students include more of the text content in their summaries, and does that effect differ depending on the difficulty of the topic being summarized?

The terms *first cosine* and *final cosine* will be used to refer to the cosine between a student's first draft and the text being summarized, and the cosine between the student's final draft and the text being summarized, respectively.

Next, the term *cosine gain* will be used to refer to the difference between the *first* and *final* cosines. Finally, the phrase *feedback effect on revision* will be used to refer to the difference in cosine gain between the No-feedback and Feedback conditions. That is, for a given text, how much larger was the cosine gain for the summaries written with feedback versus those written without feedback?

Figure 18 plots the feedback effect on revision versus topic difficulty (once again measured by the average score attained by students who summarized that topic without feedback). This time the correlation is non-significant, but the pattern is the same—for difficult topics, the feedback from SS has a larger effect on revision.

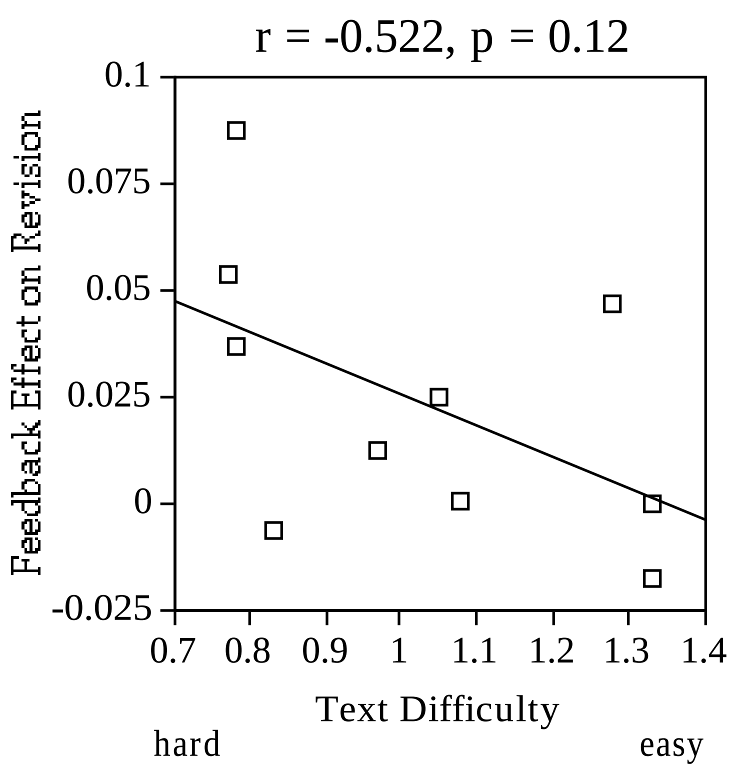


Figure 18. Text Difficulty (average content score without feedback) vs. Feedback Effect on Revision (the difference in cosine gain between the Feedback and No-Feedback conditions).

Thus far, text difficulty has been measured subjectively, using the content grades assigned by the teachers. The analysis was repeated using an objective measure of text difficulty—the average cosine between the summaries of students who summarized the text without feedback and the text being summarized. Presumably, the more difficult the text is, the more difficult it is for students to summarize, and on average, the cosine between the summaries of no-feedback students and the text will be lower.

Figure 19 shows this re-analysis of the feedback effect on revision versus topic difficulty, and once again the pattern is the same. But this time, the correlation is highly significant ($r = -.748$, $p = .013$).

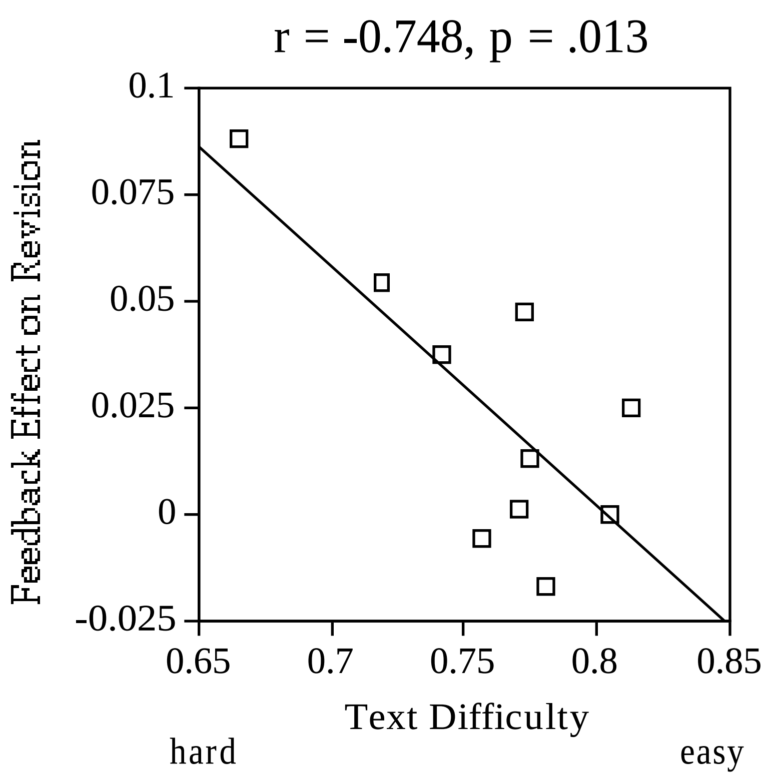


Figure 19. Text difficulty (as measured by NF cosine) vs. Feedback Effect on Revision (difference in improvement from first to last summary between F and NF condition).

Given that Figure 18 and Figure 19 exhibit the same trend, and that the correlation between content grades and cosines was high ($r = 0.70$, $p = .02$), we may reasonably conclude that both teacher judgment and the average no-feedback cosine as described above are alternative measures of text difficulty, and can be helpful in determining if the texts are appropriate for students.

TRANSFER

Transfer was examined by looking at the change in content scores from the first week to the second week. For simplicity, the students who were in the No-Feedback group during the first week will be called the NF→F group (indicating that they did not receive feedback during the first week, but the did receive feedback during the second week). The other group will be termed the F→NF group.

Figure 20 shows the change in content score for both groups. During the first week, the F→NF students performed significantly better than the NF→F students ($t_{44} = 2.55$, $p = .014$), and their performance did not decline during the second week, when they did not receive feedback ($t_{22}=1.13$, $p =.27$). During the second week, the NF→F students performed significantly better that they had during the first week ($t_{22} = 4.62$, $p = .0001$). However, it is not clear whether the increases seen in the second week are due to SS-specific transfer, or a generalized practice effect (or both).

In order to investigate further, the degree to which student summaries changed within—rather than across—revision sessions was measured. One way of doing this would be to examine the change in content scores from first

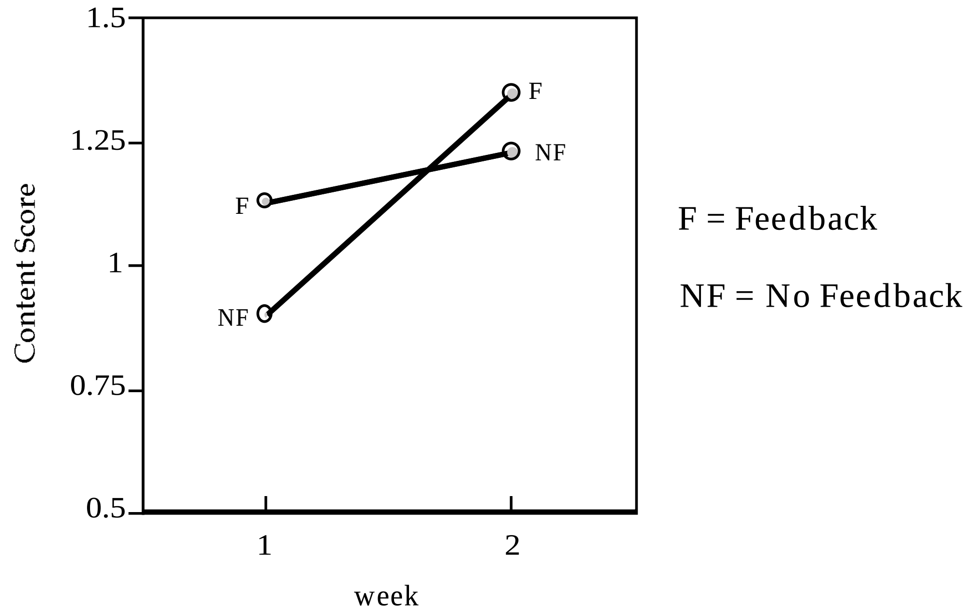


Figure 20. Change in content score from first to second week.

to final draft. However, because the teachers did not grade the students' rough drafts, the cosine gain was used instead of teacher grades.Recall that the cosine gain indicates the degree to which the student's summary became "more like the text" during the course of the revision session. In addition, the cosine between the student's first and final drafts was computed, in order to determine the amount of revision performed by each student.

Most striking was that across both groups and both revision sessions, the overall average cosine between first and final drafts was 0.94 (that is, the first and final drafts are nearly identical, on average), indicating that the amount of revision performed by the students in both groups was extremely small. Thus, the cosine gain data (omitted here) are uninformative as to the issue of transfer and practice effects. This issue will be revisited in the next chapter.

Individual Differences

In order to investigate whether SS feedback had a different effect for students of differing verbal ability, the teachers were asked to rate the verbal ability of the students on a scale of 1-3 (low / medium / high)[32], and the data from Study 3 were analyzed for these three groups, and the results are shown in Figure 21. While it appears that students of average verbal ability profit the most from SS feedback, a repeated-measures ANOVA on these data indicate no group differences ($F_{2,33} = 1.99$, $p = .15$) and therefore there is no evidence that students of differing verbal ability profit differentially from SS feedback.

---

[32] When the study was first planned, we had not considered analyzing the data by verbal ability, so we did not obtain permission from the parents to use the students' standardized test scores.

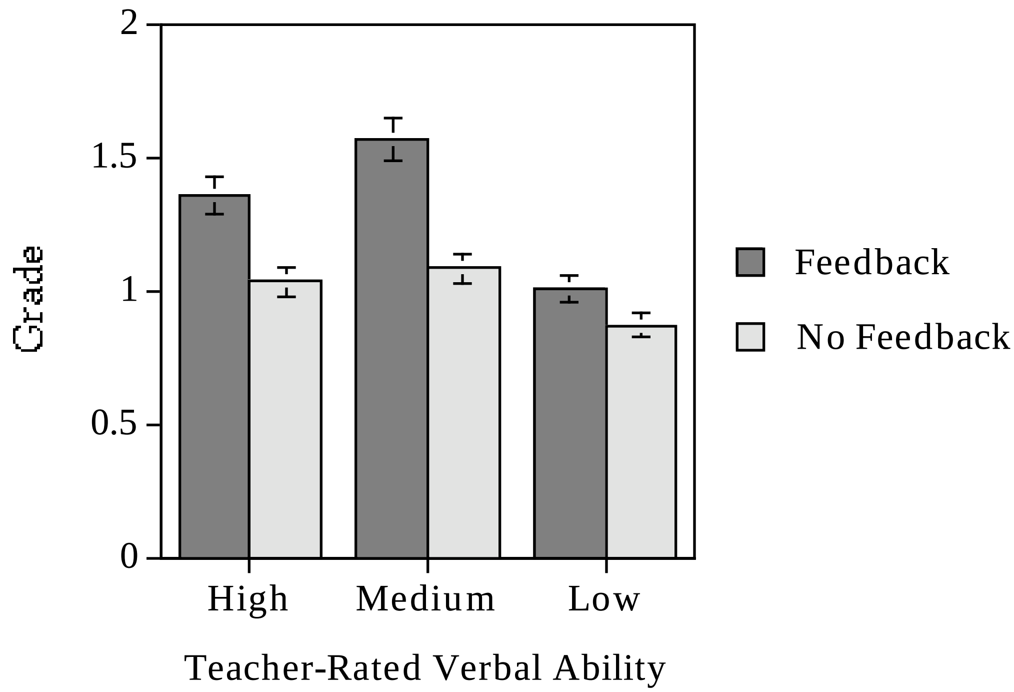


Figure 21. Effect of Summary Street feedback as a function of students' verbal ability in the Sources of Energy study.

A Further Look at Threshold / Coverage Computation

In the preceding section it was noted that determining thresholds (and therefore, section coverage) solely based upon the cosine between the student's summary and the section being considered is problematic. What is needed is a threshold that is sensitive to length as well as content coverage. The obvious candidate measure to include is the vector length, mentioned previously as being a measure of semantic content.

Incorporating the vector length into the computation should alleviate the problem whereby a small amount of text is erroneously considered adequate coverage for a section. However, the difficulty with incorporating the vector length into the computation lies in identifying which portions of the summary correspond to which sections of the text, so that the vector length of that portion may be computed, rather than computing the vector length of the entire summary.

Currently, in order to determine how well the student has covered a section of the text, SS computes the cosine between the *entire* summary and each section of the text under consideration. While it is true that this cosine can be "diluted" by portions of the summary that have nothing to do with the section in question, in practice this is not a major problem, as discussed below.

One method that could be used to determine the portion of the student's summary that pertains to each section of the text (so that its associated vector length can then be computed) is as follows. First, the student's summary is parsed into sentences. Next, for each sentence of the summary, the cosine between the sentence and each section of the text is computed. The section of the text having the maximum cosine with the sentence is considered to be the section to which the sentence corresponds[33] (because the sentence is more similar to that section than to any of the other sections). Sentences that do not seem to correspond to any of the sections (i.e., sentences whose cosine with every section is below some threshold, just as with the relevance check) are discarded[34]. The group of sentences

---

[33] Note that in this scheme we are ignoring the certain possibility that a sentence in the student's summary refers to more than one section.

[34] The relevance threshold mentioned here is actually more liberal than the one used by SS for the relevance check.

corresponding to a given section are considered to be the portion of the student's summary which corresponds to that particular section. Finally, the vector length of that portion of the student's summary which is about a given section, as well as the cosine between that portion of the summary and the text of that section are computed. To avoid confusion, the cosine computed here will be referred to as the *partial* cosine.

ASSESSING CONTENT COVERAGE

In order to determine the effectiveness of these proposed methods, they were used to predict the content grades assigned by one of our teachers, and compared with the "simple" cosine (the method currently used by SS) and the dot product (defined on page 48). The results are shown in Table 1.

Some comments are in order here. Even though most of the other measures did better, the "simple" cosine, which is currently used by SS to determine section coverage, performs reasonably well. However, the partial cosine—the cosine between that portion of the student's summary which corresponds to the section in question and the text of that section—represents a considerable improvement over the simple cosine. In addition, the vector length performed quite well on its own. However, given that the vector length does not measure content, using just vector length alone would be a risky proposition, especially if students were to find out that SS was essentially measuring the amount of writing about a given topic as opposed to whether the writing was on-topic!

PREDICTOR	r
Cosine	.43
Partial Cosine	.55
Vector Length (VL)	.52
Partial Cosine + VL	.57
Dot Product[35]	.39

Table 1. LK's grades predicted by various measures.

Finally, note that using *both* the partial cosine and vector length did not result in much improvement over the partial cosine alone. Nevertheless, a possible solution would be to compute thresholds composed of separate length (vector length) and coverage (cosine) components, requiring the summary to have an appropriate amount of information from each section as well as being on-topic. One possible way to devise the length component of the threshold would be to require that the portion of the student's summary that pertains to a given section be some reasonable percentage of the overall vector length of the section. For example, if the assignment is to summarize a 1500 word text in approximately 300 words (or 20% of the original summary length), and the first of four sections of the to-be-summarized text is 420

[35] Because the dot product is unduly affected by section length, this value actually represents an average of six correlations. That is, the dot product between each student summary and each section of the text being summarized was computed. Then, correlations between LK's grades and the dot products were computed per section (0.36, 0.26, 0.23, 0.50, 0.36, 0.66) and the average of those was reported. The section six correlation represents considerably fewer data points than the others because only two of the texts have six sections. However, all the texts have at least four sections, so the section four correlation was computed over all of the students.

words (28% of the total text length) and has a vector length of 2.57, then we might reasonably expect the student to write approximately 84 words (28% of 300) on the first section, with a vector length of 28% of 2.57, or 0.72.

In order to implement this solution, the graphical interface of SS will have to be modified to accommodate the two-part thresholds. One possible way to do this would be to indicate each section's coverage using a clock face with a single hand[36]. The degree to which the student's writing accurately summarized the section in question would be indicated by distance of the hand from the 12 o'clock position (i.e., the higher the cosine, the closer the hand would be to the 12 o'clock position), while the amount of information about the section would be indicated by the length of the hand. Having two separate indicators would make it clearer to students how much they have written versus how much of the important content material they have covered.

[36] This idea was suggested by Libby Landauer.

CHAPTER 4:

GENERAL DISCUSSION

Brown and her colleagues (Brown and Smiley, 1978; Brown and Day, 1983; Brown, Day, and Jones, 1983) have shown summarization to be a difficult task for children, as well as a gradually emerging skill requiring years of practice to perfect. Others have shown that students' summarization skills can be improved through training both for normal children (Brown, Campione, and Day, 1981; Cunningham, 1982; Hare and Borchardt, 1984), as well as children with learning disabilities (Gajria and Salvia, 1992; Palincsar and Brown, 1984).

In addition, summarization is also an efficient way for students to develop a thorough understanding of unfamiliar material. Baumann (1984) has shown that summarization improves comprehension, and Taylor (1982) has shown that summarization improves recall of text. Furthermore, summarization provides a more accurate view of students' understanding (or lack thereof) as compared to multiple choice ("multiple guess") or short answer exams. To paraphrase Palinscar and Brown (1984), if you are not able to summarize a passage, then you have not understood it.

Taking into account this body of work, as well as the opinions of the teachers with whom we worked, it seems clear that summarization skills are important and should be taught in our schools and cultivated in our children.

The potential for summarization is great, but the amount of work involved in grading multiple drafts is excessive. Especially when one considers that according to many educators, students should be writing five to ten times more than they do currently.

As a result, students do not typically receive enough practice with extended writing in school. The solution was to design Summary Street (SS) specifically to address the problem of a lack of extended writing practice. SS provides an environment in which students can prepare multiple drafts of a summary and receive immediate, automatic feedback, thereby enabling them to revise their summaries on their own. The result of using SS is that students are able to engage in a great deal of writing and revision before their teachers see the final product, ultimately allowing teachers to assign considerably more summary writing than they currently do.

Discussion of experiments and results

Given that we were primarily concerned with external validity, actual classroom studies were crucial. The experiments described herein examined the efficacy of SS using a sixth-grade curriculum in a sixth-grade classroom. Study 1, the Ancient Civilizations study, was our first field trial of SS. That study demonstrated a significant advantage for students who used SS when summarizing the Inca text, which was the most difficult text used in the study. All students, regardless of whether or not they used SS, performed considerably better on their summaries of the Aztec and Maya texts, but there were no differences in the quality of summaries produced with and without

SS. The results suggest a ceiling effect, which we believe is due to the ease with which the students were able to summarize the Aztec and Maya texts. When the task proved difficult, SS helped students write better summaries, but when it was easy, students did not need the help of SS (or presumably, teachers, parents, or peers).

Study 2 was similar to Study 1, except that students summarized two texts and were given a post-test at the end of the instructional unit. This study produced the same pattern of results. There was a significant advantage for students who summarized the more difficult Lung text using SS, while there were no differences in summaries of the easier Heart text. The post-test showed no differences in knowledge across texts as well as conditions.

Study 3 differed considerably from the others. As in Study 2, the students wrote two summaries, one with feedback, and one without, counterbalanced with respect to order. All students composed rough drafts in class, rather than at home. Instead of using a word processor or pencil and paper, the No Feedback students used a restricted version of SS which supplied length feedback, but no content feedback. Using this restricted version of SS facilitated accurate measurements of time on task for both groups, and removed variance due to differences in the manner in which students wrote their summaries.

In addition, the scoring was different in Study 3. When scoring for previous studies, the teachers told us of their difficulty in evaluating a summary's content, independent of style, mechanics, and other writing measures. This difficulty was reflected in the fact that their content and quality measures were highly correlated. In an attempt to make the grading task easier, the teachers were asked to produce an overall quality score first, and then follow up with content scores for each of the sections. The teachers indicated that this method made it easier for them to assign content scores, having already said their piece with respect to the overall quality scores.

The pattern of results from Studies 1 and 2 was repeated, in that SS once again helped the students with more difficult texts. This time, there were ten different texts, rather than two or three, and the effect of SS was nicely tied to text difficulty.

Additional results were more striking. Students who received feedback spent more than twice as much time revising their summaries as compared to those who did not receive feedback.

The students who received feedback spent more time revising their summaries, on the average, than they would have spent on other classroom activities scheduled by their teachers. Indeed, they spent more time writing and revising their summaries using SS than they would have by other means. Of course, increased time on task is not all there is to the story. Students who received feedback received significantly higher content scores than those who did not.

Not only was there a content effect as a result of feedback from SS, we also noticed an effect on revision, especially for the more difficult topics. Specifically, for difficult topics, the feedback from SS had a greater effect on revision than it does for easier topics. This result supports the earlier conclusion that for easy topics, students perform well on their own and do not need the support from SS.

The results from these three studies are promising, and suggest that SS can be a beneficial tool for students and teachers alike. The sections that follow further discuss our classroom experiences with SS, individual differences, transfer, the role of technology in the classroom, and finally, some of the possible future directions of this research.

Reflections on Summary Street in the Classroom

Several of our earlier trials with State the Essence (StE) were marred by technical difficulties. In addition, many students found StE confusing and difficult to use because the student was presented with too much feedback at once. SS owes its existence to those early trials and the helpful feedback and suggestions received from both students and teachers.

During the development and testing process, SS was heavily modified in response to this feedback, and as a result both teachers and students found SS easy to use. Consistent with its goals, SS gave students immediate feedback and a great deal of practice with a difficult writing task without teachers having to coach them through successive drafts. This was clearly a benefit for both teachers and students, and the students produced better summaries than they otherwise would have produced (and in some cases, thought they could produce).

Not only did students produce better summaries when using SS, they clearly enjoyed the process and spent much more time reviewing the source text than they would have had they worked alone[37]. In addition, Study 3 showed that students who used SS spent more than twice as much time revising their summaries compared to those students who did not use SS. While we have no direct evidence whether their content learning was improved, the extra time on task certainly cannot hurt. Indeed, Hayes & Nash (1996), who reviewed correlational studies of planning in writing, write: “Our analysis of these studies indicates that text quality is strongly and positively related to time-on-task.” (p. 53)

In addition to more time on task, the students using SS were more motivated than those who did not. Even though SS is accessed via a web browser, the students refrained from surfing the web while working on their summaries, and the vast majority of students remained focused on the task for the entirety of the revision periods, which ranged in length from 75 to 110 minutes[38]. This increased focus may have been due in part to the fact that we supervised the students during the trial, but we believe it was due to SS, and that motivation is a crucial part of getting students to write longer and better. Hayes and Nash (1996) agree, arguing that researchers need to consider motivation much more than they have in the past, in order to better understand how to keep students focused on writing tasks.

---

[37] While we did not measure “enjoyment” empirically, we did observe the students using SS, and in addition, we interviewed them after the study was completed. The students overwhelmingly preferred using SS to prepare and revise their summaries compared to the traditional methods of pen and paper or word processor.

[38] This was no small feat given that our teachers generally do not attempt to engage the students in tasks lasting longer than 45 minutes due to students’ limited attention span.

Nowhere was the increased motivation more impressive than for the several children with mild learning disabilities who participated in our studies. Learning disabled students are often excluded from mainstream activities due to their deficits in attention and ability, behavior problems, and perhaps worst of all, self-doubt. In one case, a student with attentional and behavioral problems worked for an entire revision period, and while he did not finish, he greatly improved his original summary and covered all but one of the sections of the text. When I approached him in order to observe his progress, he gleefully told me that he had never been able to do this kind of assignment before (which was verified by his teacher). Seeing this student succeed where he had otherwise failed, solely due to his interactions with SS, was truly heartwarming.

The stories of several other children with mild learning disabilities who participated in our studies are similar. The experiences of these students give us good reason to believe that SS can provide the structure and guidance needed for mildly learning disabled students to succeed when it comes to writing summaries. It is apparent that SS can be a tool for a broad range of students.

QUALITY VERSUS CONTENT REVISITED

In Study 3, the teachers were asked to provide two scores, a quality score, considering all factors, and a content score, which was intended to judge content independent of any other factors. Even though the teachers believed it was easier to provide both scores, their quality and content scores were highly correlated. In fact, after controlling for content scores, the quality difference disappeared.

There are two possible explanations for the inability of the teachers to separate quality and content. Either quality and content are two different aspects of a summary, but teachers are still having difficulty rating them independently, or quality and content are interdependent and inseparable.

If quality and content are in fact different, then what accounts for the teachers inability to tease apart them apart? Perhaps because quality and content generally appear together, teachers have a difficult time separating them when they are different. If this is the case, a possible solution might be to come up with more fine-grained measurement techniques, in order to ensure that all non-content aspects of quality can be assessed.

But what if content and quality are inseparable? While one can certainly conceive of a well-written summary having little or no content (i.e., a well-written summary about the wrong topic), or a poorly written summary having sufficient content, in practice, it seems that such summaries rarely occur. What we tend to see are well-written summaries that have the appropriate content and poorly-written summaries that do not, suggesting that a student who is able to provide the appropriate content is likely to write well, and vice versa.

Transfer

An important question regarding SS concerns whether or not its effects transfer to later summaries written by students. The results of Study 3 are suggestive, but by no means conclusive. Those students who received feedback during the first week's trial performed as well during the second week, when they did not receive feedback, but this effect could be due solely to practice. Indeed, both groups performed better during the second week, so we cannot ignore the possibility that some of the gains were simply due to using the interface a second time (with or without feedback).

Because we were not concerned with transfer when the study was designed, we did not ask the teachers to grade the students' first drafts. So instead we considered the cosine gain (the difference between the final cosine and the first cosine) between each student's first and final summary in order to look for transfer. Unfortunately, these data did not clarify the situation and instead suggested that students were benefiting from practice as well as feedback. However, by far the biggest problem was that the amount of revision performed by the students was minimal. The students who performed the most revision were those who received feedback during the first week; however, the average cosine between their first and final summaries was 0.91. Overall, the average cosine between students' first and final summaries was 0.94, indicating that on average, students were not performing substantive revisions.

HOW TO BETTER ASSESS TRANSFER? Given that Study 3 was not designed to assess transfer, it is not altogether surprising that we failed to find any clear indications of it. One obvious suggestion for further studies would be to have NF→NF and F→F groups in addition to the NF→F and F→NF groups. These additional groups should be able to give some insight into the effect of SS feedback independent of any practice effects. In addition it would be helpful to devise a study such that the students revise considerably more than they did in our studies.

What about revision? If the students who received feedback were spending an average of 72 minutes revising their summaries, and yet the amount of revision was minimal, what were they doing? One of the problems with students of this age is that they have not yet learned how to revise effectively. SS was intended to give them the opportunity to revise, but was not designed to teach students the art of revision. This is clearly a problem that will need to be addressed before SS can be seamlessly integrated into a teacher's curriculum. One solution would be for the students to use SS one or two times, then have the teacher spend time teaching the students to revise effectively, and then give the students unlimited access to SS to allow them to revise their future assignments.

The promise of transfer. One reason to be hopeful about transfer can be found in the work of Schreiner (in preparation). Her participants wrote three summaries of three short stories, and were given up to fifteen minutes to revise their summaries (with a maximum of three revisions beyond the original). Participants were placed in one of three conditions (a fourth condition, the "paper" condition was included in her study, but it is omitted here in the interest of brevity). In the first condition, the "no feedback" condition, participants simply typed their summaries into the computer, receiving no feedback in the process. In the second condition, termed the "gist" condition, subjects received a 0-10 score on their summaries. This score was derived from the LSA cosine between the summary and text being summarized, which was then mapped onto the 0-10 (continuous) scale using a linear function. In the third, or “full feedback” condition, the participants received the following feedback in addition to the point score received by the gist participants: (1) the sentence of their summary that was most related to the original text (i.e., the sentence having the highest cosine with the original text); (2) the sentence of their summary that was least related to the original text (i.e., the sentence having the lowest cosine with the original text); (3) a list of redundant sentences in their summary (measured in the same way as SS, except that the threshold was derived from LK’s grades rather than from the text itself).

The summaries of the participants in the no feedback condition progressively worsened (although not significantly), presumably due to boredom or inability to focus. On the other hand, those participants who received feedback, even the impoverished feedback of the gist condition, improved slightly over their three summaries. Even though their improvement was not significant, by the third trial they were performing significantly better than those participants who did not receive feedback. This suggests perhaps there is a motivational issue involved in their performance, in that the presence of feedback kept the participants focused on the task for a longer period of time.

These results argue that feedback, even simple feedback, contributes to improved summary writing over time. Furthermore, the more feedback one receives, the better one’s performance becomes on successive summaries. It might therefore be the case that while SS does not always help students write better summaries “in the moment,” it can, over time, help them become better summary writers.

Individual Differences

In addition to the text differences already noted, one might expect that SS is not "one size fits all" when it comes to individual differences in student ability. In the same way that students profit less from SS feedback when summarizing easy texts, one might expect better students to profit less from SS feedback because they are less likely to need the support that poorer students generally need.

During the development and experimental planning phase, we did not anticipate that individual differences in verbal ability would be a factor in terms of SS feedback. That is, we expected that SS feedback would help most if not all students equally, but did not consider any possible differential effect of feedback. When we decided to investigate individual differences post hoc, we relied upon the teachers' *low/medium/high* verbal ability ratings to divide the students into three groups.

While it appeared that SS feedback was most helpful for the *medium* kids in Study 3, these results were not statistically reliable. However, there is still reason to believe there may be individual differences in students' ability to profit from SS feedback. For one thing, the number of students of low verbal ability was small compared to the size of the other groups. But more importantly, the overall range of verbal ability was rather limited. If the study had been designed to consider individual differences from the outset, we may well have seen statistically significant differences that are only suggested by the present data.

It is therefore not unreasonable to consider how one might explain a differential effect of SS feedback, assuming it were found in a future study which was better designed. First of all, we would expect the pattern to match the results of Study 3—namely, that the medium students profit the most from SS feedback. This pattern is consistent with the assertion that if all the students in the classroom are summarizing the same text, then it is reasonable to expect that the low kids may have difficulty understanding the text and/or feedback, while the high kids are likely to be able to produce a quality summary on their own. That leaves the medium kids, who can understand the feedback and therefore understand what they need to do to improve their summaries.

The important assumption in this scenario is that all students are summarizing the same text. But that text might not be appropriate for all the students, especially if their verbal abilities vary considerably. One way to level the playing field would be to let the low students summarize easier texts, while the high students summarize more difficult texts. Assuming the low students' difficulty is with the text and not with the SS feedback, having each kid summarize a text that is at his or her level could result in all the students profiting equally from SS feedback. The low students would be able to work with a text that is at their level, while the high students would potentially be able to tackle a text which might normally be beyond their grasp.

FURTHER COMMENTS REGARDING SS AND THE ITS PARADIGM

Even though it was far from evident while SS was under development, the flaws in the redundancy and relevance checks turned out to be beneficial. We had made the mistake of assuming that SS had to be perfect (i.e., they had to agree as closely as possible with human judgments of redundancy and relevance), when in fact perfection, in addition to being an unattainable goal, is also detrimental to students who use the system. If the redundancy and relevance checks were perfect then there would be nothing left for the student to do, other than removing the irrelevant sentences, and removing or combining the redundant sentences. Instead, the redundancy and relevance checks make suggestions, some of which should be followed, and some of which should be ignored. Therefore the students are given some guidance by the system, but forced to decide on their own which sentences are redundant or irrelevant.

The crucial lesson here is that there is a danger in creating tutors that are too intelligent, lest the student fail to think and instead merely follow the advice of the tutor. The student must be required to think critically, evaluate the advice given by the tutor, and be an active participant in the process. Indeed, Nathan (1998) suggests that "Relegating diagnosis to the tutoring environment may implicitly teach students that evaluation and reflection upon a solution is not considered relevant to the problem-solving process." (p. 137). We must therefore strike a delicate balance between intelligence and support or encouragement, in much the same way a human tutor does. Even though human tutors generally possess all the knowledge that the tutee is trying to acquire, they do not simply blurt out the information, robbing the tutee of the process of discovery.

Of course, in the case of SS, it does not know unequivocally which sentences are redundant (relevant) and which are not. However, it can make reasonable guesses, and let the student be the final arbiter of what is and is not redundant (relevant). One might argue that SS might already be too intelligent. Rather than suggesting sentences that are redundant or irrelevant, the student might be required to think more critically if SS simply flagged sentences that it identified as problematic, without elaborating further. Arguably, this could be a problem for the poorer students, so perhaps the level of elaboration should be adjustable by the teacher in order to best fit the student using the system.

Individual Differences Revisited

Like students, not all texts are created equal. The three studies presented in the preceding chapter show SS to be most effective with difficult texts. That is, we assume that for easier texts, students have little or no difficulty producing an adequate summary on their own, and therefore do not need feedback from SS. When the texts are difficult, students profit the most from the SS feedback, but it is possible that the relationship is not linear. While the degree of difficulty of the texts in the Study 3 varied, none of the texts was overly difficult—they were all within the grasp of the majority of the 6^{th} graders who participated in our studies. Had more difficult texts been used, it is possible that the effectiveness of SS would have decreased. That is, it might be the case that SS works best with texts that are neither too easy nor too difficult, but just right.

This notion has a basis in the work of Kintsch (1994), who argued that in order to learn from a text, readers must connect information in the text with their prior knowledge. Kintsch further argued that in order to optimize learning from a text, readers need texts that are neither too difficult nor too easy, but instead are just right. Wolfe, Schreiner, Rehder, Laham, Foltz, Kintsch, and Landauer (1998) built on Kintsch's work and argued that learning is maximized within a "zone-of-learnability," which occurs when texts are just beyond the knowledge base of the reader, but not too far. This idea is based on Vygotsky's (1978) theory of cognitive development in children. He argued that learning should be matched to the developmental level of the child. He defined the "zone of proximal development" as the distance between the child's actual developmental level (as defined by what the child can perform on his own) and potential developmental level (as defined by what the child can perform given the help of adults or more capable peers).

It is therefore necessary for the child to interact with external mediators (peers or adults) who facilitate the emergence of a skill in the child. The crucial part of the theory is that the degree to which the child masters the skill with the help of external mediators exceeds what the child could develop on his or her own.

We might therefore liken SS to an external mediator with regards to the skill of summarization. Of course we would not argue that the skill would fail to emerge without SS. Instead, we would argue that summarization is mastered more quickly through the use of SS, and that SS provides the support, feedback, and extended practice necessary for mastery.

FUTURE DIRECTIONS

In spite of the fact that the studies discussed here showed SS to be an effective tool for bettering students' summarization skills, it clearly has some distance to go before it could be considered a standalone system that teachers could use without assistance. This section describes the modifications to SS that are needed in this regard.

THRESHOLDS. Most important are the threshold and section coverage computations. As noted earlier, we strove to develop an automated technique for setting the section thresholds and were somewhat successful, but further work is needed. The use of two thresholds, one for content coverage and one for length coverage (or a single threshold combining both measures, e.g., the dot product) as discussed in the previous section, may prove to be the solution. By determining the source section of each sentence in the student's summary, SS would be able to do a significantly better job determining how much the student has written about each section of the text. The system described in the previous section is no doubt less than ideal, but it is sufficient in that it will alleviate the problem of uneven coverage that exists in the current system.

SEMANTIC SPACES. Another important issue that needs to be resolved concerns the semantic spaces being used by SS. This has been an ongoing source of frustration and concern with the system. Of the three studies discussed in the previous chapter, only the Sources of Energy study was conducted using the general knowledge space. For the other two studies (Human Circulatory System and Mesoamerican Civilizations), specialized spaces had to be built. Building a new semantic space, however, is not the main problem, even though it can often entail a considerable amount of effort. The problem is that while specialized spaces contain the necessary knowledge to enable SS to evaluate a summary about a specialized topic, they tend to contain too little general knowledge. Specialized spaces are tantamount to a grader who is an expert on a topic but does not possess general, real-world knowledge. On the other hand, the general knowledge space contains too little specialized knowledge, tantamount to a grader who possesses general knowledge, but lacks deep knowledge of the topic at hand.

One potential solution to this problem of coverage is to combine two or more spaces into a single space. Numerous attempts to augment a specialized space by adding documents from the general knowledge space[39] have been unsuccessful, perhaps because the addition of documents "dilutes" the knowledge of the specialized space. An alternative possibility, which is being explored by the LSA Research Group, is to *rotate* one semantic space onto another, with the goal being to preserve the semantic knowledge in both spaces, rather than diluting it, which is the case when the documents from the two spaces are simply combined.

---

[39] In pursuit of this problem, I attempted to create an augmented Mesoamerican space by adding random, semantically close, or semantically distant documents from the TASA-all space, and in every case, the new space proved to be worse than the Mesoamerican space on its own.

Additional Features. Several features have been suggested in the course of developing SS that are not included in the current version, but could potentially be included in a future version of the system. Perhaps most important is a coherence check, which would attempt to identify points in the summary where coherence breaks down. This check would operate by computing the cosine between each pair of neighboring sentences and flagging points where the inter-sentence cosine drops. We should see a drop at paragraph boundaries, but within paragraphs, we would expect that each successive sentence is related to the one that preceded it.

Organizational Check. Another useful feature would be an organizational check. We expect a properly organized student summary to follow the organization of the text (expert summaries, on the other hand, are less likely to mirror the organization of the original text, given that experts tend to synthesize information and generalize across several sections). Therefore, this feature would operate by computing the cosine between each sentence of the summary and each section of the text in order to determine which section the sentence is about (in the same way as the alternative threshold computation described earlier). If, for example, the summary starts out with several sentences summarizing the first section of the text, then continues on to summarize the rest of the text and concludes with a sentence that is once again about the first section of the text, it might have organizational problems. Even if the summary does not follow the organization of the text, we would expect there to be topic clusters present within the summary. Of course, just as with the redundancy and relevance checks, SS need not be perfect. If only some of the perceived organizational problems are genuine, they can still be beneficial to the student, and as is the case with the redundancy and relevance checks, the student should be the final arbiter of correctness.

Grammar Check. SS does not currently evaluate grammar or language mechanics. While a complete grammar checker is clearly beyond the scope of SS, a rudimentary grammar checker could possibly be built using the statistical technique of bigram probabilities. That is, given a large corpus of English, the probability of any two words (a bigram) appearing next to one another can be computed. Once those bigram probabilities are known, potentially ungrammatical sentences can be identified because they will contain bigrams that are either not found in English or are found with low probability. Any sentences containing such bigrams could be flagged as potential grammatical errors. The additional computation required for such a check would be not be excessive; depending on the length of the summary, it would be comparable to that required for the existing redundancy and relevance checks.

Plagiarism Check. State the Essence and an early version of SS contained a plagiarism check, but the teachers indicated that it was less of a priority for them, and it was removed. A plagiarism check could operate in one of two fashions. The most obvious way to check for plagiarism would be to look for word overlap between the student's summary and the text being summarized. This method could be computationally expensive in that the entire text must be scanned each time a section of the student's summary is to be checked for plagiarism.

Another way to implement a plagiarism check would be on a sentence by sentence basis. Any sentences having a high cosine with a sentence of the text could be flagged as potential plagiarism. The problem with this method is that, as we have seen, it is possible for the cosine between two sentences to be quite high, when in fact the sentences are neither redundant, nor, in this case, plagiarized.

IMPROVED SPELLING CHECKER. One final potential feature that has been discussed is a spelling checker that makes suggestions as opposed to simply flagging misspelled words. This was a suggestion we heard countless times when interviewing students after the studies were completed. While we want to avoid having SS do too much for the student, there is a definite need for a spelling checker that can make suggestions. In our experience, when SS flags a word as misspelled, students are often stymied and must therefore ask their teacher (or, as was the case during our studies, one of us) how to correctly spell the misspelled word. If SS is eventually to be used by students on their own, without adult guidance, it will definitely need a better spelling checker. One possible solution would be to use LSA to guide the suggestion process[40]. That is, rather than suggesting words that are orthographically similar to the misspelled word, the spell checker could scan the semantic space and suggest words that have a high cosine with the other words in the sentence. One can certainly imagine sentences for which this method would fail, but the method has not yet been tested, so we have yet to find out how well it will work in general.

---

[40] This method was suggested by Missy Schreiner.

Conclusion

Summarization is a difficult skill to master, but one that clearly enhances learning, especially for novel and difficult material. SS can help students become better summarizers without increasing the load on our already overburdened teachers. SS gives students the opportunity for extended writing practice, while providing fully automatic feedback that enables students to improve the content of their summaries. In addition, SS motivates students to work harder and keeps them focused on writing and revising longer than traditional means of composition, ultimately resulting in better quality writing. As a result, teachers can give out more writing assignments, students receive more writing practice, and ultimately become better writers.

Of course, several issues have yet to be resolved. In our classroom trials so far, students engaged in relatively minor revisions. Those students who used SS wrote better summaries than those who did not, but the revisions performed by the students who used SS were clearly not substantive. Arguably, this has more to do with the young age of the students and their minimal amount of writing experience than it does with SS, but it is an issue that will need to be explored further.

Additionally, the issue of transfer and/or practice effects must be resolved through future studies. While we have good reason to believe that SS will contribute to better summary writing over time, more evidence is needed if SS is to be adopted on a wide scale. The fact that SS generates enthusiasm among students, while at the same time helping them produce better summaries is compelling, and time on task is certainly an important factor which affects writing quality, but it behooves us to determine more precisely what the benefits (and costs) of SS are.

The effects of text difficulty should also be addressed. We believe that text difficulty can be manipulated to compensate for individual differences in verbal ability, but do not currently have evidence to support this assertion. It certainly seems intuitive that students who struggle with texts that are beyond their ability will not be helped much by SS. However, given easier texts, we would expect that these students would be working at a level whereby SS feedback would be instructive.

Likewise, students whose abilities are well beyond the texts being summarized by the rest of the class will succeed independent of SS. If these students are given a chance to summarize more difficult texts, SS could potentially bridge the gap and enable them to summarize texts that might otherwise lie beyond their grasp.

Because it is based on Latent Semantic Analysis (LSA), SS is considerably simpler than the traditional Intelligent Tutoring System (ITS). However, the simplicity of SS is misleading. Even though the SS architecture deviates considerably from that of a traditional ITS, SS is able to perform natural language comprehension by evaluating arbitrary summary content, as well as provide users with detailed, individualized feedback. Furthermore, these features of SS are performed without complicated student and expert modules which are difficult to build, and not easily extensible. Instead, the intelligence of SS stems from LSA, enabling SS to accurately assess the knowledge contained in student summaries by comparing them to the text being summarized.

The use of LSA as a foundation for an ITS clearly holds great promise, because LSA-based tutors can be created relatively quickly and efficiently, even when new semantic spaces must be built. Instead of complex expert modules which must be rebuilt for each new application, LSA provides a general purpose “semantic engine” which can be utilized in a wide variety of tutoring applications.

Having successfully demonstrated the efficacy of SS, the future of SS looks bright. In its current state, SS is not yet ready for deployment in classrooms across the country, but once the outstanding problems are solved, there is good reason to believe SS can and will be an important instructional tool. In addition to the instructional uses outlined here, SS could also be used for assessment within a school or district, enabling students to demonstrate proficiency in summarization in order to supplement the current battery of assessment tools.

References

Anderson, R.C., & Biddle, W.B. (1975). On asking people questions about what they are reading. In G. Bower (Ed.), *Psychology of learning and motivation, Vol. 9.* New York: Academic Press.

Anderson, J.R., Boyle, C.F., & Reiser, B.J. (1985). Intelligent tutoring systems. *Science*, 228, 456-462.

Anderson, J.R., Farrell, R., & Sauers, R. (1984). Learning to program in LISP. *Cognitive Science*, 8, 87-129.

Baumann, J.F. (1984). The effectiveness of a direct instruction paradigm for teaching main idea comprehension. *Reading Research Quarterly*, 20, 93-115

Bean, T.W., & Steenwyk, F.L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Reading Behavior,* 16(4), 297-306.

Bloom, B.S. (1984). The 2 Sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher,* 13(6), 4-16.

Bonar, J.G., & Cunningham, R. (1988). Bridge: Tutoring the programming process. In J. Psotka, L.D. Massey, & S.A. Mutter (Eds.), *Intelligent tutoring systems: Lessons learned.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A.L., Campione, J.C., & Day, J.D. (1981). Learning to learn: On training students to learn from texts. *Educational Researcher,* 10, 14-21.

Brown, A.L., & Day, J.D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior,* 22(1), 1-14.

Brown, A.L., Day, J.D., & Jones, R.S. (1983). The development of plans for summarizing texts. *Child Development,* 54, 968-979.

Brown, A.L., & Smiley, S.S. (1977). Rating the importance of structural units of prose passages: A problem of metacognitive development. *Child Development,* 48, 1-8.

Brown, A.L., Smiley, S.S. (1978). The effects of experience on the selection of suitable retrieval cues for studying texts. *Child Development*, 49, 829-835.

Carbonell, J.G. (1970). AI in CAI: An artificial intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems*, 11, 190-202.

Corbett, A., Koedinger, K.R., & Anderson, J.R. (1997). Intelligent tutoring systems. In M. Helander, T. K Landauer, P. Prabhu (Eds.), *Handbook of human-computer interaction 2nd Edition*. New York: Elsevier.

Cunningham, J.W. (1982). Generating interactions between schemata and text. In J.A. Niles & L.A. Harris (Eds.), *New inquiries in reading research and instruction*. Rochester, NY: National Reading Conference.

Fox, A. (1982). Quiz time in the classroom: Spelling Bee. *Creative Computing*, 8(10), 18-20.

Gajria, M., & Salvia, J. (1992). The effects of summarization instruction on text comprehension of students with learning disabilities. *Exceptional Children*, 58(6), 508-516.

Hare, V.C., & Borchardt, K.M. (1984). Direct instruction of summarization skills. *Reading Research Quarterly*, 20, 62-78.

Hayes, J.R., & Nash, J.G. (1996). On the nature of planning in writing. In C.M. Levy & S. Ransdell (Eds.), *The Science of Writing: Theories, Methods, Individual Differences, and Applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Jitendra, A.K., Cole, C.L., Hoppes, M.K., & Wilson, B. (1998). Effects of a direct instruction main idea summarization program and self-monitoring on reading comprehension of middle school students with learning disabilities. *Reading & Writing Quarterly*, 14(4), 379-396.

Judd, C.M., Kenny, D.A., & McClelland, G.H. (2001). Estimating and testing mediation and moderation in within-participant designs. *Psychological Methods*, in press.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, 49, 294-303.

Kintsch, E., Steinhart, D., Stahl, G., Matthews, C., Lamb, R., & the LSA Research Group (2001). Developing Summarization Skills through the Use of LSA-Based Feedback. *Interactive Learning Environments*, 8(2), 87-109.

Koschmann, T. (1996). Paradigm shifts and instructional technology: An introduction. In T. Koschmann (Ed.), *CSCL: Theory and practice.* Mahwah, NJ: Lawrence Erlbaum Associates.

Kuhn, T.S. (1972). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: Chicago University Press.

Kurland, L.C., & Tenney, Y.J. (1988). Issues in developing an intelligent tutor for a real-world domain: Training in radar mechanics. In J. Psotka, L.D. Massey, & S.A. Mutter (Eds.), *Intelligent tutoring systems: Lessons learned.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Landauer, T.K., & Dumais, S. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review,* 104, 211-240.

Lepper, M., Woolverton, M., Mumme, D., & Gurtner, J. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S. Lajoie & S. Derry (Eds.), *Computers as cognitive tools*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Littman, D., & Soloway, E. (1988). Evaluating ITSs: The cognitive science perspective. In M.C. Polson & J.J. Richardson (Eds.), *Foundations of intelligent tutoring systems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Malone, L.D, & Mastropieri, M.A. (1992). Reading comprehension instruction: Summarization and self-monitoring training for students with learning disabilities. *Exceptional Children*, 58(3), 270-279.

McCann, P.H. (1975). *Training mathematics skills with games.* Navy Personnel Research and Development Center: San Diego, CA.

McNeil, J., & Donant, L. (1982). Summarization strategy for improving reading comprehension. In J.A. Niles & L.A. Harris (Eds.), *New inquiries in reading research and instruction.* Rochester, NY: National Reading Conference.

Nakatani, L.H., Egan, D.E., Ruedisueli, L.W., Hawley, P.M., & Lewart, D.K. (1986). *TNT: A talking tutor 'n' trainer for teaching use of interactive computer systems*. CHI '86 Proceedings.

Nathan, M.J. (1998). Knowledge and situational feedback in a learning environment for algebra story problem solving. *Interactive Learning Environments,* 5, 135-159.

Nathan, M.J., Kintsch, W., and Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction*, 9(4), 329-389.

Neilson, I., Thomas, R., Smeaton, C., Slater, A., & Chand, G. (1996). Education 2000: Implications of W3 Technology. *Computers and Education*, 26(1-3), 113-122.

Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension fostering and comprehension-monitoring activities. *Cognition and Instruction*, 2, 117-175.

Pressey, S.L. (1927). A machine for automatic teaching of drill material. *School and Society*, 25, 549-552.

Pylyshyn, Z. (1989). Computing in cognitive science. In M. Posner (Ed.), *Foundations of cognitive science.* Cambridge, MA: MIT Press.

Reis, R., & Leone, P.E. (1987). Teaching reading and study skills to mildly handicapped learners. *The Pointer,* 31(2), 41-43.

Reusser, K. (1993). Tutoring systems and pedagogical theory. In S. Lajoie & S. Derry (Eds.), *Computers as cognitive tools.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Rinehart, S.D., Stahl, S.A., and Erickson, L.G. (1986). Some effects of summarization training on reading and studying. *Reading Research Quarterly*, 21(4), 422-437.

Rinehart, S.D., & Thomas, K.F. (1993). Summarization ability and text recall by novice studiers. *Reading Research and Instruction*, 32(4), 24-32.

Ross, S.M., & DiVesta, F.J. (1976). Oral summary as a review strategy for enhancing recall of textual material. *Journal of Educational Psychology*, 68(6), 689-695.

Schreiner, M.E. (2001). *The Role of Automatic Feedback in the Summarization of Narrative Text.* Unpublished doctoral dissertation, University of Colorado, Boulder.

Shute, V.J., & Regian, J.W. (1993). Principles for evaluating intelligent tutoring systems. Journal of Artificial Intelligence and Education, 4(2-3), 245-271.

Slavin, R.E. (1983). *Cooperative learning.* New York: Longman.

Stevens, A.L., & Collins, A.M. (1977). The structure of a Socratic tutor. Proceedings of the SAMPO Conference. New York: Association for Computing Machinery.

Taylor, B.M. (1982). Text structure and children’s comprehension and memory for expository material. *Journal of Educational Psychology*, 74, 323-340.

Taylor, B.M., & Beach, R.W. (1984). The effect of text structure instruction on middle-grade students' comprehension and production of expository text. *Reading Research Quarterly*, 19, 132-146.

van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Vygotsky, L. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.

Weisberg, R., & Balajthy, E. (1990). Development of disabled readers' metacomprehension ability through summarization training using expository text: Results of three studies. *Reading, Writing, and Learning Disabilities*, 6, 117-136.

Wolfe, M. B., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). Learning from text: Matching readers and text by Latent Semantic Analysis. *Discourse Processes*, 25, 309-336.

Woolf, B.P., & McDonald, D.D. (1984). Building a computer tutor: Design issues. *IEEE Computer*, 17, 61-73.

APPENDIX A

TEXTS FROM THE SOURCES OF ENERGY STUDY

BIOMASS

WHAT IS BIOMASS?

Biomass is any organic matter--wood, crops, seaweed, animal wastes--that can be used as an energy source. Biomass is probably our oldest source of energy. For thousands of years, people have burned wood to heat their homes and cook their food.

Biomass gets its energy from the sun. All organic matter contains stored energy from the sun. During a process called photosynthesis, sunlight gives plants the energy they need to convert water, carbon dioxide, and minerals into oxygen and sugars. The sugars, called carbohydrates, supply plants (or the animals that eat plants) with energy. Foods rich in carbohydrates (like spaghetti) are a good source of energy for the human body!

Biomass is a renewable energy source because its supplies are not limited. We can always grow trees and crops, and people will always produce garbage.

Using Biomass Energy

Usually we burn wood and use its energy for heating. Burning, though, is not the only way to convert biomass energy into a usable energy source. There are four ways:

*Burning***.** We can burn biomass in special plants to produce steam for making electricity, or we can burn it to provide heat for industries and homes.

Bacterial Decay. Bacteria feed on dead plants and animals, producing a gas called methane. This is a natural process that happens whenever waste decays. Methane is the same thing as natural gas, the gas sold by natural gas utilities.

Fermentation. Adding a yeast to biomass produces an alcohol called ethanol. This is how wine, beer, and liquor are made. Wine is just fermented grape juice.

Conversion. Biomass can be converted into gas or liquid fuels by using chemicals or heat. In India, cow manure is converted to methane gas to produce electricity. Methane gas can also be converted to methanol, a liquid form of methane.

Types of Biomass

We use four types of biomass today: 1) wood and agricultural products; 2) solid waste; 3) landfill gas; and 4) alcohol fuels.

Wood and Agricultural Biomass. Most biomass used today is home grown energy. Wood-logs, chips, bark, and sawdust-accounts for about 79 percent of biomass energy. But any organic matter can produce biomass energy. Other biomass sources include agricultural waste products like fruit pits and corn cobs.

Solid Waste. There is nothing new about people burning trash. What's new is burning trash to generate electricity. This turns waste into a usable form of energy. A ton (2,000 pounds) of garbage contains about as much heat energy, as pounds of coal.

Power plants that burn garbage for energy are called waste-to-energy plants. These plants generate electricity much as coal-fired plants do except that garbage-not coal-is the fuel used to fire an industrial boiler. Making electricity from garbage costs more than making it from coal and other energy sources. The main advantage of burning solid waste is it reduces the amount of garbage dumped in landfills by 60 to 90 percent, and reduces the cost of landfill disposal.

Landfill Gas. Bacteria and fungi are not picky eaters. They eat dead plants and animals, causing them to rot or decay. Even though this natural process is slowed in the artificial environment of a landfill, a substance called methane gas is still produced as the waste decays.

New regulations require landfills to collect methane gas for safety and environmental reasons. Methane gas is colorless and odorless, but it is not harmless. The gas can cause fires or explosions if it seeps into nearby homes and is ignited.

Landfills can collect the methane gas, purify it, and then use it as an energy source. Methane, which is the same thing as natural gas, is a good energy source. Most gas furnaces and gas stoves use methane supplied by natural gas utility companies. The city landfill in Florence, Alabama recovers 32 million cubic feet of methane gas a day. The city purifies the gas and then pumps it into natural gas pipelines.

Today only a tiny portion of landfill gas is used to provide energy. Most is burned off at the landfill. Why? With today's low natural gas prices, this higher-priced "biogas" has a hard time competing.

Alcohol Fuels. Wheat, corn, and other crops can be converted into a variety of liquid fuels including ethanol and methanol.

Using ethanol as a motor fuel is nothing new. Its use is almost as old as the automobile. In the early 20th century, automobile pioneer Henry Ford advocated using gasohol, a mixture of ethanol and gasoline, to run his cars.

Today ethanol is a high cost fuel and its use has become a controversial issue. It is estimated that a barrel of oil will have to more than double in price before ethanol can compete with gasoline as a transportation fuel.

In spite of this, the ethanol industry has continued to grow, mainly because the federal government exempts ethanol fuels from the federal highway tax. This exemption has been extended to the year 2000.

Because ethanol is expensive, and because car engines must be modified to run on pure ethanol, ethanol is usually mixed with gasoline to produce gasohol. (Cars can run on gasohol without adjustments.)

Gasohol is 10 percent ethanol and 90 percent gasoline. In 1994, 12 percent of the nation's motor fuel consisted of this ethanol and gasoline mixture. However, in some corn-growing states, gasohol use is as high as 50 percent.

Gasohol does have some advantages over gasoline. It has a higher octane rating than gasoline (provides your car with more power), and it is cleaner-burning than unleaded gasoline, with one-third less carbon monoxide emissions. Gasohol may also help reduce America's dependence on foreign oil.

USE OF BIOMASS AND THE ENVIRONMENT

Until the mid-1800s, wood gave Americans 90 percent of the energy they used. Today biomass gives us only 3.2 percent of the energy we use. Biomass was largely replaced by coal, natural gas, and petroleum.

Seventy-nine percent of the biomass we use today comes from burning wood and wood scraps - The rest of the biomass comes from crops, garbage, landfill gas, and alcohol fuels.

Who uses biomass energy? Industry is the biggest user of biomass. Seventy-seven percent of biomass is used by industry.

Homes are the next biggest users of biomass energy. About one-fifth of American homes burn wood for heating. Three percent of homes use wood as their main heating fuel.

Electric utilities also use biomass energy to produce electricity. One percent of biomass is used to make electricity. Still, biomass produces only a tiny amount of the electricity we use in this country.

Environmentally, biomass has some advantages over fossil fuels such as coal and petroleum. Biomass contains little sulfur and nitrogen, so it does not produce the pollutants that cause acid rain. Growing plants for use as biomass fuels may also help keep global warming in check. That's because plants remove carbon dioxide--one of the greenhouse gases-from the atmosphere when they grow.

COAL

WHAT IS COAL?

Coal is a fossil fuel created from the remains of plants that lived and died about 100 to 400 million years ago when parts of the earth were covered with huge swampy forests. Coal is classified as a nonrenewable energy source because it takes millions of years to form.

The energy we get from coal today comes from the energy that plants absorbed from the sun millions of years ago. All living plants store energy from the sun through a process known as photosynthesis. After the plants die, this energy is released as the plants decay. Under conditions favorable to coal formation, however, the decay process is interrupted, preventing the further release of the stored solar energy.

Millions of years ago, dead plant matter fell into the swampy water and over the years, a thick layer of dead plants lay decaying at the bottom of the swamps. Over time, the surface and climate of the earth changed, and more water and dirt washed in, halting the decay process. The weight of the top layers of water and dirt packed down the lower layers of plant matter. Under heat and pressure, this plant matter underwent chemical and physical changes, pushing out oxygen and leaving rich hydrocarbon deposits. What once had been plants gradually turned into coal.

Seams of coal--ranging in thickness from a fraction of an inch to hundreds of feet-may represent hundreds or even thousands of years of plant growth. One important coal seam, the seven-foot thick Pittsburgh seam, may represent 2,000 years of rapid plant growth. One acre of this seam contains about 14,000 tons of coal, enough to supply the electric power needs of 4,500 American homes for one year.

HISTORY OF COAL IN AMERICA

North American Indians used coal long before the first settlers arrived in the New World. Hopi Indians, who lived in what is now Arizona, used coal to bake the pottery they made from clay.

European settlers discovered coal in North America during the first half of the 1600s. They used very little coal at first. Instead, they relied on water wheels and burning wood to power colonial industries.

Coal became a powerhouse by the 1800s. People used coal to manufacture goods and to power steamships and railroad engines. By the American Civil War, people also used coal to make iron and steel. And by the end of the 1800s, people even used coal to make electricity.

When America entered the 1900s, coal was the energy mainstay for the nation's businesses and industries. Coal stayed America's number one energy source until the demand for petroleum products pushed petroleum to the front. Automobiles needed gasoline. Trains switched from coal power to diesel fuel. Even homes that used to be heated by coal turned to oil or gas furnaces instead. Coal production reached its low point in the early 1950s. Since then, coal production has steadily increased, reaching record highs again. Today coal supplies 22 percent of the nation's energy needs. Its major use today is for electricity production.

Mining, Processing, and Transporting Coal

Coal Mining. There are two ways to remove coal from the ground: surface mining and underground mining.

Surface mining is used when a coal seam is relatively close to the surface, usually within 200 feet. The first step in surface mining is to remove and store the soil and rock covering the coal (called the "overburden"). Workers use a variety of heavy equipment--draglines, power shovels, bulldozers, and front-end loaders-to expose the coal seam for mining.

After surface mining, workers replace the overburden, grade it, cover it with topsoil, and fertilize and seed the area. These steps help restore the biological balance of the area and prevent erosion. The land can then be used for croplands, wildlife habitats, recreation, or as sites for commercial development.

Although only about 32 percent of the nation's coal can be extracted by surface mining, some 63 percent of all U.S. coal is mined using this method today. Why? Because surface mining is typically much cheaper than underground mining.

Underground mining is used when the coal seam is buried several hundred feet below the surface. In underground mining, workers and machinery go down a vertical "shaft" or a slanted tunnel called a "slope" to remove the coal. Mine shafts may sink as much as 1,000 feet deep.

One underground mining method is called room-and-pillar mining. With this method, much of the coal must be left behind to support the mine's roofs and walls. Sometimes as much as half the coal is left behind in large column formations to keep the mine from collapsing.

A more efficient and safer underground mining method, called longwall mining, uses a specially shielded machine which allows a mined-out area to collapse in a controlled manner. This method is called "longwall" mining because huge blocks of coal up to several hundred feet wide can be removed.

Processing and Transporting Coal. After coal comes out of the ground, it typically goes on a conveyor belt to a preparationplant that is located at the mining site. A "prep" plant cleans and processes coal to remove dirt, rock, ash, sulfur, and other impurities. Removing the impurities increases the heating value of coal.

After the coal is mined and processed, it is ready to go to market. Transportation is a very important consideration in coal's competitiveness with other fuels because sometimes transporting the coal can cost more than mining it.

Underground pipelines can easily move petroleum and natural gas to market. But that's not so for coal. Huge trains transport most coal (almost 60 percent) for at least part of its journey to market. It is cheaper to transport coal on river barges, but this option isn't always available. Coal can also be moved by trucks and conveyors if the coal mine Is close by. Ideally, coal-fired electric power plants are built near coal mines to minimize transportation costs.

COAL RESERVES, PRODUCTION AND USE

Coal Reserves. When scientists estimate how much coal, petroleum, natural gas, or other energy sources there are in the United States, they use the term reserves. Reserves are coal deposits that can be mined using today's mining methods and technology. Experts estimate that the United States has about 265 billion tons of coal reserves. If we continue to use coal at the same rate as we do today, we will have enough coal to last 285 years. This vast amount of coal makes the United States the world leader in known coal reserves.

Where is all this coal located? Coal deposits can be found in 38 states. Montana has the most coal--about 120 billion menial tons. Other top coal states in order of known reserves are: Illinois, Wyoming, Kentucky, West Virginia, Pennsylvania, Ohio, Colorado, Texas, and Indiana. Western coal generally contains less sulfur than eastern coal (which is good for the air when coal is burned), but not always.

The federal government is by far the largest owner of the nation's coalbeds. In the west. the federal government owns 60 percent of the coal and indirectly controls another 20 percent. Coal companies must lease the land from the federal government in order to mine this coal.

Coal Production. Coal production is the amount of coal mined and taken to market. Where does mining take place in the United States? Although coal is mined in 27 states, more coal is mined in eastern states, especially coal that is taken from underground mines, than in western states. However, the West's share of total coal production has increased steadily since 1968 when it provided just five percent of U.S. production. Today the West provides 45 percent of the nation's total production.

Total U.S. production of coal reached one billion tons in 1990, an historic high. The leading coal producing states are Wyoming, Kentucky, West Virginia, Pennsylvania, and Texas.

Some coal produced in the United States is exported to other countries. Last year, foreign countries imported seven percent of all the coal produced in the U.S. The five biggest foreign markets for U.S. coal are Japan, Canada, Italy, Brazil, and Belgium.

How Coal Is Used. What do we use coal for? Electricity is the main use. Last year 88 percent of all the coal used in the United States was for electricity production. (Other energy sources used to generate electricity include nuclear power, hydropower, and natural gas.)

Another major use of coal is in iron and steelmaking. The iron industry uses coke ovens to melt iron ore. Coke, an almost pure carbon residue of coal, is used as a fuel in smelting metals. The United States has the finest coking coals in the world. These coals are shipped around the world for use in coke ovens.

Coal is also used by other industries. The paper, brick, limestone, and cement industries all use coal to make their products.

Contrary to what many people think, coal is no longer a major energy source for heating American homes or other buildings. Less than one percent of the coal produced in the U.S. today is used for heating. Coal furnaces, which were popular years ago, have largely been replaced by oil or gas furnaces or by electric heat pumps.

COAL AND THE ENVIRONMENT

When coal became an important energy source for American industry over a century ago, concern for the environment was not at the forefront of public attention. For years, smokestacks from electrical and industrial plants emitted pollution into the air. Coal mining left some land areas barren and destroyed. Automobiles, coming on strong after World War II, contributed noxious gases to the air. Eventually, as the effects of pollution became more and more noticeable, Americans decided it was time to balance the needs of industry and the environment.

Federal laws passed in the 1960s and 70s, namely the Clean Air Act and the Clean Water Act, required industries to reduce pollutants released into the air and the water. Laws also were passed that required coal companies to reclaim the land destroyed by strip mining. Since the passage of these laws, much progress has been made toward cleaning up the environment.

The coal industry's most troublesome problem today is removing *organic sulfur, a* substance that is chemically bound to coal. All fossil fuels, such as coal, petroleum, and natural gas, contain sulfur. When these fuels are burned, the organic sulfur is released into the air where it combines with oxygen to form sulfur dioxide. Sulfur dioxide is an invisible gas that has been shown to have adverse- effects on the quality of air we breathe. It also contributes to acid rain, an environmental problem that many scientists think adversely affects wildlife (especially fish) and forests.

However, the coal industry is doing something to solve this problem. One method uses "scrubbers" to remove the sulfur in coal smoke. Scrubbers are installed at coal-fired electric and industrial plants where a water and limestone mixture reacts with sulfur dioxide to form a sludge. Scrubbers eliminate up to 98 percent of the sulfur dioxide, but they are very expensive to build.

The coal industry is also concerned about the carbon dioxide that is produced when coal is burned. Carbon from burning coal reacts with air to form carbon dioxide. When carbon dioxide and other gases, such as those emitted from automobiles, accumulate in the earth's atmosphere, they form a shield that allows the sun's light and heat in, but doesn't let it out. This condition is called the greenhouse effect.

Scientists and others are concerned about the greenhouse effect because it could cause a change in the earth's climate. Some say the earth is already experiencing a warming trend due to the greenhouse effect; others are not so sure yet. While warmer weather might be appreciated by some in northern climates, it could cause drought in some areas of the world (the American grain belt, for example) and the erosion of ocean coasts due to rising sea levels in all areas.

The coal industry is currently researching ways to lower carbon dioxide emissions. But a wholesale approach will be needed to stop the greenhouse effect. This approach must look not only at the burning of fossil fuels as a problem, but also at automobile emissions, the deforestation of the world's forests, and several other possible contributors.

WHAT IS GEOTHERMAL ENERGY?

Geothermal energy comes from the heat within the earth. The word "geothermal" comes from the Greek words geo, meaning earth," and *therme,* meaning "heat." People around the world use geothermal energy to produce electricity, to heat buildings and greenhouses, and for other purposes.

The earth's core lies almost 4,000 miles beneath the earth's surface. The double-layered core is made up of very hot molten iron surrounding a solid iron center. Estimates of the temperature of the core range from 5,000 to 11,000 degrees Fahrenheit (F). Heat is continuously produced within the earth by the slow decay of radioactive particles that is natural in all rocks.

Surrounding the earth's core is the mantle, thought to be partly rock and partly magma. The mantle is about 1,800 miles thick. The outermost layer of the earth, the insulating crust, is not one continuous sheet of rock, like the shell of an egg, but is broken into pieces called plates. These slabs of continents and ocean floor drift apart and push against each other at the rate of about one inch per year in a process called continental drift.

Magma (molten rock) may come quite close to the surface where the crust has been thinned, faulted, or fractured by plate tectonics. When this near-surface heat is transferred to water, a usable form of geothermal energy is created.

Geothermal energy is called a renewable energy source because the water is replenished by rainfall, and the heat is continuously produced by the earth.

History of Geothermal Energy

Many ancient peoples, including the Romans, Chinese, and Native Americans, used hot mineral springs for bathing, cooking, and heating. Water from hot springs is now used world-wide in spas, for heating buildings, and for agricultural and industrial uses. Many people believe hot mineral springs have natural healing powers.

Using geothermal energy to produce electricity is a relatively new industry. It was initiated by a group of Italians who built an electric generator at Lardarello in 1904. Their generator was powered by the natural steam erupting from the earth.

The first attempt to develop geothermal power in the United States came in 1922 at The Geysers steam field in northern California. The project failed because the pipes and turbines of the day could not stand up to the abrasion and corrosion of the particles and impurities that were in the steam. Later, a small but successful hydrothermal plant opened at the Geysers in 1960. Today 28 plants are operating there.

Electricity is now produced from geothermal energy in 21 countries, including the United States.

Where Is Geothermal Energy Found?

What does geothermal energy look like? Some visible features of geothermal energy are volcanoes, hot springs, geysers, and fumaroles. But you cannot see most geothermal energy. Usually geothermal energy is deep underground. There may be no clues above ground to what exists below ground.

Geologists use many methods to find geothermal resources. They may study aerial photographs and geological maps. They may analyze the chemistry of local water sources and the concentration of metals in the soil. They may measure variations in gravity and magnetic fields. Yet the only way they can be sure there is a geothermal resource is by drilling wells to measure underground temperatures.

The earth is a hotbed of geothermal energy. The most active geothermal resources are usually found along major plate boundaries where earthquakes and volcanoes are concentrated. Most of the geothermal activity in the world occurs in an area known as the "Ring of Fire." The Ring of Fire rims the Pacific Ocean and is bounded by Japan, the Philippines, the Aleutian Islands, North America, Central America, and South America.

TODAY'S GEOTHERMAL ENERGY

There are four main kinds of geothermal resources: hydrothermal, geopressured, hot dry rock, and magma. Today hydrothermal resources are the only kind in wide use. The other three resources are still in the infant stages of development.

Hydrothermal resources have the common ingredients of water (hydro) and heat *(thermal).* These geothermal reservoirs of steam or hot water occur naturally where magma comes close enough to the surface to heat ground water trapped in fractured or porous rocks, or where water circulates at great depth along faults. Hydrothermal resources are used for different energy purposes depending on their temperature and how deep they are.

*Low Temperature: "Direct Use" or Heating***.** When the temperature of a hydrothermal resource is around 50F and up, it can be used directly in spas or to heat buildings, grow crops, warm fish ponds, or for other uses. Hydrothermal resources suitable for heating occur throughout the United States and in almost every country in the world. Most of the people in Iceland and over 500,000 people in France use geothermal heat for their public buildings, schools, and homes. In the United States, geothermal heat pumps are used in 45 states to heat and cool homes and buildings. Idaho, Oregon, Nevada, and some other states use geothermal energy to heat entire districts.

Heat from geothermal resources is also used to dry ceramics, lumber, vegetables, and other products.

High Temperature: Producing Electricity. When the temperature of a hydrothermal resource is around 220F and up, it can be used to generate electricity. Most electricity-producing geothermal resources have temperatures from 300 to 700F, but geothermal reservoirs can reach nearly 1,000F.

Two main types of hydrothermal resources are used to generate electricity: dry steam (vapor-dominated) reservoirs, and hot water (liquid-dominated) reservoirs.

Dry steam reservoirs are rare but highly efficient at producing electricity. The Geysers in California is the largest and best-known dry steam reservoir. Here, steam is obtained by drilling wells from 7,000 to 10,000 feet deep. In a dry steam reservoir, the natural steam is piped directly from a geothermal well to power a turbine generator. The spent steam (condensed water) can be used in the plant's cooling system and injected back into the reservoir to maintain water and pressure levels.

Hot water geothermal reservoirs are the most common type. In a liquid-dominated reservoir, the hot water has not vaporized into steam because the reservoir is saturated with water and is under pressure. To generate electricity, the hot water is piped from geothermal wells to one or more separators where the pressure is lowered and the water *flashes* into steam. The steam then propels a turbine generator to produce electricity. The steam is cooled and condensed and either used in the plant's cooling system or injected back into the geothermal reservoir.

A binary cycle power plant is used when the water in a hot water reservoir is not hot enough to flash into steam. Instead, the lower-temperature hot water is used to heat a fluid that expands when warmed. The turbine is powered from the expanded, pressurized fluid. Afterwards, the fluid is cooled and recycled to be heated over and over again.

Geothermal Energy Production and Economics

Geothermal energy is put to work in many places around the world. The best-known geothermal energy sources in the United States are located in western states and Hawaii. Some moderately hot geothermal resources also exist in the Dakotas, along the Atlantic coast, and in Arkansas and Texas. Someday we may be able to use these resources too.

Most geothermal energy is produced in four states—California, Nevada, Utah, and Hawaii. Today the total installed capacity of geothermal power plants in the United State is 3,200 megawatts (MW) That's the energy equivalent of three nuclear power plants. American geothermal power plants range in size from a few hundred kilowatts to more than 130 megawatts.

In 1994, geothermal energy produced 18 billion kilowatt hours (kWh) of electricity, or 0.3 percent of the electricity used in this country. Still, this was enough to serve the electrical energy needs of over three million households. California gets six percent of its electricity from geothermal energy, more than any other state.

Geothermal supporters say geothermal energy production will grow in the 1990s despite the fact that geothermal energy production peaked in 1987 and has since declined. Geothermal supporters say at least 400 MW more capacity is planned for the next five years and estimate that geothermal energy could provide 10 percent of the electrical capacity of the western United States by the turn of the century.

Economics of Geothermal Energy. Geothermal power plants can produce electricity as cheaply as some conventional power plants. It costs 4.5 to seven cents per kWh to produce electricity from hydrothermal systems. In comparison, new coal-fired plants produce electricity at about four cents per kWh.

Initial construction costs for geothermal power plants are high because geothermal wells and power plants must be constructed at the same time.

But the cost of producing electricity over time is lower because the price and availability of the fuel is stable and predictable. The fuel does not have to be imported or transported to the power plant. The power plant literally sits on top of its fuel source.

Geothermal power plants are also excellent sources of baseload power. Baseload power is power that electric utility companies must deliver all day long. Baseload geothermal plants sell electricity all the time, not only during peakuse times when the demand for electricity is high.

Until recently, utilities were required to buy the least-cost electricity, without regard to environmental impacts. Federal and state energy and environmental agencies are studying ways to give preference to nonpolluting energy sources such as geothermal energy.

GEOTHERMAL ENERGY AND THE ENVIRONMENT

Geothermal energy is a renewable energy source that does little damage to the environment. Geothermal steam and hot water do contain naturally occurring traces of hydrogen sulfide (a gas that smells like rotten eggs) and other gases and chemicals that can be harmful in high concentrations. Geothermal power plants use "scrubber" systems to clean the air of hydrogen sulfide and the other gases. Sometimes the gases are converted into marketable products, such as liquid fertilizer. Newer geothermal power plants can even inject these gases back into the geothermal wells.

Geothermal power plants do not burn fuels to generate electricity as do fossil fuel plants. Geothermal power plants release less than one to four percent of the amount of carbon dioxide (C02) emitted by coal plants.

Emissions of sulfur compounds from motor vehicles and fossil fuel plants are also major contributors to acid rain. Geothermal power plants, on the other hand, emit only about one to three percent of the sulfur compounds that coal and oil-fired power plants do. Well-designed binary cycle power plants have no emissions at all.

Geothermal power plants are compatible with many environments. They have been built in deserts, in the middle of crops, and in mountain forests.

Geothermal development is often allowed on federal lands because it does not significantly harm the environment. Before permission is granted, however, studies must be made to determine what effect a plant may have on the environment. Geothermal features in national parks, such as the geysers and fumaroles in Yellowstone and Lassen National Parks, are protected by law, so geothermal energy is not tapped in these areas.

HYDROPOWER

WHAT IS HYDROPOWER?

Hydropower (from *hydro* meaning water) is energy that comes from the force of moving water.

The fall and flow of water is part of a continuous natural cycle. The sun draws moisture up from the oceans and rivers, and the moisture then condenses into clouds in the atmosphere. The moisture falls as rain or snow, replenishing the oceans and rivers. Gravity drives the water, moving it from high ground to low ground. The force of moving water can be extremely great. Anyone who has been white water rafting knows that!

Hydropower is called a renewable energy source because it is replenished by snow and rainfall. As long as the rain falls, we won't run out of this energy source.

HISTORY OF HYDROPOWER

Hydropower has been used for centuries. The Greeks used water wheels to grind wheat into flour more than 2000 years ago. In the early 1800s, American and European factories used the water wheel to power machines.

The water wheel is a simple machine. The water wheel picks up flowing water in buckets located around the wheel. The weight of the water causes the wheel to turn. Water wheels convert the **kinetic energy** (energy pertaining to motion) of water to mechanical energy. The mechanical energy can then be used to grind grain, drive sawmills, or pump water.

In the late 19th century, the force of falling water was used to generate electricity. The first hydroelectric power plant was built at Niagara Falls in 1879. In the following decades, many more hydroelectric plants were built. At its height in the early 1940s, hydropower provided 33 percent of this country's electricity.

But by the late 1940s, the best sites for big dams had been developed. Inexpensive fossil fuel (coal, oil) plants also entered the picture. At that time, these plants could make electricity more cheaply than hydro plants. Soon they began to underprice the smaller hydroelectric plants. It wasn't until the oil shocks of the 1970s that people showed a renewed interest in hydropower.

HYDROELECTRIC POWER PLANTS

As people discovered centuries ago, the flow of water represents a huge supply of kinetic energy that can be put to work. Water wheels are useful for generating mechanical energy to grind grain or saw wood, but they are not practical for generating electricity. Water wheels are too bulky and slow.

Hydroelectric plants (or hydro plants, as they are usually called) are very different. They use mode turbine generators to produce electricity just as thermal (coal, oil, nuclear) power plants do.

How a Hydro Plant Works. A hydro plant uses the force of falling water to make electricity. A typical hydro plant is a system with three parts:

- an **electric plant** where the electricity is produced.
- a **dam** that can be opened or closed to control water flow.

- a **reservoir** (artificial lake) where water can be stored.

To make electricity, a dam opens its gates to allow water from the reservoir to flow through a large tube called a penstock. At the bottom of the penstock, the fast-moving water spins the blades of a turbine. The turbine is connected to a generator to produce electricity. The electricity is then transported via huge transmission lines to a local utility company.

Head and Flow. The amount of electricity that can be generated at a hydro plant is determined by two factors: head and flow. **Head** is how far the **water drops.** It is the distance from the highest level of the dammed water to the point where it goes through the power-producing turbine.

Flow is how much water moves through the system. The more water moving through a system, the higher the flow. Generally, a high-head plant needs less water flow than a low-head plant to produce the same amount of electricity.

More about Dams. It's easier to build a hydro, plant where there is a natural waterfall. That's why the first hydro plant was built at Niagara Falls. Dams, which are artificial waterfalls, are the next best way.

Dams are built on rivers where the terrain will produce an artificial lake or reservoir above the dam. Today there are about 80,000 dams in the United States, but only three percent have power-generating hydro plants. Most dams are built for flood control and irrigation, not electric power generation.

A dam serves two purposes at a hydro plant. First, a dam increases the head or height of a waterfall. Second, it controls the flow of water. Dams release water when it is needed for electricity production. (Special gates called "spillway gates" release excess water from the reservoir during heavy rainfalls.)

Storing Energy. One of the biggest advantages of a hydro plant is its ability to store energy. The water in a reservoir is, after all, stored energy.

Water can be stored in a reservoir and released when needed for electricity production. During the day when people use more electricity, water can flow through aplant to generate electricity. Then, during the night when people use less electricity, water can be held back in the reservoir. Storage also makes it possible to save water from winter rains for summer generating power, or to save water from wet years for generating electricity during dry years.

*Pumped Storage Systems***.** Some hydro plants also use **pumped storage systems.** A pumped storage system operates much as a public fountain does. The same water is used again and again.

At a pumped storage hydro plant, flowing, water is used to make electricity and then stored in a lower pool. Depending on how much electricity is needed, the water may or may not be pumped back to an upper pool. Pumping water to the upper pool requires electricity so hydro plants usually use pumped storage systems when there is a big demand for electricity.

Pumped hydro is the most reliable energy storage system used by American electric utilities. Coal and nuclear power plants have no energy storage systems. They must turn to expensive gas and oil-fired generators when people demand lots of electricity. They also have no way to store any extra energy they might produce during normal generating periods.

Hydropower Production

How much electricity do we get from hydropower today? Quite a bit. Depending on whether the year has been wet or dry, hydro plants produce from eight to 10 percent of the electricity produced in this country (almost 8.5 percent in 1994), far more than any other renewable energy source. In Oregon and Washington, hydropower supplies over 85 percent of the electricity each year.

Currently, there are about 75 million kilowatts of hydroelectric generating capacity in the United States. That's equivalent to the generating capacity of 70 large nuclear power plants. The biggest hydro plant in the country is located at the Grand Coulee dam on the Columbia River in northern Washington.

The United States also gets some hydropower electricity from Canada. Some New England utility companies buy this imported electricity.

What does the future look like for hydropower? The best sites for hydropower dams have already been developed so the development of big hydro plants is unlikely. But existing plants could be enlarged to provide additional generating capacity. Plus, many flood-control dams not equipped for electricity production could be outfitted with generating equipment. The Federal Energy Regulatory Commission estimates 60 thousand megawatts of additional generating capacity could be developed in the United States.

Good source of baseload power. Demand for electricity is not steady; it goes up and down. People use more electricity during the day when they are awake and using electrical appliances, and less at night when they are asleep. People also use more electricity when the weather is very cold or very hot.

Electric utility companies have to produce electricity to meet these changing demands. **Baseload power** is the electricity that utilities have to generate all the time. For that reason, baseload power should be cheap and reliable. Hydropower meets both these requirements. Generating, electricity from flowing water is the cheapest way to generate electricity in the United States, and the fuel supply-flowing water is always available, especially at plants with pumped storage systems.

Hydro plants are more energy efficient than thermal power plants too. That means they waste less energy to produce electricity. In thermal power plants, a lot of energy is lost as heat.

Hydro plants also run 85 percent of the time, about 50 percent more than thermal plants.

Economics of Hydropower and the Environment

Hydropower is the *cheapest* way to generate electricity today. No other energy source, renewable or nonrenewable, can match it. In 1994 it cost less than *one cent* per kWh (kilowatt-hour) to produce electricity at a typical hydro plant. In comparison, it costs coal plants about four cents per kWh and nuclear plants two cents per kWh to generate electricity.

Producing electricity from hydropower is cheap because, once a dam has been built and the equipment installed, the energy source—flowing water—is free.

Another reason hydro plants produce power cheaply is due to their sturdy structures and simple equipment. Hydro plants are dependable and long-lived, and their maintenance costs are low compared to coal or nuclear plants.

There is one thing that may increase hydropower's costs in the future. The procedure for licensing a dam has become a lengthy and expensive process. Many environmental impact studies must be undertaken. And some times as many as 13 state and federal agencies must be consulted. It takes anywhere from five to seven years just to get a license to build a dam.

Hydropower and the Environment. Hydropower does present a few environmental problems. Damming rivers may destroy or disrupt wildlife and natural resources. Fish, for one, may no longer be able to swim upstream.

Hydro plant operations may also affect water quality by churning up dissolved metals that may have been deposited by industry long ago. Hydro plant operations may increase silting, change water temperatures, and lower the levels of dissolved oxygen. To some degree, these problems can be managed by constructing fish ladders, dredging silt, and carefully regulating plant operations.

On the plus side, hydropower's fuel supply (flowing water) is clean and is renewed yearly by snow and rainfall. Unlike fossil fuel plants, hydro plants do not emit any pollutants into the air because they bum no fuel.

Hydropower is also the only energy source that offers a whole range of added benefits. Dams control flood waters, and reservoirs provide lakes for boating, fishing, and swimming.

Natural Gas

What Is Natural Gas and History of Use

Natural gas is generally considered a nonrenewable fossil fuel. (There are some *renewable* sources of natural gas; we'll discuss these later.) Natural gas is called a fossil fuel because most scientists believe that natural gas was formed from the remains of tiny sea animals and plants that died 200-400 million years ago.

When these tiny sea animals and plants died, they sank to the bottom of the oceans where they were buried by layers of sand and silt. Over the years, the layers of sand and silt became thousands of feet thick, subjecting the energy-rich plant and animal remains to enormous pressure. Most scientists believe that the pressure, combined with the heat of the earth, changed this organic mixture into petroleum and natural gas. Eventually, concentrations of natural gas became trapped in the rock layers much like a wet household sponge traps water.

Raw natural gas is a mixture of different gases. Its main ingredient is **methane,** a natural compound that is formed whenever plant and animal matter decays. By itself, methane is odorless, colorless, and tasteless. As a safety measure, natural gas companies add a chemical odorant (it smells like rotten eggs) so escaping gas can be detected. Natural gas should not be confused with gasoline, which is made from petroleum.

History of Natural Gas. The ancient peoples of Greece, Persia, and India discovered natural gas many centuries ago. The people were mystified by the burning springs created when natural gas seeping from cracks in the ground was ignited by *lightning.* They sometimes built temples around these eternal flames so they could worship the fire.

About 2,500 years ago, the Chinese recognized that natural gas could be put to work. The Chinese piped the gas from shallow wells and burned it under large pans to evaporate seawater for salt.

Natural gas was first used in America to illuminate the streets of Baltimore in 1816. Soon after, in 1821, William Hart dug the first successful American natural gas well in Fredonia, New York. His well was 27 feet deep, quite shallow compared to today's wells. The Fredonia Gas Light Company opened its doors in 1858 as the nation's first natural gas company. By 1900, natural gas had been discovered in 17 states. In the past 40 years, the use of natural gas has grown dramatically. Today, natural gas accounts for about a quarter of the energy we use.

PRODUCING NATURAL GAS

Natural gas can be hard to find since it can be trapped in porous rocks deep underground. Scientists use many methods to find natural gas deposits. They may look at surface rocks to find clues about underground formations. They may set off small explosions or drop heavy weights on the surface and record the sound waves as they bounce back from the rock layers underground. They also may measure the gravitational pull of rock masses deep within the earth.

If test results are promising, the scientists may recommend drilling to find the natural gas deposits. Natural gas wells average 6,000 feet deep and can cost more than $75 per foot to drill, so it's important to choose sites carefully. On average, 27 out of every 100 exploratory wells produce gas. The others come up "dry." (The odds are better f or *developmental* wells-wells drilled on known gas fields. On average, 80 out of every 100 developmental wells yield gas.) Natural gas can be found in pockets by itself or in petroleum deposits.

After natural gas comes out of the ground, it goes to a processing plant where it is cleaned of impurities and separated into its various components. Approximately 90 percent of natural gas is composed of methane, but it also contains small amounts of other gases such as propane and butane.

Natural gas may also come from several other sources. One source is the gas found in coalbeds. Until recently, coalbed gas was just considered a safety hazard to miners, but now it is a valuable source of natural gas.

Another source of natural gas is the gas produced in landfills. Landfill gas is considered a *renewable* source of natural gas since it comes from decaying garbage. The gas from coalbeds and landfills accounts for three percent of the total gas supply today, yet their contribution could double by the year 2010.

Today natural gas is produced in 32 states, though just three states--Texas, Louisiana, and Oklahoma-produce 61 percent of the country's natural gas. Altogether, the United States produces nearly 22 percent of the world's natural gas each year. In 1994 the United States imported 12 percent of its natural gas from other countries-mostly from Mexico and Canada.

*Transporting and Storing Natural Gas***.** How does natural gas get to you, the consumer? Usually by pipeline. More than one million miles of underground pipelines link natural gas fields to major cities across the United States. Natural gas is sometimes transported thousands of miles by pipeline to its final destination. A machine called a compressor increases the pressure of the gas, forcing the gas to move along the pipelines. Compressor stations, which are spaced about 50 to 100 miles apart, move the gas along the pipelines at about 15 miles per hour.

Some gas moved along this subterranean highway is temporarily stored in huge underground reservoirs. The underground reservoirs are typically filled in the summer so there will be enough natural gas during the winter heating season.

Eventually, the gas reaches the "city gate" of a local gas utility. Here, the pressure is reduced and an odorant is added so leaking gas can be detected. Local gas companies use smaller pipes to carry gas the last few miles to homes and businesses. A gas meter measures the volume of gas a consumer uses.

WHO USES NATURAL GAS AND HOW MUCH?

Just about everyone in the United States uses natural gas. Natural gas ranks number three in energy use, right after petroleum and coal. Twenty-three percent of the energy we use in the United States comes from natural gas.

Industry is the biggest consumer of natural gas, using it mainly to manufacture goods. Industry also uses natural gas as an ingredient in fertilizer, photographic film, ink, glue, paint, plastics, laundry detergent, and insect repellents. Synthetic rubber and man-made fibers like nylon also could not be made without the chemicals derived from natural gas.

Residences are people's homes. Residences are the second biggest users of natural gas. Six in ten homes use natural gas for heating. Many homes also use gas water heaters, stoves, and clothes dryers.

Like residences, commercial use of natural gas is mostly for heating. Commercial use includes stores, offices, schools, churches, and hospitals.

Natural gas is also used to make electricity. Just as the heat energy in coal is used to make electricity, so can the heat energy in natural gas. Many people in the energy industry believe natural gas will play a bigger role in electricity production as the demand for electricity increases in the future. Why? Because natural gas power plants are cheaper and cleaner than coal plants. Natural gas plants produce electricity about 20 percent more efficiently than new coal plants, and they produce it with far fewer air-polluting emissions.

To a lesser degree, natural gas is making inroads as a transportation fuel. Natural gas can be used in any vehicle with a regular internal combustion engine, although the vehicle must be outfitted with a special carburetor and fuel tank. Natural gas is cleaner burning than gasoline, costs less, and has a, higher octane (power boosting) rating. Today more than 30,000 cars, trucks, and buses run on natural gas in the United States.

Natural Gas Reserves. People in the energy industry use two special terms when they talk about how much natural gas there is- resources and reserves. Natural gas **resources** include all the deposits of gas that are still in the ground waiting to be tapped.

Natural gas **reserves** are only those gas deposits that scientists know, or strongly believe, can be recovered given today's prices and drilling technology. In other words, when scientists estimate the amount of known gas reserves, they do not include gas deposits that may be discovered in the future or gas deposits that are not economical to produce given today's prices. (You can think of reserves this way. if it cost you $10 to manufacture a box of yoyos that you could sell for $8, would you make the yoyos? Of course not! You would lose $2 on every box.)

The United States has large reserves of natural gas. Most reserves are in the Gulf of Mexico and in the following states: Texas, Louisiana, Oklahoma, New Mexico, Wyoming, Kansas, and Alaska. If we continue to use natural gas at the same rate as we use it today, the United States has about a 50-year supply of natural gas, though another 200 years of additional gas supplies could be produced if people are willing to pay more for the gas they use.

New Ways to Use and Produce Natural Gas. Because natural gas is cleaner than coal or petroleum, scientists are researching new ways to use and produce it.

Fuel Cells. Many scientists are interested in using natural gas to generate electricity. Engineers have already developed ways to use coal/petroleum and natural gas together to generate electricity, but a device called a fuel cell can use natural gas alone. A fuel cell is similar to a battery. It uses a chemical process rather than combustion (burning) to convert the energy of a fuel into electricity. The chemical process is much more energy-efficient than combustion and it emits no air pollutants. Unfortunately, the technology to generate electricity from fuel cells must be improved if it is to be commercially successful.

Biomass. Scientists are also researching new ways to obtain natural (methane) gas from **biomass**—a fuel source derived from plant and animal wastes. Methane gas is naturally produced whenever organic matter decays. Today we can drill shallow wells into landfills to recover the methane gas. Landfills are already required to collect methane gas as a safety measure. Typically, landfills collect the gas and bum it to get rid of it. But the gas can be put to work. Last year over four billion cubic feet of landfill methane gas was used for heating and electricity production.

There are other ways to convert biomass into natural gas. One method converts aquatic plants, such as sea kelp, into methane gas. In the future, huge kelp farms could also produce renewable gas energy.

Liquid Natural Gas. Another successful development has been the conversion of natural gas into a liquid state. In its liquid state, natural gas is called **LNG**, or **liquid natural gas.**

LNG is made by cooling natural gas to a temperature of minus 260 degrees F. At that temperature, natural gas becomes a liquid and its volume is reduced 615 times. (A car reduced 615 times would fit on your thumbnail.) Liquid natural gas is easier to store than the gaseous form since it takes up much less space. LNG is also easier to transport. People can put LNG in special tanks and transport it on trucks or ships. Today more than 100 LNG storage facilities are operating in the United States.

NATURAL GAS PRICES AND THE ENVIRONMENT

Since 1985, natural gas prices have been set by the market. The federal government sets the price of transportation for natural gas that crosses state lines. State public utility commissions will continue to regulate natural gas utility companies-just as they regulate electric utilities.

These commissions regulate how much utilities may charge their customers, and they monitor the utilities' policies.

So how much does it costs to heat your home with natural gas? Compared to other energy sources, natural gas is a good buy. Heating your home with natural gas is cheaper than any other major heating source. It is more than four times less expensive than electricity when you use resistance heat. It is 25 percent less expensive than electricity when you use a heat pump.

Natural Gas and the Environment. All the fossil fuels coal, petroleum, and natural gas-release pollutants into the atmosphere when burned to provide the energy we need. The list of pollutants they release reads like a chemical cornucopia-carbon monoxides, reactive hydrocarbons, nitrogen oxides, sulfur oxides, and solid particulates (ash or soot).

The good news is that natural gas is the most environmentally friendly fossil fuel. It is cleaner burning than coal or petroleum because it contains less carbon than its fossil fuel cousins. Natural gas also has less sulfur and nitrogen compounds, and it emits less ash particulates into the air when it is burned than coal or petroleum fuels.

NUCLEAR

What Is Nuclear Energy?

Nuclear energy is energy that comes from the nucleus (core) of an atom. Atoms are the particles that make up all objects in the universe. Atoms consist of neutrons, protons, and electrons.

Nuclear energy is released from an atom through one of two processes: nuclear fusion or nuclear fission. In **nuclear fusion,** energy is released when the nuclei of atoms are combined or *fused* together. This is how the sun produces energy. In **nuclear fission,** energy is released when the nuclei of atoms are split apart. Nuclear fission is the only method currently used by nuclear plants to generate electricity.

The fuel most widely used by nuclear power plants for fissioning is **uranium.** Uranium is the heaviest of the 92 naturally occurring elements and is classified as a metal. It is also one of the few elements that is easily fissioned. Uranium was formed when the earth was created and is found in rocks all over the world. Rocks that contain a lot of uranium are called uranium ore, or pitchblende. Uranium, although abundant, is a nonrenewable energy source.

Two forms (isotopes) of uranium are found in nature, uranium-235 and uranium-238. These numbers refer to the number of neutrons and protons in each atom.

Uranium-235 is the form commonly used for energy production because, unlike uranium-238, its nucleus splits easily when bombarded by a neutron. During fissioning, the uranium-235 atom absorbs a bombarding neutron, causing its nucleus to split apart into two atoms of lighter weight. At the same time, the fission reaction releases energy in the form of heat, radiation, and more neutrons. The newly released neutrons go on to bombard other uranium atoms, and the process repeats itself over and over. This is called a **chain reaction.**

HISTORY OF NUCLEAR ENERGY

Compared to other energy sources, nuclear energy is a very new way to produce energy. It wasn't until the early 1930s that scientists discovered that the nucleus of an atom made up of protons an neutrons. Then just a few years later, in 1938, two German scientists split the nucleus of the atom apart by bombarding it with a neutron--the process called "fission." Soon after a Hungarian scientist discovered the "chain reaction" and its ability to produce enormous amounts of energy. Under the dark cloud of World War II, nuclear fission was first used to make a bomb. After that war, nuclear fission was developed for generating electricity.

The first commercial nuclear power plant came on line in Shippingport, Pennsylvania in 1957. Since then, the industry has experienced dramatic shifts in fortune. Throughout the mid 1960s, government and industry experimented with demonstration and small commercial plant. A period of rapid expansion followed between 1965 and 1975. The Three Mile Island accident in 1979 abruptly stalled the growth. The public opposition that organized in response to the Three Mile Island accident has been so successful that no new plants have been ordered since then.

THE NUCLEAR FUEL CYCLE

The chain of steps from mining the uranium ore, through its use in a nuclear reactor, through disposal--is called the **nuclear fuel cycle.**

Mining. The first step in the nuclear fuel cycle is mining the uranium ore. Workers mine uranium ore much as coal miners mine coal-in deep underground mines or in open-pit surface mines. A ton of uranium ore in the United States typically contains three to four pounds of uranium.

Milling. After it has been mined uranium ore is crushed. The crushed ore is usually poured into an acid, which dissolves the uranium, but not the rest of the crushed rock. The acid solution is drained off and dried, leaving a yellow powder called "yellowcake," consisting mostly of uranium. This process of removing uranium from the ore is called uranium milling.

Conversion. The next step in the cycle is the conversion of the yellowcake into a gas called uranium hexafluoride, or UF6. The uranium hexafluoride is then shipped to a gaseous diffusion plant for enrichment.

Enrichment. Because less than one percent of uranium ore contains uranium-235, the form used for energy production, uranium must be treated to increase the concentration of uranium-235. This treatment process--called uranium enrichment--increases the percentage of uranium-235 from one to three percent. It typically takes place at a gaseous diffusion plant where the uranium hexafluoride is pumped through filters that contain extremely tiny holes. Because uranium-235 has three fewer neutrons and is one percent lighter than uranium-238, it moves through the holes more easily than uranium-238. This method increases the percentage of uranium-235 as the gas passes through thousands of filters.

Fuel Fabrication. The enriched uranium is then taken to a fuel fabrication plant where it is prepared for the nuclear reactor. Here, the uranium is made into a solid ceramic material and formed into small barrel-shaped pellets. These ceramic **fuel pellets** can withstand very high temperatures, just like the ceramic tiles on the space shuttle. Fuel pellets are about the size of your fingertip, yet each one can produce as much energy as 120 gallons of oil. The pellets are then stacked and sealed in 12-foot metal tubes called **fuel rods.** Finally, the fuel rods are bundled into groups called **fuel assemblies.**

Nuclear Reactor. The uranium fuel is now ready for use in a nuclear **reactor.** Fissioning takes place in the reactor **core.** Surrounding the core of the reactor is a shell called the reactor pressure vessel. To prevent heat or radiation leaks, the reactor core and the vessel are housed in an air-tight containment building made of steel and concrete several feet thick.

The reactor core houses approximately 200 fuel assemblies. Spaced between the fuel assemblies are movable **control rods.** Control rods absorb neutrons and slow down the nuclear chain reaction. They are called control rods because they help control the fissioning process. Water also flows up through the fuel assemblies and control rods to remove some of the heat of the nuclear chain reaction.

The nuclear reaction generates heat energy just as burning coal or oil generate heat energy. Likewise, the heat is used to boil water into steam which turns a turbine generator to produce electricity. Afterward, the steam is condensed back into water and cooled in a separate structure called a **cooling tower.** In this way, the water can be used again and again.

Spent Fuel Storage. Like most industries, nuclear power plants produce waste. One of the main concerns about nuclear power plants is not the amount of waste created, which is quite small compared to other industries, but the radioactivity of some of that waste. The fission process creates radioactive waste products. After three or so years, these waste products build up in the fuel rods making the chain reaction more difficult. Utility companies generally replace one-third of the fuel rods every 12 to 18 months to keep power plants in continuous operation. The fuel that is taken out of the reactor is called **spent fuel.** The spent fuel contains both radioactive waste products and some unused nuclear fuel.

The spent fuel is typically stored near the reactor in a deep pool of water called the **spent fuel pool.** During storage, the spent fuel cools down and begins to lose most of its radioactivity through radioactive decay. In three months, the spent fuel will lose 50 percent of its radiation; in one year, 80 percent; in 10 years, 90 percent. The spent fuel pool is intended as a temporary method for storing nuclear waste. Eventually, the spent fuel will be reprocessed and/or transported to a permanent federal disposal site.

Reprocessing. Spent fuel contains both radioactive waste products and unused nuclear fuel. In fact, about one-third of the nuclear fuel remains unused when the fuel rod must be replaced. Reprocessing separates the unused nuclear fuel from the waste products so that it can be used in a reactor again and again.

But currently American nuclear power plants store the spent fuel in spent fuel pools without reprocessing. Why? Mainly because reprocessing is more expensive than making new fuel from uranium ore.

NUCLEAR WASTE REPOSITORIES, USE AND POWER PLANTS

Most scientists believe the safest way to store nuclear waste is in rock formations deep underground--called geological **repositories**. In 1982, the U.S. Congress agreed and passed the Nuclear Waste Policy Act. This law directed the U.S. Department of Energy to site, design, construct, and operate America's first repository by 1998. The repository will store radioactive waste from nuclear power plants and from defense weapons plants.

The same law also established the Nuclear Waste Fund to pay for the repository. People who use electricity from nuclear power plants contribute 1/10 of a cent for each kilowatt-hour of nuclear-generated electricity they use. An average American household, which uses about 7,500 kilowatt-hours a year, would contribute $7.50 a year to the fund if it got all its electricity from nuclear power. The nation contributed $600 million to the fund in 1993.

More recently, Congress passed the Nuclear Waste Policy Amendments Act in 1987. Among other things, this act proposed Yucca Mountain, Nevada as the nation's first repository site.

If the current plan is approved (it's not a done deal), nuclear waste will be sealed in steel canisters and stored in underground vaults located 1,000 feet below the surface by the year 2003.

Yucca Mountain is being studied as a repository site because it is dry (water won't percolate through the repository) and geologically stable (the chance of erupting volcanoes or damaging earthquakes is extremely slim). Another plus about the Yucca Mountain site is its isolation. Hardly anyone lives near it.

Although utility companies currently store their nuclear waste in pools of water at the power plant, some companies will run out of storage space by 1998--the original deadline for the opening of the repository. Utility companies are asking the Department of Energy to accept responsibility for the waste in 1998. The Department of Energy would need to store the waste in a temporary facility prior to its final burial atthe repository.

How Much Nuclear Energy Do We Use? Nuclear energy is an important source of electricity in the United States-it's second only to coal. Nuclear power provides about 21 percent of this country's electricity.

At the end of 1994, there were 109 nuclear power plants operating in the United States. No new plants are planned for the remaining decade.

In other parts of the world, nuclear energy is a growing source of electrical power. Nuclear energy now provides about 17 percent of the world's electricity. The United States, France, Japan, and Germany are the world leaders in producing electricity from nuclear power. France gets 75 percent of its electricity from nuclear power.

Licensing Nuclear Power Plants. Nuclear power plants must obtain a permit to start construction and, later, a license to begin operation.

Researchers conduct many studies to find the best site for a nuclear power plant. Detailed plans and reports are submitted to the **Nuclear Regulatory Commission,** the federal government agency responsible for licensing nuclear power plants and overseeing their construction and operation. When the builders of a nuclear power plant apply for a license, local hearings are held so people can testify and air their concerns and opinions. After a plant is built, the Nuclear Regulatory Commission places inspectors at the site to assure the plant is operating properly.

Economics of Nuclear Energy and the Environment

The cost of producing electricity from nuclear energy is somewhat higher than the cost of producing electricity from coal. Much of the cost of producing electricity at a nuclear plant comes not from the fuel source-uranium is very inexpensive at just $11 a ton-but from the cost of building the plant. Building a nuclear power plant is very expensive because of the high costs of licensing, construction, and inspection.

The cost of producing nuclear electricity is about two cents per kilowatt-hour (kWh). In comparison, the cost of producing electrical power from new coal plants is approximately four cents per kWh.

Uranium is an abundant natural resource that is found all over the world. At current rates of use, uranium resources could last more than 500 years. (A process called "breeding," which converts uranium into plutonium-an even better fuel--could extend uranium reserves for millions of years. Breeder reactors are already being used in France, but they are not planned for use in this country.) And because uranium is an extremely concentrated fuel source, it requires far less mining and transportation than other fuel sources for the energy it furnishes.

Nuclear Energy and the Environment. One of the best kept secrets about nuclear power its impact on the environment. There is very little. Generating electricity from nuclear power produces no air pollution because fuel is not burned as it is in coal and oil or gas-fired plants. In fact, using nuclear energy may become one way to solve pollution problem linked to acid rain and the greenhouse effect.

People are using more and more electricity. Some experts predict that we will have to use nuclear energy to produce the amount of electricity people need at cost they can afford. Whether or not we should use nuclear energy to produce electricity has become a controversial and sometimes highly emotional issue.

PETROLEUM

WHAT IS PETROLEUM AND HISTORY OF OIL

Petroleum is a fossil fuel. It is called a fossil fuel because it was formed from the remains of tiny sea plants and animals that died millions of years ago. When the plants and animals died, they sank to the bottom of the oceans. Here, they were buried by thousands of feet of sand and silt. Over time, this organic mixture was subjected to enormous pressure, and heat as the layers increased. The mixture changed, breaking down into compounds made of hydrogen and carbon atoms--**hydrocarbons.** Finally, an oil-saturated rock-much like a wet household sponge was formed.

All organic material does not turn into oil. Certain geological conditions must exist within the oil-rich rocks. There must be a trap of non-porous rock that prevents the oil from seeping out, and a seal (such as salt or clay) that keeps the oil from rising to the surface. Under these conditions, only two percent of the organic material is transformed into oil.

A typical petroleum reservoir is mostly sandstone or limestone in which oil is trapped. The oil in it may be as thin as gasoline or as thick as tar.

Petroleum is called a **nonrenewable** energy source because it takes millions of years to form. We cannot make new petroleum reserves.

History of Oil. People have used petroleum since ancient times. The ancient Chinese and Egyptians burned oil for lighting.

Before the 1850s, Americans often used whale oil to light their homes and shops. When whale oil became scarce, people began looking for other oil sources. In some places, oil seeped naturally to the surface of ponds and streams. People skimmed this oil and made it into kerosene. Kerosene was commonly used to light America's homes before the arrival of the electric light bulb.

As demand for kerosene grew, a group of businessmen hired Edwin Drake to drill for oil in Titusville, Pennsylvania. After much hard work and slow progress, he discovered oil in 1859. Drake's well was 69.5 feet deep, very shallow compared to today's wells.

Drake refined the oil from his well into kerosene for lighting. Gasoline and other products made during refining were simply thrown away because people had no use for them. In 1892, the "horseless carriage" solved this problem since it required gasoline. By 1920 there were nine million motor vehicles in this country and gas stations were opening everywhere.

PRODUCING OIL

Although research has improved the odds since Edwin Drake's days, petroleum exploration today is still a gamble. Geologists study underground rock formations to find areas that might yield oil. Still, even with advanced methods, only about 33 in every 100 exploratory wells have oil. The rest come up "dry."

When oil is found, a petroleum company brings in a 50 to 100-foot drilling rig and raises a derrick that houses the tools and pipes that go into the well. Today's oil wells average 5,000 feet deep and may sink below 20,000 feet.

To safeguard the environment, oil drilling and oil production are regulated by state and federal governments. Oil companies must get permission to explore for oil on new lands. Many experts believe that 85 percent of our remaining oil reserves are on land owned by the federal government. Oil companies lease the land from the federal government, which, in return, receives rental payments for the land as well as percentage payments from each barrel of oil.

Top Producers. Texas produces more oil than any other state. The other top producing states are Alaska, California, Louisiana, and Oklahoma--in that order. In all, 31 states produce petroleum.

From Well to Market. We cannot use crude oil in the state it's in when it comes out of the ground. The processis a little more complicated than that. So, how does thick, black crude oil come out of the ground and eventually get into your car as a thin, amber-colored liquid called gasoline? Let's find out.

*Oil Refineries***.** Oil's first stop outside the well is an oil refinery. A refinery is a plant where crude oil is processed. Sometimes refineries are located near oil wells, but usually the crude oil has to be delivered to the refinery by ship, barge, pipeline, or train.

After the crude oil has reached the refinery, huge round tanks store the oil until it is ready to be processed. **Tank farms** are sites with many storage tanks.

An oil refinery cleans and separates the crude oil into various fuels and byproducts. The most important one is gasoline. Some other petroleum products are diesel fuel, heating oil, and jet fuel.

Refineries use many different methods to make these products. One method is a heating process called **distillation.** Since oil products have different boiling points, the end products can be distilled or separated. For example, asphalts have a higher boiling point than gasolines, allowing the two to be separated.

Refineries have another job. They remove contaminants from the oil. A refinery removes sulfur from gasoline, for example, to increase its efficiency and to reduce air pollution.

Shipping Petroleum Products. After processing at the refinery, gasoline and other petroleum products are usually shipped out through pipelines. There are about 230,000 miles of pipeline in the United States. Pipelines are the safest and cheapest way to move large quantities of petroleum across land. Pump stations, which are spaced 20 to 100 miles apart along the underground pipelines, keep the petroleum products moving at around five miles per hour. At this rate, it takes 15 days to move a shipment of gasoline from Houston, Texas to New York City.

Distributing Petroleum Products. Companies called "jobbers" handle the wholesale distribution of oil. There are 15,000 jobbers in the U.S., and they sell just about everything that comes out of a barrel of crude oil. Jobbers fill bulk orders for petroleum products from gasoline stations, industries, utility companies, farmers, and so on.

The final link in the chain is the retailer. A retailer may be a gasoline station or a home heating oil company. The story ends when you pump gasoline into your car's tank, and the engine converts the gasoline's heat energy into mechanical energy to make your car move!

Demand for Oil and Imported Oil.

Since World War II, petroleum has replaced coal as the United States' leading source of energy. Petroleum supplies more than 39 percent of the energy used in the United States. (Coal supplies 22 percent of our energy needs, and natural gas supplies 23 percent.)

Americans use almost 17 million barrels of oil (more than 700 million gallons) every day of the year. And experts say we will be using more and more oil, especially for transportation, in the coming years. Even now, we use almost 13 percent more oil for transportation than we did in 1973 when the first oil crisis hit the United States. This is true even though today's automobiles get more than 1.5 times as many miles to the gallon as their 1970s counterparts. There are 50 percent more vehicles on the road today than in the 1970s. Today we use about two out of every three barrels of oil to keep us on the move.

Imported Oil. To satisfy our appetite for petroleum, the United States has become increasingly dependent upon other countries for petroleum. In 1994 we purchased 45 percent of our petroleum from other countries.

Americans know this dependence can be dangerous. We were first alerted to the danger in 1973 when some Arab countries stopped shipping oil (called an **embargo)** to the United States. These countries belonged to an international trade group called the **Organization of Petroleum Exporting Countries** or OPEC for short.

OPEC members try to set production levels for petroleum. As a rule, the less oil they produce, the higher the price of oil on the world market. The more oil they produce, the lower the price. The OPEC countries don't always agree with each other. Some OPEC countries want to produce less oil to raise prices. Other OPEC countries want to flood the market with petroleum to reap immediate returns.

Americans learned some lessons from the 1973 oil shock. Gas-guzzling cars became about as wanted as last week's leftovers, and the nation raced to find ways to conserve energy. But more oil shocks followed.

The next shock came in 1978-79 when the Iranian Revolution cut off oil production. Again, world oil prices raced up. Our most recent crisis was the Persian Gulf War. Iraq invaded Kuwait, and again, Americans worried about oil shortages and skyrocketing oil prices.

The U.S. has taken some steps to prevent another big oil crisis. For one thing, the U.S. has almost a three-month supply of oil tucked away in the **Strategic Petroleum Reserve (SPR).** Established in 1975, the SPR is only to be tapped during an energy emergency. The SPR was first used in January 1991, during the Persian Gulf Crisis.

The United States has also turned to non-Arab and non-OPEC countries for oil imports. Today, we import much of our oil from Canada and Mexico. This is good for us because we have friendly relations with our neighbors, and because the oil doesn't have to be shipped so far. Still, the amount of oil that we can import from Canada and Mexico is limited. By law, Mexico can only export half the oil it produces to the United States.

Even with the SPR and friends in the right places, the United States is not out of the woods. We still buy more than half our imported oil from OPEC countries, 11 percent of which comes from volatile Arab OPEC countries.

Some economists believe the United States is setting itself up for another oil crisis. Other analysts say a true oil shock--like those of the 1970s--is unlikely because the producing nations don't want to drive their customers away or encourage a shift to other forms Of energy.

Still, there are more steps we can take to help ensure our energy security. Depending on whom you talk to--whether an oil company representative or an environmentalist--opinions vary on the one or more steps we should take. Some experts believe we should decrease our demand for oil through increased conservation. Others say we should increase oil production and exploration in the United States, particularly in the Arctic National Wildlife Refuge (ANWR) in northern Alaska. Still others say we should use alternative fuels, especially for transportation. Some experts believe will need to do all three to dodge another big oil crisis.

Oil Prices and the Environment

Most of the world moves on petroleum--gasoline for cars; jet fuel for planes; diesel fuel for trucks. Then there is petroleum for running factories or manufacturing goods. That's why the price of oil is so important. Oil prices have declined for the fourth straight year in a row. (In 1994, the price of a barrel of oil was $15.50.)

Low oil prices are good for the consumer and the world economy. Low oil prices act as a check on inflation.

But there is another side--the U.S. oil industry does not prosper during periods of low oil prices. Oil industry workers lose their jobs, many small wells are permanently sealed, and the exploration for new oil sources drops off.

Low oil prices have another side effect. People use more oil when oil products are cheap. This can create a Catch-22 situation because we may depend more and more on foreign oil to keep up with demand.

Oil and the Environment. Petroleum products--gasoline for cars, fertilizers, etc.--have brought untold benefits to Americans and the rest of the world's people. We depend on these products, and, as consumers, we demand them.

But there is a flipside—pollution. Petroleum production and petroleum products can contribute to air and water pollution. Drilling for oil may disturb fragile ecosystems whether on land or sea. Transporting oil may endanger wildlife and the environment if it's spilled on rivers or oceans. Leaking underground storage tanks may pollute groundwater and create noxious fumes. Processing oil at the refinery may contribute to air and water pollution. Burning gasoline to fuel our cars contributes to air pollution. Even the careless disposal of waste oil drained from the family car can pollute streams, rivers, and lakes.

The situation is far from hopeless though, and many improvements have been made since the passage of the Clean Air Act in 1970. Oil companies have cleaned up their refineries by reducing air and water emissions.

Gasolines have been reformulated to bum cleaner, dramatically cutting the levels of lead, nitrogen oxide, carbon monoxide, and hydrocarbons released into the air. And, in response to the oil tanker spill in Prince William Sound, Alaska, in 1989, oil companies formed the Marine Spill Response Corporation to ensure fast and effective cleanup of oil spills.

The situation presents a challenge. The future must balance the growing demand for petroleum products with protection of the global environment.

PROPANE

WHAT IS PROPANE AND HISTORY OF PROPANE

Propane is a kissing cousin of natural gas and petroleum. Propane is usually found mixed with natural gas and petroleum deposits in rocks deep underground. Propane is called a **fossil fuel** because it was formed millions of years ago from the remains of tiny sea animals and plants.

When the plants and animals died, they sank to the bottom of the oceans where they were buried by layers of sand and silt. Over the years, the layers became thousands of feet thick. The layers were subjected to enormous heat and pressure, changing the energy-rich remains into petroleum and natural gas deposits. Eventually, pockets of these fossil fuels became trapped in rock layers much as a wet household sponge holds water.

Propane is just one of the many fossil fuels that are included in the **liquefied petroleum (LP) gas** family. Because propane is the type of LP-gas most commonly used in the United States, "propane" and "LP-gas" are often used synonymously. The chemical formula for propane is C3H8.

Just as water can change its physical state and become a liquid or a gas (steam vapor), so can propane. Under normal atmospheric pressure and temperature, propane is a gas. Under moderate pressure and / or lower temperatures, however, propane changes into a liquid. And that's the beauty of it.

Propane is easily stored as a liquid in pressurized tanks. (Think of the small tanks you see attached to a gas barbecue grill, for example.)

Propane takes up much less space in its liquid form. It is 270 times more compact in its liquid state than it is as a gas. A thousand gallon tank holding gaseous propane would provide a family enough cooking fuel for one week. A thousand gallon tank holding *liquid* propane would provide enough cooking fuel for almost ten years! Liquid propane instantly vaporizes into a gas when it is released from its tank to fuel propane gas appliances and equipment. Propane has been nicknamed the "portable gas" because it is easier to store and transport than natural gas.

Like its close cousin natural gas, propane is colorless and odorless. An odorant is added to propane (as it is to natural gas) to serve as a warning agent for escaping gas. And like all the fossil fuels---coal, natural gas, and petroleum--propane is a **nonrenewable energy source.**

History of Propane. Propane does not have a long history. It wasn't discovered until 1912 when people were trying to find a way to store gasoline. The problem with gasoline was that it evaporated when stored under normal conditions.

Dr. Walter Snelling, directing a series of experiments for the U.S. Bureau of Mines, discovered that several evaporating gases could be changed into liquids and stored at moderate pressure. The most plentiful of these gases was propane. Dr. Snelling developed a way to "bottle" the wet (liquid) gas. One year later, the commercial propane industry began heating American homes.

Producing and Transporting Propane

Propane comes from natural gas and petroleum wells. Fifty-five percent of the propane used in the United States is extracted from raw natural gas. (Raw natural gas is natural gas that hasn't been cleaned and processed yet.) Raw natural gas contains about 90 percent methane, five percent propane, and five percent other gases. The propane is separated from the other gases at a natural gas processing plant.

The remaining 45 percent is extracted from petroleum. Petroleum is separated into its various parts at a processing plant called a refinery.

Transporting Propane. How does propane get from natural gas processing plants or oil refineries to the consumer? Generally, propane first moves through underground pipelines to distribution terminals across the nation. Distribution terminals, which are operated by propane companies, function similarly to warehouses that store merchandise before shipping it to stores and shops. Sometimes, especially in the summer when less energy is needed for heating, propane is stored in large underground storage caverns.

After storage at distribution terminals, propane is transported via railroad tank cars, trucks, barges, and tanker ships to bulk plants. A bulk plant is where local propane dealers fill their small tank trucks.

People who use very little propane-backyard barbecue cooks, for example--must bring their propane cylinders to the dealer to be filled.

How Propane Is Used

Propane is used by homes, farms, business, and industry-and mostly for heating.

Homes. Propane is used mostly by homes in rural areas that do not have natural gas service. Propane appliances include ranges, ovens, space heaters, furnaces, water heaters, clothes dryers, and air conditioners. Millions of backyard cooks use gas (that's *propane* gas) grills for cooking. And recreational vehicles (RV's) usually have propane-fueled appliances, giving them a portable source of energy for cooking, hot water, and refrigeration.

Farms. Half of America's farms use propane to meet their energy needs, too. Farmers use propane to dry crops, brood chickens, power tractors, and warm greenhouses.

Business. Business and commercial establishments--from grocery stores to laundromats--use propane for heating and cooking.

Industry. Certain industries find propane well-suited to their special needs. Metal workers use small propane tanks to fuel their cutting torches and other equipment. Portable propane heaters give construction and road workers warmth in cold weather. Propane heaters at construction sites are used to dry concrete, plaster, and fuel pitch. Propane also heats asphalt for highway construction and repairs. And because propane is a very low-pollution fuel, fork-lift trucks powered by propane can operate safely inside factories and warehouses.

The United States uses more propane gas than any other country in the world. Propane supplies one percent of our total energy needs and ranks as the seventh most important source of energy in the country today, just after hydroelectric power and biomass.

Nearly 90 percent of the propane used in this country is produced in the United States. The other 10 percent is imported from Canada, Venezuela, and Middle Eastern countries.

PROPANE—A TRANSPORTATION FUEL TOO

Did you know that propane has been used as a transportation fuel for more than half a century? Taxicab companies, government agencies, and school districts often use propane, instead of gasoline, to fuel their fleets of vehicles. Today about six percent of propane's use is for transportation.

There are some interesting characteristics about propane that make it an ideal engine fuel. First, propane is clean-burning, much more so than gasoline. Propane leaves no lead, varnish, or carbon deposits that cause the premature wearing of pistons, rings, valves, and spark plugs. The engine stays clean, free of carbon and sludge. This means less maintenance and an extended engine life.

Also, propane is all fuel. It doesn't require additives usually blended into some grades of gasoline. Even without additive boosters, propane's octane rating of 110 is equal to and, in most cases, higher than available gasoline.

Propane-fueled engines produce less air pollution than gasoline engines. Carbon monoxide emissions from engines using propane are 50 percent to 92 percent lower than emissions from gasoline-fueled engines. Hydrocarbon emissions are 30 percent to 62 percent lower.

So why isn't propane used as a transportation fuel more often? For one reason, it's not as conveniently available as gasoline. Second, an automobile engine has to be adjusted to use propane fuel, and the cost of converting an engine to use propane is often prohibitive. Third, there is a slight drop in miles per gallon when propane is used to fuel vehicles.

WHAT IS SOLAR ENERGY?

Solar energy is energy that comes from the sun. Every day the sun radiates, or sends out, an enormous amount of energy. The sun radiates more energy in one second than people have used since the beginning of time!

Where does all this energy come from? It comes from within the sun itself. Like other stars, the sun is a big gas ball made up mostly of hydrogen and helium. The sun generates energy in its core in a process called **nuclear fusion.** During nuclear fusion, the sun's extremely high pressure and hot temperature cause hydrogen atoms to come apart and their nuclei (the central cores of the atoms) to *fuse* or combine. Four hydrogen nuclei fuse to become one helium atom. But the helium atom weighs less than the four nuclei that combined to form it. Some matter is lost during nuclear fusion. The lost matter is emitted into space as radiant energy.

It takes millions of years for the energy in the sun's core to make its way to the solar surface, and then just a little over eight minutes to travel the 93 million miles to earth. The solar energy travels to the earth at a speed of 186,000 miles per second, the speed of light.

Only a small portion of the energy radiated by the sun into space strikes the earth, one part in two billion. Yet this amount of energy is enormous. Every day enough energy strikes the United States to supply the nation's energy needs for one and a half years!

Where does all this energy go? About 15 percent of the sun's energy that hits the earth is reflected back into space. Another 30 percent is used to evaporate water, which, lifted into the atmosphere, produce's rain-fall. Solar energy also is absorbed by plants, the land, and the oceans. The rest could be used to supply our energy needs.

HISTORY OF SOLAR ENERGY.

People have harnessed solar energy for centuries. As early as the 7th century B.C., people used simple magnifying glasses to concentrate the light of the sun into beams so hot they would cause wood to catch fire. Over 100 years ago in France, a scientist used heat from a solar collector to make steam to drive a steam engine.

In the beginning of this century, scientists and engineers began researching ways to use solar energy in earnest. One important development was a remarkably efficient solar boiler invented by Charles Greeley Abbott, an American astrophysicist, in 1936.

The solar water heater gained popularity at this time in Florida, California, and the Southwest. The industry started in the early 1920s and was in full swing just before World War 11. This growth lasted until the mid-1950s when low-cost natural gas became the primary fuel for heating American homes. The public and world governments remained largely indifferent to the possibilities of solar energy until the oil shortages of the 1970s. Today people use solar energy to heat buildings and water and to generate electricity.

Solar Collectors and Solar Space Heating

Heating with solar energy is not as easy as you might think. Capturing sunlight and putting it to work is difficult because the solar energy that reaches the earth is spread out over a large area. The sun does not deliver that much energy to any one place at any one time . How much solar energy a place receives depends on several conditions. These include the time of day, the season of the year, the latitude of the area, and the clearness or cloudiness of the sky.

A **solar collector** is one way to collect heat from the sun. A closed car on a sunny day is like a solar collector. As sunlight passes through the car's glass windows, it is absorbed by the seat covers, walls, and floor of the car. The light that is absorbed changes into heat. The car's glass windows let light in, but don't let all the heat out. (This is also why greenhouses work so well and stay warm year-round.)

So, a solar collector does three things:

1. it allows sunlight inside the glass (or plastic);
2. it absorbs the sunlight and changes it into heat;
3. it traps most of the heat inside.

Solar Space Heating. Space heating means heating the space inside a building. Today many homes use solar energy for space heating. There are two general types of solar space heating systems: passive and active. A "hybrid" system is a mixture of the passive and active systems.

Passive Solar Homes. In a **passive** solar home, the whole house operates as a solar collector. A passive house does not use any special mechanical equipment such as pipes, ducts, fans, or pumps to transfer the heat that the house collects on sunny days. Instead, a passive solar home relies on properly oriented windows. Since the sun shines from the south in North America, passive solar homes are built so that most of the windows face south. They have very few or no windows on the north side.

A passive solar home converts solar energy into heat just as a closed car does. Sunlight passes through a home's windows and is absorbed in the walls and floors.

To control the amount of heat in a passive solar house, the doors and windows are closed or opened to keep heated air in or to let it out. At night, special heavy curtains or shades are pulled over the windows to keep the daytime beat inside the house. In the summer, awnings or roof overhangs help to cool the house by shading the windows from the high summer sun.

Heating a house by warming the walls or floors is more comfortable than heating the air inside a house. It is not so drafty. And passive buildings are quiet, peaceful places to live. A passive solar home can get 50 to 80 percent of the heat it needs from the sun.

Many homeowners install equipment (such as fans to help circulate air) to get more out of their passive solar homes. When special equipment is added to a passive solar home, the result is called a **hybrid** system.

Active Solar Homes. Unlike a passive solar home, an **active** solar home uses mechanical equipment, such as pumps and blowers, and an outside source of energy to help heat the house when solar energy is not enough.

Active systems use special solar collectors that look like boxes covered with glass. Dark-colored metal plates inside the boxes absorb the sunlight and change it into heat. (Black absorbs sunlight more than any other color.)

Air or a liquid flows through the collectors and is warmed by this heat. The warmed air or liquid is then distributed to the rest of the house just as it would be with an ordinary furnace system.

Solar collectors are usually placed high on roofs where they can collect the most sunlight. They are also put on the south side of the roof where no tall trees or tall buildings will shade them.

Storing Solar Heat. The challenge confronting any solar heating system--whether passive, active, or hybrid--is heat storage. Solar heating systems must have some way to store the heat that is collected on sunny days to keep people warm at night or on cloudy days.

In passive solar homes, heat is stored by using dense interior materials that retain heat well--masonry, adobe, concrete, stone, or water. These materials absorb surplus heat and radiate it back into the room after dark. Some passive homes have walls up to one foot thick.

In active solar homes, heat may be stored in one of two ways--a large tank may store a hot liquid, or rock bins beneath a house may store hot air.

Houses with active or passive solar heating systems may also have furnaces, wood-burning stoves, or another heat source to provide heat in case there is a long period of cold or cloudy weather. This is called a backup system.

Solar Hot Water Heating

Solar energy is also used to heat water. Water heating is usually the second leading home energy expense, costing the average family over $400, a year. Depending on where you live, and how much hot water your family uses, a solar water heater can pay for itself in as little as five years. A well-maintained system can last 15-20 years, longer than a conventional water heater.

A solar water heater works in the same way as solar space heating. A solar collector is mounted on the roof, or in an area of direct sunlight. It collects sunlight and converts it to heat. When the collector becomes hot enough, a thermostat starts a pump. The pump circulates a fluid, called a heat transfer fluid, through the collector for heating. The heated fluid then goes to a storage tank where it heats water. The hot water may then be piped to a faucet or showerhead. Most solar water heaters that operate in winter use a heat transfer fluid, similar to antifreeze, that will not freeze when the weather turns cold.

Today over 1.5 million homes in the U.S. use solar heaters to heat water for their homes or swimming pools.

Solar Electricity

Besides heating homes and water, solar energy also can be used to produce electricity. Two ways to generate electricity from solar energy are photovoltaics and solar thermal systems.

Photovoltaic Electricity. **Photovoltaic** comes from the words *photo* meaning "light" and *volt*, a measurement of electricity. Sometimes photovoltaic cells are called **PV cells** or **solar cells** for short. You are probably already familiar with solar cells. Solar-powered calculators, toys, and telephone call boxes all, use solar cells to convert light into electricity.

A photovoltaic cell is made of two thin slices of **silicon** sandwiched together and attached to metal wires. The top slice of silicon, called the N-layer, is very thin and has a chemical added to it that provides the layer with an excess of free electrons. The bottom slice, or P-layer, is much thicker and has a chemical added to it so that it has very few free electrons.

When the two layers are placed together, an interesting thing happens-an electric field is produced that prevents the electrons from traveling from the top layer to the bottom layer. This one-way junction with its electric field becomes the central part of the PV cell.

When the PV cell is exposed to sunlight, bundles of light energy known as photons can knock some of the electrons from the bottom P-layer out of their orbits through the electric field set up at the P-N junction and into the N-layer.

The N-layer, with its abundance of electrons, develops an excess of negatively charged electrons. This excess of electrons produces an electric force to push the additional electrons away. These excess electrons are pushed into the metal wire back to the bottom P-layer, which has lost some of its electrons.

This electrical current will continue flowing as long as radiant energy in the form of light strikes the cell and the pathway, or circuit, remains closed.

Current PV cell technology is not very efficient. Today's PV cells convert only about 10 to 14 percent of the radiant energy into electrical energy. Fossil fuel plants, on the other hand, convert from 30-40 percent of their fuel's chemical energy into electrical energy. The cost per kilowatt-hour to produce electricity from PV cells is presently three to four times as expensive as from conventional sources. However, PV cells make sense for many uses today, such as providing power in remote areas or other areas where electricity is difficult to provide. Scientists are researching ways to improve PV cell technology to make it more competitive with conventional sources.

Solar Thermal Electricity . Like solar cells, **solar thermal systems** use solar energy to make electricity. But as the name suggests, solar thermal systems use the sun's heat to do it.

Most solar thermal systems use solar collectors with mirrored surfaces to concentrate sunlight onto a receiver that heats a liquid. The super-heated liquid is used to make steam that drives a turbine to produce electricity in the same way that coal, oil, or nuclear power plants do.

Solar thermal systems may be one of three types: central receiver, dish, or trough. A central receiver system uses large mirrors on top of a high tower to reflect sunlight onto a receiver. This system has been dubbed a "solar power tower." Another system uses a dish-shaped solar collector to collect sunlight. This system resembles a television satellite dish. A third system uses mirrored troughs to collect sunlight. Until recently, trough systems seemed the most promising.

The world's first solar electric plant used mirrored troughs. LUZ, as the plant was called, was perfectly situated in the sunny Mojave desert of

California. LUZ was the only solar plant to generate electricity economically. Dollar for dollar, it had always been cheaper to use conventional sources of energy (coal, oil, nuclear) to generate electricity. But the LUZ solar plant turned that around, producing electricity as cheaply as many new coal plants, and with no hidden pollution costs. The future looked bright for this pioneering solar plant and then the dream cracked. LUZ closed its doors at the end of 1992 because of a drop in oil prices and an over-budget construction project at LUZ's home-base.

LUZ may be gone, but most solar energy engineers believe solar power towers will be ready to take the place of trough systems very soon.

Solar Energy and the Environment

In the 1970s, the push for renewable energy sources was driven by oil shortages and price increases. Today, the push for renewable energy sources is driven by a renewed concern for the environment.

Solar energy is the prototype of an environmentally friendly energy source. It consumes none of our precious energy resources, makes no contribution to air, water, or noise pollution, does not pose a health hazard, and contributes no harmful waste products to the environment.

There are other advantages too. Solar energy cannot be embargoed or controlled by any one nation. And it will not run out until the sun goes out.

Wind

What Is Wind?

Wind is simply air in motion. It is caused by the uneven heating of the earth's surface by the sun. Since the earth's surface is made up of land, desert, water, and forest areas, the surface absorbs the sun's radiation differently.

During the day, air above the land heats more quickly than air above water. The hot air over the land expands and rises, and the heavier, cooler air over a body of water rushes in to take its place, creating local winds. At night, the winds are reversed because air cools more rapidly over land than over water.

Similarly, the large atmospheric winds that circle the earth are created because land near the equator is heated more by the sun than land near the North and South Poles.

Today people can use wind energy to produce electricity. Wind is called a *renewable* energy source because we will never run out of it.

History of Wind Machines

Throughout history people have harnessed the wind. Over 5,000 years ago, the ancient Egyptians used wind power to sail their ships on the Nile River. Later people built windmills to grind their grain. The earliest known windmills were in Persia (the area now occupied by Iran). The early windmills looked like large paddle wheels.

Centuries later, the people in Holland improved the windmill. They gave it propeller-type blades and made it so it could be turned to face the wind. Windmills helped Holland become one of the world's most industrialized countries by the 17th century.

American colonists used windmills to grind wheat and corn, to pump water, and to cut wood at sawmills.

In this century, people used windmills to generate electricity in rural areas that did not have electric service. When power lines began to transport electricity to rural areas in the 1930s, the electric windmills were used less and less.

Then in the early 1970s, oil shortages created an environment eager for alternative energy sources, paving the way for the re-entry of the electric windmill on the American landscape.

Today's Wind Machine

Today's **wind machine** is very different from yesterday's windmill. Along with the change in name have come changes in the use and technology of the windmill.

While yesterday's machines were used primarily to convert the wind's kinetic energy into mechanical power to grind grain or pump water, today's wind machines are used primarily to generate electricity.

Like old-fashioned windmills, today's wind machines still use blades to collect the wind's kinetic energy. Windmills work because they slow down the speed of the wind. The wind flows over the airfoil shaped blades causing lift, like the effect on airplane wings, causing them to turn. The blades are connected to a drive shaft that turns an electric generator to produce electricity.

Modern wind machines are still wrestling with the problem of what to do when the wind isn't blowing. Large turbines are connected to the utility power network-some other type of generator picks up the load when there is no wind. Small turbines are often connected to diesel / electric generators or sometimes have a battery to store the extra energy they collect when the wind is blowing hard.

Types of Wind Machines. Two types of wind machines are commonly used today:

- **horizontal,** which has blades like airplane propellers; and
- **vertical,** which looks like an egg-beater.

Horizontal-axis wind machines are used the most. They make up 95 percent of all wind machines. A typical horizontal wind machine stands as tall as a 10-story building and has two or three blades that span 60 feet across. The largest wind machines in the world have blades longer than a football field! Wind machines stand tall and wide to capture more wind.

Vertical-axis wind machines make up just five percent of the wind machines used today. The typical vertical wind machine stands 100 feet tall and 50 feet wide. The most popular vertical wind machine today is the Darrieus wind turbine, named after its inventor, J.G.S. Darrieus, a French engineer.

Which Type Is Better? Each wind machine has its advantages and disadvantages. Horizontal-axis machines need a way to keep the rotor facing the wind. This is done with a tail on small machines. On large turbines, either the rotor is located down wind of the tower acting like a weather vane, or drive motors are used. Vertical-axis machines accept wind from any direction.

Both types of rotors are turned by air flowing over their wing shaped blades. Vertical axis blades lose energy as they turn out of the wind, while horizontal-axis blades work all the time. Also, at many sites, the wind increases as you go higher-above the ground, giving an advantage to tall horizontal-axis turbines. The small tower and ground mounted generators on vertical-axis turbines make them less costly and easier to maintain.

Wind Power Plants. Wind power plants, or wind farms as they are sometimes called, are clusters of wind machines used to produce electricity. A wind farm usually has hundreds of wind machines in all shapes and sizes.

Unlike coal or nuclear plants, most wind plants are not owned by public utility companies. Instead they are owned and operated by business people who sell the electricity produced on the wind farm to electric utilities. These private companies are known as **Independent Power Producers.**

Operating a wind power plant is not as simple as plunking down machines on a grassy field. Wind plant owners must carefully plan where to locate their machines. They must consider wind availability (how much the wind blows), local weather conditions, nearness to electrical transmission lines, and local zoning codes.

Wind plants also need a lot of land. One wind machine needs about two acres of land to call its own. A wind power plant takes up hundreds of acres. On the plus side, farmers can grow crops around the machines once they have been installed.

After a plant has been built, there are still maintenance costs. In some states, maintenance costs are offset by tax breaks given to power plants that use renewable energy sources. The Public Utility Regulatory Policies Act, or PURPA, also requires utility companies to purchase electricity from independent power producers at rates that are fair and nondiscriminatory.

Wind Resources and Energy Production

Where is the best place to build a wind plant? There are many good sites for wind plants in the United States including California, Alaska, Hawaii, the Great Plains, and mountainous regions. Scientists say there is enough wind in 37 states to produce electricity from the wind. Generally, an average wind speed of 14 mph is needed to convert wind energy into electricity economically. The average wind speed in the United States is 10 mph.

Scientists use an instrument called an **anemometer** to measure how fast the wind is blowing. An anemometer looks like a modern-style weather vane. It has three spokes with cups that spin on a revolving wheel when the wind blows. It is hooked up to a meter that tells the wind speed. (By the way, a weather vane tells you the direction of the wind, not the speed.)

As a rule, wind speed increases with altitude and over open areas with no wind breaks. Good sites for wind plants are the tops of smooth, rounded hills, open plains or shorelines, and mountain gaps that produce wind funneling. The three biggest wind plants in California are located at mountain gaps.

Wind speed varies throughout the country. It also varies from season to season. In Tehachapi, California, the wind blows more from April through October than it does in the winter. This is because of the extreme heating of the Mojave desert during the summer months. The hot desert air rises, and the cooler, denser air from the Pacific Ocean rushes through the Tehachapi mountain pass to take its place. In a state like Montana, on the other hand, the wind blows more during the winter.

By happy coincidence, these seasonal variations perfectly match the electricity demands of the regions. In California, people use more electricity during the summer when air conditioners are used for cooling. Conversely, more people use electricity in Montana during the winter heating months.

Wind Energy Production. How much energy can we can get from the wind? We will use two terms to describe wind energy production: efficiency and capacity factor.

Efficiency refers to how much useful energy (electricity, for example) we can get from an energy source. A 100 percent energy efficient machine would change all the energy put into the machine into useful energy. It would not waste any energy. (You should know there is no such thing as a 100 percent energy efficient machine. Some energy is always "lost" or wasted when one form of energy is converted to another. The "lost" energy is usually in the form of heat.)

How efficient are wind machines? Wind machines are just as efficient as coal plants. Wind plants convert 30 percent of the wind's kinetic energy into electricity. A coal-fired power plant converts about 30-35 percent of the heat energy in coal into electricity.

It is the capacity factor of wind plants that puts them a step behind other power plants. Capacity factor refers to the capability of a plant to produce energy. A plant with a 100 percent capacity rating would run all day, every day at full power. There would be no down time for repairs or refueling, an impossible dream for any plant.

Wind plants have about a 25 percent capacity rating because wind machines only run when the wind is blowing around nine mph or more. In comparison, coal plants typically have a 75 percent capacity rating since they can run day or night, during any season of the year.

One wind machine can produce 275-500 thousand kilowatt-hours (kWh) of electricity a year. That is enough electricity for about 50 homes per year.

In this country, wind machines produce about three billion kWh of energy a year. Wind energy provides 0. 12% of the nation's electricity, a very small amount. Still, that is enough electricity to serve more than 300,000 households, as many as in a city the size of San Francisco or Washington, D.C.

California produces more electricity from the wind than any other state. It produces 98 percent of the electricity generated from the wind in the United States. Some 16,000 wind machines produce more than one percent of California's electricity. (This is about half as much electricity as is produced by one nuclear power plant.) In the next 15 years, wind machines could produce five percent of California's electricity.

Why is California outproducing every other state? More than any other reason, wind energy has taken off in this state because of California's progressive state policies that support renewable energy sources. Other states have just as good wind resources as California.

The United States is the world's leading wind energy producer. The U.S. produces about half of the world's wind power. Other countries that have invested heavily in wind power research are Denmark, Japan, Germany, Sweden, The Netherlands, United Kingdom, and Italy.

What does the future look like for wind energy? Using a best-case scenario, the American Wind Energy Association (AWEA) estimates wind energy could produce more than 10 percent of the nation's electricity within the next 30 years.

So, wind energy may be an important alternative energy source in the future, but it will not be the sole answer to our energy problems. We will still need other energy sources to meet our growing demand for electricity.

WIND ENERGY ECONOMICS AND THE ENVIRONMENT

On the economic front, there is a lot of good news for wind energy. First, a wind plant is far less expensive to construct than a conventional energy plant. Wind plants can simply add wind machines as electricity demand increases.

Second, the cost of producing electricity from the wind has dropped dramatically in the last two decades. Electricity generated by the wind cost 30 cents per kWh in 1975, but now costs less than five cents per kWh. In comparison, new coal plants produce electricity at four cents per kWh.

Wind and the Environment. In the 1970s and 1980s, oil shocks and shortages pushed the development of alternative energy sources. In the 1990s, the push may come from something else, a renewed concern for the earth's environment.

Wind energy is clean. Wind machines produce no air or water pollution because no fuel is burned to generate electricity.
The only environmental drawbacks to wind energy may be a wind plant's effect on bird populations and its visual impact on the surrounding landscape. To some, the glistening blades of wind machines are an eyesore; to others, they're a beautiful alternative to smog-filled skies.