tecWAVE Project Summary

*Principal Investigators:*
Professor Judith Scott, Education Department, University of California Santa Cruz
Professor Yi Zhang, School of Engineering, University of California Santa Cruz

This summary for the TEaching with Computers: Word Annotations for Vocabulary Education (tecWAVE) project (IES: Grant #R305A00596 titled: *Explicit Scaffolding for Word Learning in Context through Multimedia Multilingual Word Annotation*) highlights the findings of our collaborative research project, involving researchers at the University of California, Santa Cruz in both the Education Department and the Baskin School of Engineering. A team of engineers and graduate students worked with Professor Zhang to develop the tecWAVE tool using artificial intelligence algorithms. Another team of educational researchers and graduate students worked with Professor Scott, and a set of middle school teachers, to create appropriate educational applications and reports, develop assessment tools, pilot test the application, provide iterative feedback on its development, and finally to test its completed form in middle school science and English Language Arts classes in three school districts.

**tecWAVE: An Overview**

The tecWAVE innovation is a major advancement towards alleviating the difficulty of encountering unknown words in electronically available texts. The technology developed through this grant allows students to click on unfamiliar words in context while reading on a computer screen. The annotation program, triggered by students’ clicks, quickly accesses streams of context appropriate pictures as well as easy-to-read definitions of the words in English, Spanish and Chinese. This project addresses the grave concern that many struggling readers and English learners cannot comprehend their grade level texts and novels in part because they are unfamiliar with the vocabulary used in these resources. Our technological solution blends cutting edge computer science and educational research to provide students with “one-click” access to definitions for nouns that match the context of the text (e.g., *a river bank* instead of *bank* as a financial institution) read on computers, match their reading level, match their home language if they wish, and provide images for the words.

Both qualitative and quantitative assessments indicate that tecWAVE significantly facilitates vocabulary learning. Measureable and significant word learning for all students in both subject areas occurred when they had the opportunity to access information about previously unknown words using tecWAVE. Those who were learning English were just as likely to learn word meanings from tecWAVE as English proficient students. In addition, students, teachers and parents were enthusiastic about the product and interested in its widespread application across subject areas and age groups both inside and outside of schools.

**Specific results derived from the project:**

*A. We created a computer program that provides explicit scaffolding for word learning in context through multimedia multilingual word annotations.*
The tecWAVE technology allows students to click on unfamiliar words in context while reading on a computer screen. Triggered by students’ clicks, tecWAVE quickly provides pop-ups of context-appropriate pictures and definitions of nouns in Spanish, Chinese and English. The key components include 1) identifying vocabulary from the text that may be unfamiliar to students; 2) disambiguating the meaning of the word based on the context of the surrounding sentences; 3) searching the data repository for definitions of the word in English and other languages (when appropriate) and related images; and 4) creating a personalized system based on the student’s metadata, and reading, writing and testing history.

tecWAVE is based on cutting edge research and artificial intelligence technologies that automate the annotation process. These technologies include 1) an English Part-Of-Speech Tagger (POS Tagger), a software program that reads text in English and assigns parts of speech (such as noun, verb, adjective, etc.) to each word based on both its definition, as well as its context; 2) a stemmer to conflate morphological variants, such as plurals, into the root form; 3) a word sense disambiguation engine; 4) a context-based image search engine; and 5) a context-based word translation engine. We were able to integrate the various components, generate an annotation for each word, and cache the annotated results with the text. When a user clicks a word, the cached annotation is shown to the user without significant time delay.

An important contribution of this technology is the integrated and easy-to-use nature of the program that provides context appropriate definitions and images. Going to a dictionary for definitions is a complex metacognitive task, even when the dictionary is online but not integrated with the text. It involves accessing the word, finding the right sense of the word for the context of the sentence, reinserting the definition of the word into the sentence, and judging whether or not the definition makes sense in context. tecWAVE eliminates major hurdles in processing the information from definitions by providing the correct word sense for each particular context immediately as it is needed in the reading process.

Given feedback from students, teachers and parents, we went beyond the scope of the goal to also provide a prototype for an Android tablet with the integration of an audio pronunciation component. The audio component can help students recognize words in their oral vocabularies that they may not recognize in print, as well as move them toward using these words orally on their own. Deploying the technology onto alternative platforms will make it more readily available to students both inside and outside of school.

A. Quantitative results indicate that tecWAVE is effective.

As part of the iterative development process, we used tecWAVE in limited pilot studies during the second and third year of the project. Then, in spring of the third year, we conducted our final trial of tecWAVE in 24 intact classes in three middle schools in three school districts in California. Thirteen of the classes were studying short stories in English language arts (ELA), and 11 were studying biology. The program was tailored to individual students upon login, so that words were annotated according to a pretest. The sets of words students selected as
“unknown” were randomly split in half, with an equal number annotated/unannotated. A multiple-choice vocabulary posttest on unknown words followed the reading.

Five hundred and forty-six students participated in the study. Of these, the 373 students with complete data include 76 grade six students, 257 grade seven students, and 40 grade eight students. Seventeen total home languages were identified, with Spanish and English as the most common languages (90%). Thirty-two percent of the students were English Only (EO), 14% were English Learners (EL), 22% were IFEP (Initial Fluent English Proficient) and 29% were RFEP (Reclassified Fluent English Proficient).

The results indicate that the probability of learning words well enough to pass a rigorous multiple choice test increases significantly if students have the opportunity to access information about the annotated words digitally while reading (F(1,353) = 22.98, p < .000). Neither home language nor status as an English learner interacted significantly with this factor, although there was a main effect of EL status. In addition, there was a main effect of instructional context with students obtaining higher posttest scores, on average, in science. However, the instructional content area did not interact significantly with percent correct due to annotation.

We interpret these results to indicate that tecWAVE works equally well for all types of middle school students regardless of EL status, home language background, or the type of content being read (fiction or informational text).

The results provide strong evidence that approaching vocabulary development with explicit scaffolding through technology is effective. In other words, giving students the opportunity to access information about words they said that they didn't know provided enough support for them to learn more than when this opportunity wasn't available (words weren't annotated).

**B. Qualitative results indicate that students, teachers and parents are enthusiastic about tecWAVE.**

Qualitative interviews, conducted with fifteen middle school students during the study, revealed that tecWAVE was well received by the participants and considered a valuable tool for learning that was both fun and easy. Students felt that tecWAVE was particularly well suited for classes with dense vocabulary loads such as Science, Math and Physical/Health Education. In addition to clicking on unknown words, the interviews also indicated that students were curious about words that were partially known and clicked on them as well. Whereas it is unlikely that efforts would be made to confirm word meanings using conventional dictionaries or even online dictionaries, clickable annotations reduced the time and effort required to verify accuracy of word knowledge.

Teachers and students who used tecWAVE in the study classrooms, as well as teachers and parents exposed to it for the first time in focus group discussion, are enthusiastic about its potential to enhance word learning and reading comprehension. Several indicated that they are eager to see this become readily available for the educational community and for their own use.
C. We developed a technique to measure gains in word knowledge using carefully constructed pretests and posttests, along with a series of reports for teachers and researchers.

Because traditional vocabulary assessments are unwieldy, time consuming, and lack of alignment between theory and practice (Pearson, Hiebert & Kamil, 2007; Scott, Lubliner & Hiebert, 2006), we needed to develop assessments that would allow quick assessment of word knowledge at the beginning of reading and assess word knowledge accurately post reading. In addition, we wanted to use the pretest results to manipulate exposure to annotations of unknown words. The distractors developed for the posttest were all within the same semantic field, allowing a more sophisticated assessment of understanding than might otherwise be found.

Reports to teachers and researchers can provide click information and post test information about words both by individual students and collectively for different language groups and by classes. This can facilitate teachers’ vocabulary instruction and provide monitoring for conceptual content knowledge that students may missed or misunderstood.

Conclusion:

The tecWAVE project has made a major contribution to the field of educational technology and vocabulary research by illuminating the possibilities of blending cutting edge computer science with high level educational research to address a significant and persistent educational problem: that many struggling readers and English learners cannot comprehend their grade level texts and novels in part because they are unfamiliar with the vocabulary used in these resources. We built a system that provides multilingual multimodal word annotations for nouns that brings up definitions, images and translations tied to the context in which the words are found. And, this system created measureable and significant word learning for a diversity of middle school students in both science and English Language Arts. Giving students the opportunity to quickly access easy-to read, contextually appropriate information about words they said that they didn't know provided enough support for them to learn more than when this opportunity wasn't available (words weren't annotated).