

2000

# A quantitative interaction analysis of an authentic CALL environment: New Dynamic English

Dustin E. Tower  
*Iowa State University*

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>



Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#)

---

## Recommended Citation

Tower, Dustin E., "A quantitative interaction analysis of an authentic CALL environment: New Dynamic English" (2000). *Retrospective Theses and Dissertations*. 278.

<https://lib.dr.iastate.edu/rtd/278>

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

A quantitative interaction analysis of an authentic CALL environment: *New Dynamic English*

Dustin Everett Tower

Major Professor: Volker Hegelheimer  
Iowa State University

This study describes second language learners' use of interactional options (e.g., a glossary usage) within *New Dynamic English*, a listening-based computer-assisted language learning (CALL) program produced by DynEd International. The study then elaborates on the students' use of options and how it relates to the students' performances within the program and their second language (L2) learning. Student choices in an interactive multimedia environment and modification of input are discussed. Results from this study, though they conflict with past research, have important implications for future L2 teachers and researchers in second language acquisition (SLA) and CALL. This study was performed in an authentic learning environment and points out the benefits, drawbacks, and the need for much more authentic learning research.

A quantitative interaction analysis of an authentic CALL environment: *New*

*Dynamic English*

by

Dustin Everett Tower

A thesis submitted to the graduate faculty

In partial fulfillment of the requirements for the degree of

MASTER OF ARTS

Major: English (Teaching English as a Second Language/Applied Linguistics)

Major Professor: Volker Hegelheimer

Iowa State University

Ames, Iowa

2000

Graduate College  
Iowa State University

This is to certify that the Master's thesis of

Dustin Everett Tower

has met the thesis requirements of Iowa State University

---

Major Professor

---

For the Major Program

---

For the Graduate College

## TABLE OF CONTENTS

<b>CHAPTER 1. INTRODUCTION</b>	<b>1</b>
Purpose of the Study	1
Rationale	2
Research Questions	3
Organization of this study	4
<b>CHAPTER 2. LITERATURE REVIEW</b>	<b>6</b>
Artificial Learning Environments	7
Experimental Learning Environments	10
Authentic Learning Environments	12
Modification of Input	16
<b>CHAPTER 3. METHODOLOGY</b>	<b>19</b>
Participants	19
Materials	19
Procedure	25
Analysis	28
<b>CHAPTER 4. RESULTS</b>	<b>33</b>
Frequency of Option Selection	33
Statistical Results	36
Group Ranking Analysis	41
<b>CHAPTER 5. DISCUSSION AND CONCLUSION</b>	<b>44</b>
Insights for Researchers, Teachers, and Developers	44
Conflict with Past Research	49
Limitations of the Study	50
Suggestions for Future Research	52
Benefits and Drawbacks	54
Conclusion	55
<b>APPENDIX A: PLACEMENT TEST DESCRIPTION</b>	<b>57</b>
<b>APPENDIX B: INTERACTION DESCRIPTION</b>	<b>63</b>
<b>APPENDIX C: PRELIMINARY MULTIPLE REGRESSION ANALYSES</b>	<b>67</b>
<b>APPENDIX D: SYSTEM REQUIREMENTS</b>	<b>68</b>
<b>BIBLIOGRAPHY</b>	<b>69</b>
<b>ACKNOWLEDGEMENTS</b>	<b>73</b>

## CHAPTER 1. INTRODUCTION

Though computer-assisted language learning (CALL) was first implemented in the 1960s, much of the learning was more or less “drill and practice” in its approach (Warschauer, 1996). It was not until the early 1980s that more communicative approaches to language learning influenced CALL to become more interactive, take on multimedia capacities, and offer students options as to how they interacted with a program (ibid.). Since that time, due to the rapid increases in technology and importance of the microcomputer in society, CALL has now become a tool that many language teachers are unable to overlook (Levy, 1996; Warschauer, 1996). Interactive multimedia CALL, and CALL in general, has become a technology which is expanding faster than language teachers and researchers can keep up (Levy, 1996). By the time a program has existed long enough to be tested and thoroughly researched, technological advances have encouraged the use of different and faster programs. In view of the current situation, studies have shifted from the 1980s where much research was focused on differences between CALL and non-CALL environments to now, where research is more focused on the effectiveness of software and the choices learners make (Kolich, 1991).

### **Purpose of the Study**

This study describes second language learners' use of interactional options (e.g., a glossary usage) within *New Dynamic English*, a listening-based CALL program produced by DynEd International, Inc. The study then elaborates on the students' use of options and how it relates to the students' performances within the program. It draws upon correlational data

gathered from student interaction to analyze the effects of the software on the students' second language (L2) learning.

### **Rationale**

This study is useful for both theoretical and practical purposes. Theory in second language acquisition (SLA) suggests that learners need to interact with the target language to acquire it. Computer programs like the one analyzed in this study, which offer students options and allow them to interact with the language, may be able to help students begin to use the language effectively and draw closer to understanding how to use the language in actual environments (Harless, Zeir, & Duncan, 1999). CALL, because it facilitates tracking and accurately recording student actions within a lesson, offers researchers a medium with which to inquire about how students use various options and then make inferences about language learning. CALL research that identifies and evaluates the effectiveness of what learners actually do while they are using the software (e.g., Hegelheimer, 1998; Chun & Plass, 1996) will be most beneficial for increasing general knowledge as to how CALL can best be used.

Practically speaking, software developers should benefit from a corpus of empirical research to guide their improvement of the programs' effectiveness. As Hegelheimer and Chapelle (2000) have pointed out, few studies have analyzed the effectiveness of CALL programs that are actually being used by students. Though many studies have researched outcomes, and some studies have used descriptive analyses, few have researched real CALL environments to investigate their benefit by analyzing what actually happens within a lesson. More research has been done in artificial environments, but according to Hegelheimer and

Chapelle, “Although such experiments carefully model the desired cognitive characteristics for formal learning, critical elements of learner motivation and communicative language use are likely to be missing” (41). This study, because it analyzes a real language learning environment, contains these critical elements, and thus provides a specific guide for software developers and researchers. An understanding of this relationship may aid teachers and software developers in deciding which options are most beneficial to students and which options are less beneficial, and in what context. This may, in turn, help the students, teachers, and developers make decisions that will increase the effectiveness of the software, and validate the options the developers have already provided. To guide this research in describing the software and analyzing student interaction and performance, three research questions were constructed.

### **Research Questions**

These questions are used throughout the study and serve as an organizational framework for describing use of the program, for finding important relationships between student options and student performance, and for describing behaviors of top-performing students versus bottom-performing students.

1. Which options do learners use the most often, the least often, repeatedly, or only once?
2. Is there a correlation between (a) the learner’s choice of options during a lesson, time spent using the program, number and length of phrases heard, and (b) the learner’s performance in answering questions within the program?



3. What differences can be observed between the group of students who perform at the highest level and the group of students who perform at the lowest level? Which options appear to be used more frequently by students who perform at the highest level?

### **Organization of this Study**

This study begins by reviewing the literature in the area of SLA research and its interaction with CALL. Specifically, this review focuses on empirical research that addresses questions about what learners actually do while interacting with CALL programs and how this information can be used to guide SLA theory. The review begins by providing an introduction to how SLA and CALL can be used to complement each other. It then discusses studies that were performed in experimental controlled environments using artificial and semi-artificial languages as material. Then, it documents research that has examined CALL in more authentic environments using natural language. Still, most of the studies in this second category have analyzed environments designed, controlled, or at least influenced, by the researcher. The review concludes by addressing some of the reasons for conducting research in even more authentic — however less controlled — environments, giving examples from past research.

The following section, Chapter 3, describes the demographic data about the learners as well as the environment in which the program was used. This chapter also discusses the placement test used and the program and describes some of the lessons from which data for this research was collected. Then, it explains the data collection and how this information relates to the study. It ends with an explanation of the types of analyses used to examine the data.

Chapter 4 examines the results of the analyses and further discusses the modifications that were employed to reach the results that were attained. The final chapter suggests possible future considerations for CALL developers as they seek to implement programs that help learners to acquire the language with optimal efficiency. It also explains the study's specific limitations, and makes recommendations for future research in authentic CALL environments.

## CHAPTER 2. LITERATURE REVIEW

Though there is no lack of research in CALL environments, relatively few studies have empirically examined and analyzed learners' use of CALL programs that are currently being used. Many studies have investigated SLA in artificial CALL environments, which allow researchers to control variables (e.g., prior knowledge of the user), thereby increasing their internal validity. However, as Hegelheimer and Chapelle (2000) have pointed out, lack of learner motivation and distance from authentic L2 situations present a serious problem in these studies and limit their external validity. Nevertheless, the need for studies in more authentic learning environments exists not only because researchers should know what students are doing with programs currently available, but because CALL programmers may benefit from research for the improvement of existing programs and the development of new ones.

Within SLA and CALL research, there is a continuum that spans from completely *artificial* learning environments to completely *authentic* learning environments. In this chapter, I will define what I will use these terms to mean and then give several examples of studies which are performed in artificial or semi-artificial learning environments, *experimental* learning environments, and (more or less) authentic learning environments, along with explanations of the benefits and drawbacks of each category. The chapter will proceed with an explanation of some of the benefits of authentic learning research and conclude with a discussion of two areas that need further examination within the context of authentic learning research, namely modification of input and simultaneous audio-visual-textual input (e.g., video captioning).

At one end of a continuum are *artificial* learning environments, which I will use to refer to classes or private language study situations designed solely for researching SLA which use artificial or semi-artificial languages or pseudo-words. This is done in order to study acquisition of structures while controlling for subjects' prior knowledge of the input.

I will use *experimental* (as an intermediate category) to refer to classes or private language study situations designed by the researcher for the purpose of gathering data for analysis. I will use this to refer to environments that use natural or only slightly modified language, but may be – to some degree – in unnatural settings for the students, settings that would not occur without the researcher having influence on the activity of the students.

On the other end of the continuum, *authentic* learning environments, are classes or private language study situations where L2 learning is the primary purpose. Because of the CALL component these situations can be used for research, but they were not designed beforehand or adjusted in progress for the purpose of research. Therefore, the researcher has little or no influence on the students' activities throughout the course of the study. Most studies (including the ones cited here) have not been performed in a purely authentic learning environment, but approximate one when compared to more experimental or artificial studies. This shows the current state of this continuum, where several studies have been performed in highly artificial environments, but very few are performed in truly authentic learning environments.

### **Artificial Learning Environments**

Studies in CALL, perhaps to a greater extent than previous studies in SLA, enable researchers to control variables in hopes of increasing the reliability of their results. This has

been advocated and demonstrated in many cases (e.g., de Graaff, 1997; DeKeyser, 1995, 1997; Issidorides, 1988; Robinson, 1997; Yang and Givón, 1997) where the researchers used artificial languages or pseudo-words in order to have complete control over the type and amount of input. Because the researchers rely on the use of artificial input and environment, however, there are also drawbacks, as several researchers have also noted (Hegelheimer & Chapelle, 2000; Hulstijn, 1997; and Yang & Givón, 1997).

Besides the ability to control subjects' prior knowledge of the language being studied, artificial environments offer researchers the ability to study discrete individual variables. For example, de Graaff (1997) studied the effects of implicit and explicit instruction on morphology/syntax in eXperanto (a simplified version of Esperanto), and found that the group with explicit instruction performed better on immediate and delayed posttests of proficiency. He noted that the simplification and artificial environment allowed for the fast learning of the target structures, but admitted

...the more controlled the design and the more specific the learning task, the more we bear the risk of not studying L2 acquisition anymore, but only participants' capacity to carry out some kind of cognitive puzzle. Only a careful combination of the advantages of both realistic and optimally controlled L2 learning environments can provide real opportunities for studying the effect of instruction on L2 acquisition (272).

DeKeyser (1995) found similar results dealing with implicit vs. explicit instruction. In a later study, DeKeyser (1997) instructed subjects equally in the morphosyntactic rules of Implexan (a miniature linguistic system designed for the experiment). He found that eight weeks of practice in comprehension led to greater improvement in comprehension, practice in production led to greater improvement in production, and that practice in both led to improvement in both, but to a lesser degree than focused practice in one area or another.

However, he noted that, “no specific recommendations for teaching methodology should be made before the findings of this study have been put to the test in the more ecologically valid context of real second language classrooms” (215).

Yang and Givón (1997) performed a study to test the difference between instruction using a pidginized version and a fully grammatical version of Keki (an artificial language). Givón had hypothesized that vocabulary and grammar compete for learners’ attention in the early phases of acquisition and that learners exposed to a pidgin would excel in vocabulary and rapidly catch up in grammar after being exposed to the fully grammatical version. The hypothesis was rejected after results clearly indicated that there was no advantage for the pidgin group over the grammar group in lexical recall.

Issidorides (1988) found results similar to Yang and Givón’s when exposing different test groups to language without function words. She used a “twin” approach where she performed one study with Dutch speakers learning a miniature artificial language (MAL) and another on Greek speakers learning Dutch, to which they had never been exposed. This twin approach, advocated by Hulstijn (1997) because it deals with natural and artificial languages, has the advantage of increasing external validity, but possible drawbacks in ethical considerations. As Yang and Givón have explained:

From an ethical standpoint, the use of an artificial language, which has no obvious future utility for the L2 learner, allows the researcher to employ conditions that might inhibit learning of the L2 in the short term and perhaps even in the long term.... Take, for example, our study. We do not know how long the negative effects of pidgin input would last.... (189)

Thus, because artificial language research environments allow complete control over input, researchers should be cautious about choosing research methodology that can ethically be replicated in natural language environments.

Other methods of balancing between control of input and external validity have been demonstrated in studies like Robinson's (1997) study in which he used pseudo-verbs in ESL instruction to Japanese speakers. He tested four subject groups' abilities in grammaticality judgment tasks after they had been exposed to implicit, incidental, enhanced, and instructed conditions. He found that the reaction times were lowest and grammaticality judgments most accurate for the instructed learners, which had been taught to focus on form (Long, 1991). The other conditions were taught to focus mostly on meaning or some combination of meaning and form. Robinson notes the limitations of the lab environment by explaining that, "artificial verbs [may] have interfered with the process of interest in the incidental and enhanced conditions by making it difficult to process the sentences for meaning in the way natural language samples are processed."

Other artificial language experiments have also noted the inevitable change in learner motivation. Though several studies claim that the subjects were motivated (Yang and Givón, 1997), none claim that the motivation is the same as it would be in a natural language environment. Partly for this reason, researchers have also conducted experiments in natural language environments, even though they do not offer the researcher the same degree of control over input. In order to gain a closer application to real L2 environments, many researchers have examined experimental environments that use predominantly natural language, even though the environment may not be realistic (e.g., in a classroom setting).

### **Experimental Learning Environments**

Carroll and Swain (1993) have advocated the use of an experimental environment using natural language. In their study of feedback, which was not conducted in a CALL

environment, they tested 100 Spanish-speaking learners of English to determine which group produced the most correct responses dealing with alternation. They found that all the groups with feedback outperformed the group with no feedback and that the group with the most explicit feedback outperformed all the others.

Carroll and Swain defended the use of the experimental context for this study when they stated the following:

In the classroom context, the teacher sometimes presents data in the form of exercises that are misleading as well as grammatical “rules” that are inaccurate with respect to the knowledge to be learned. In contrast, experimental contexts can be constructed so that the learning objective is well defined, the data are homogeneous and ordered, and consequently the generalization of an induced organizing principle ought to be straightforward. If feedback does not work in an experimental situation, it is highly unlikely that it would work elsewhere. (362)

As is mentioned in the conclusion, however, this study was conducted in a short time span (1 week). One could argue that just because feedback may not work in a short experimental environment, this may not mean that, given the time span of a natural class setting, feedback could not be understood. Furthermore, CALL environments, on the other hand, because they offer researchers consistency of presentation to learners and have the capacity to dramatically reduce the negative aspects of authentic learning research *can* produce generalizable results.

Doughty (1991) has also conducted research in a natural language CALL context in which she pre and post-tested twenty ESL students’ abilities to understand relativization after they were divided into three treatment groups. She found that the control group (COG), which interacted with the same readings as the two instructed groups, improved to a significantly lesser degree than the others. The instructed groups, one of which was called



the “meaning-oriented instructional group,” (MOG) and the other the “rule-oriented instructional group” (ROG) both made improvements (see table 2.1). Thus, she concluded that L2 instruction did make a difference, but she did not determine which instruction condition was better for optimal SLA.

Table 2.1: Group mean gain scores by relativization category for MOG, ROG, and COG subjects (Doughty, 1991).

Group	subject	Direct object	indirect object	Object of a preposition	Possessive	Object of a comparison
MOG	11.53	29.73	35.69	39.96	23.58	26.03
ROG	20.86	32.12	24.08	31.69	15.61	24.80
COG	7.67	10.80	15.55	13.24	9.65	18.82

The studies in SLA and CALL that have adopted artificial or experimental learning environments have all to some degree or another used what Chaudron (1988) referred to as “the psychometric tradition” as their research paradigm. They have all exposed subjects to two or more treatments to determine the differential effects. Though this provides insight into specific variables, it removes the study from reality and the variables that will inevitably be present in a real L2 environment. Because of the required control on the experiment necessary for research that adheres to the psychometric tradition, this method is seldom used for more authentic learning research.

### **Authentic Learning Environments**

The studies that follow, because they have tended to focus on more authentic learning environments, and because they have not tested one group of subjects against a control group, have relied on interaction analyses in order to determine exactly how the learner

interacts with real L2 learning materials in a CALL environment. This type of research is necessary to determine how learners actually acquire language in authentic settings and to evaluate CALL materials that are already in use. Because these studies encounter real L2 environments, they have the capacity to study learners for more extended periods of time. Most of all, in many authentic learning studies, learner motivation should be near to equal what it would be in other real L2 environments.

This proximity to reality does come with the cost of lack of researcher control over the input — students have control over the input they choose, and this input will inevitably differ from student to student. However, CALL environments offer a consistency of presentation not possible with human instructors, who may respond differently from time to time to the same student requests. Finally, a CALL environment allows interference from researchers' observations to be minimized. Therefore, CALL facilitates useful authentic learning research that was not previously possible.

Hulstijn (1993) conducted research to determine the translation look-up behavior of 44 Dutch students of EFL under two conditions. He found that subjects did not differ significantly in their look-up behavior depending on whether they were assigned to answer comprehension questions or to write a global summary. In another part of the study, his research indicated that 38 other subjects tended to look up the meanings of non-inferable pseudo-words and pseudo-words relevant to answers to comprehension questions more than they did words that were inferable or irrelevant. In the third part, subjects who tested as having greater inferring ability did not look up fewer words than students with less inferring ability, but students with greater initial vocabulary knowledge did look up fewer words. Finally, Hulstijn made note of individual look up behavior and determined that (1) subjects

went through the reading linearly by paragraph at least once (meaning, they looked up words only in the order they appeared in the text), (2) ten subjects looked up nearly every word, and (3) nine subjects looked up none or only one of the eight minus relevant target words. His research used some methodology of the psychometric tradition in that it tested the effects of two different conditions. However, this study is different from the previous studies mentioned because it made use of a more authentic CALL environment and examined student behavior as it occurred during student interaction with the program.

Chun and Plass (1996) performed an even more authentic study in which they had 160 second year German students use an interactive multimedia CALL reading program called *Cyberbuch* for two consecutive class periods. The program used a reading of 762 words on 11 pages. 82 of the words in the text had been annotated and students were able to look up the words by clicking on the word and holding the mouse button down. The annotations included three types, textual English definitions alone, text with video, and text with pictures.

Their first research question was to determine the degree to which learners acquired vocabulary incidentally through reading in a hypertext environment. Their study reported a 24.1%-26.5% range for the subjects' performances on post-study retention measures. This represents the percentage of the time that students were able to provide the correct English equivalent words in a list of German words they were given immediately after the reading and two weeks later.

They also analyzed students' look-up behavior to find that the correlation between the annotation mode (text only, picture + text, or video + text) and recall. The static pictures + text had the most beneficial effect for recall of the vocabulary. This study explained the

benefit of research in a authentic learning environment: “the complexity of language acquisition is such that studying it requires simulating a realistic learning situation, which in turn requires attention to multiple factors and variables” (186).

Others (Davis and Lyman-Hager, 1997; Hegelheimer, 1998; Lomicka, 1997; Koren, 1999) have also used more or less authentic learning environments to examine the look-up behavior of readers who have access to various gloss or hypertext material. Each performed research with reading texts observing learner choices within real CALL multimedia environments and analyzing the relationship between learners’ look-up behavior and learner performance. As noted by Chun and Plass (1996), the importance of these studies is that they include all or most of the variables that are present in real L2 learning environments.

All of these studies can be classified into categories based on their interface with L2 teaching (Pica, 1997). Some research *coexists* with teaching, having little real application for teachers. Other research *collaborates* with L2 teaching, where teachers and researchers work toward similar goals of understanding SLA. Research that *complements* teachers, however, is research that (as Chapelle [1998] has stated) is the most beneficial for CALL. In this type of research, researchers work together with teachers to develop activities that best aid in L2 development and then test their effectiveness within the classroom environment. Whereas this type of interaction between teaching and research is unlikely in artificial language laboratory experiments using psychometric measures, it has the potential to become common in authentic interactionist CALL research because the expressed aim of such research is to understand what learners actually do with real CALL materials.

Authentic CALL research and its interaction with SLA research has been examined by Chapelle (1998). She outlined seven questions for empirically evaluating multimedia

CALL. These questions, drawn from hypotheses in current SLA theory, can be used to guide future research in CALL, especially research on software used in real L2 environments. The second question of the seven, “Do learners choose to see the modifications of linguistic input?” (30) holds special relevance for studies that have examined interactional options and learner choices within multimedia CALL programs. This modification “can consist of such features as simplification, elaboration, or added redundancy” (24). Research dealing with modification of input, which is discussed in the following section, has produced evidence for advantages of certain types of modifications over others, but much more research is needed in this area, as very few authentic learning and longitudinal studies have been performed.

### **Modification of Input**

Not only is there a need to understand whether or not learners choose to see modifications of input, but researchers may still benefit from continued research as to whether and to what degree learners benefit from modification of input. Perhaps more importantly, teachers and researchers still have much to learn about *what types of* modification of input are the most beneficial for learners.

Hsu (1994) studied the behavior of 15 beginning ESL students using a CD-ROM listening program that allowed them to request repetition, repetition with text, or use of a dictionary. She classified each of these as modifications of input and noted that each of the requests for modification was correlated with higher learner scores on post-tests. However, she found that the repetition with displayed text had the greatest effect in helping students to comprehend the information and that students believed repetition with text to be the most helpful.

Other studies performed in less authentic learning environments have also observed the effects of simultaneous audiovisual and text input. Garza (1991), Borrás (1993), and Borrás & Lafayette (1994) have all reported positive effects using video-captioning materials for increasing student comprehension and recall of information. The studies of video captioning materials were not based in multimedia CALL, but they did represent a modification of input in at least one sense: the input the experimental groups received was elaborated.

Past research in the area of elaborated input seems to indicate positive effects overall for simultaneous audiovisual and textual input. However, these studies all were restricted in that they were, at least to some extent, experimental. Hsu used an actual CALL program, but none of these studies occurred within the context of an actual classroom, with all the variables that occur. All of the subjects in each of these studies participated voluntarily in the studies on their own time, so learner motivation, although perhaps significantly greater than in artificial environments, was not under the same conditions as would be found in an actual classroom environment.

Therefore, strong need exists for studies which research CALL in authentic learning environments, including all the variables present in an actual classroom because they occur in actual classroom environments. CALL research that addresses modification of input and seeks to find answers about the most beneficial types of modification will be very valuable for SLA research and CALL research and development. Most specifically, studies that seek to find out more about the effects of simultaneous audiovisual and textual input in an interactive CALL environment may have important implications for future language teaching and research.

Therefore, the present study is important to the fields of SLA and CALL research because it analyzes the activity of students within the actual classroom environment. It also provides information about modification of input and the relationships between student requests for certain types of modification and student performance. This information is important for teachers who use CALL materials and software developers as well as for researchers.

### CHAPTER 3. METHODOLOGY

This chapter addresses the actual interaction examined in the study and is divided into 1) participants – the subjects who participated in this class; 2) materials – a description of the software used in the study; 3) procedure - the software's implementation and the environment in which the interaction took place; and 4) analysis – the analyses and hypotheses used to examine the interactions.

#### Participants

Data were collected from 94 female EFL students working with *New Dynamic English* at Zayed University in the United Arab Emirates whose ages ranged from 18 – 24, though most of the students were either 18 or 19. The class size ranged from 8 to 18 students, and there were 7 different classes. The students were all Emirati nationals and had five years of prior English instruction in middle and high school. However, most of the students were at the false beginner level (Mitchell, 2000a). All the subjects took the program's placement test (see Appendix A) before beginning and as a group averaged in the high beginner level. The mean score was 0.88 (SD = .44) on the scale where 0.0 represents a beginner, 4.0 represents an advanced non-native speaker, and 5.0 is an educated native speaker.

#### Materials

Though the software offers much potential for future studies which have a complementary relationship with second language teaching, the program was designed



primarily for second language acquisition in a classroom environment. The courseware is contained on 8 CDs with two CDs for each level. There are four levels ranging from beginning to advanced (See Appendix D for system requirements).

The students in this study were only working with levels 1 (for beginners and false beginners) and 2 (for low-intermediate students). The contents of levels one and two are displayed in Table 3.1. Each Unit contains several individual activities.

Within “Names and Places” (Level 1, Disk 1) there are the following activities: 1) *Hello Max and Kathy*, 2) *Where is France?* 3) *Who Speaks English?* 4) *Question Practice*, and 5) *Focus Exercises*. The first three activities introduce the characters and content of the unit aurally and include occasional comprehension check questions; the second two activities review ideas presented in the lesson through quiz. The units are organized around communicative goals and introduce vocabulary and grammar over the course of the program. For a description of the activities involved, see Appendix B.

Table 3.1: Contents of *New Dynamic English* (from Fitzgerald, et al. 1999)

Level 1	
Disk 1	Disk 2
Names & Places Jobs & Family Numbers & Time Review Exercises Video Interactions	Family Schedule Matrix Vocabulary Likes and Dislikes Review Exercises Video Interactions
Level 2	
Disk 1	Disk 2
Daily Activities Our World Locations Review Exercises Video Interactions	Planning Ahead Matrix Vocabulary Biography Review Exercises Video Interactions

Comprehensive guides for users and for record management accompany the software. In addition, there is a study guide for each level that is intended for extending the course content into classroom activity apart from the computer interface, and for explaining the use of the program's interface to the user. Figure 3.1 displays a description of the interactional options available to the user.









Icon	Function	Icon	Function
	<b>Record</b>		<b>Repeat</b>
	<b>Playback</b>		<b>Fast Forward</b>
<b>abc</b>	<b>ABC</b>		<b>Rewind</b>
	<b>Translation</b>		<b>Pause</b>
<b>Exit</b>	<b>Exit</b>		<b>Play</b>

Figure 3.1: from *New Dynamic English User's Guide* (14)

Figure 3.2 shows a typical screen from the program. The control bar at the bottom of the screen remains constant throughout and consists of eight buttons. The "Exit" button leaves the program or returns the user to the previous menu. The "microphone" button allows the user to record her voice for practice, and the "headphone" button plays back the user's voice so that she can compare it with the native speakers' voices in the program. The "repeat" button repeats the previous statement aurally, and the ABC button repeats the

File Options Speech Help

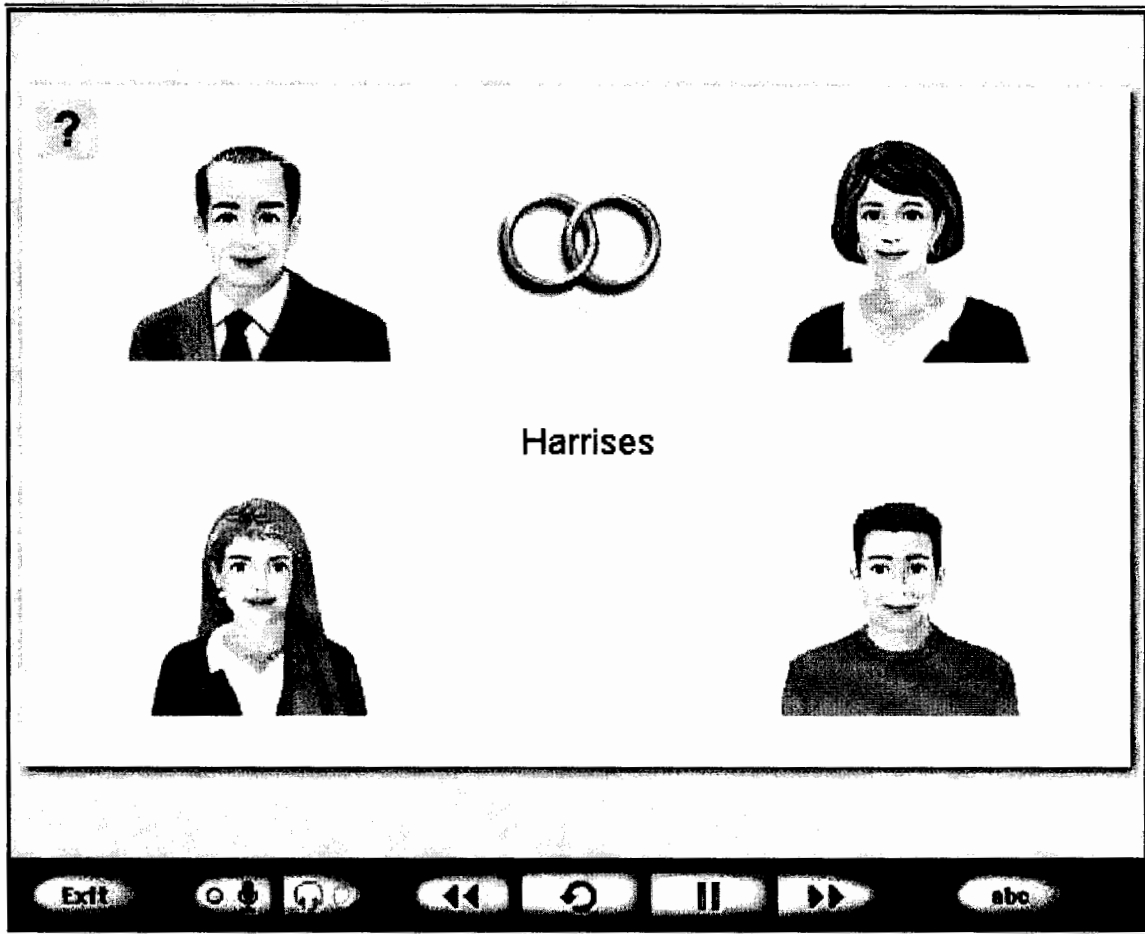


Figure 3.2: Screen display from "Family Schedule," (Level 1, Disc 2, from Fitzgerald, et al. 1999)

previous statement aurally with accompanying text. In addition to the options available in the control bar, the "options" menu offers users a glossary consisting of the words used in the unit. The glossary may be accessed by pulling down the options menu, clicking on "glossary," and choosing the word sought from an alphabetical list, or by clicking on a highlighted word in displayed text. The glossary provides a brief English definition for the word, several example sentences, and sometimes example diagrams or pictures. With the

exception of the “exit” button, all of these options are recorded in the records manager and used to analyze user options in this study.

The “rewind” and “fast forward” buttons, which repeat the previous sentence or play the next sentence (respectively) with accompanying text, are not recorded in the records manager and are therefore unavailable as a variable for entry in this research. Also, the “pause” button stops the program and allows the user to think about the question before answering. Use of this option is also not recorded, but if the student does not answer a question in a given amount of time, the program prompts the user, “Are you there?” and allows her to click “yes” or “no,” and gives the user more time regardless of the answer. So while use of the pause button may benefit the user, it may not have a significant effect in predicting the accuracy of the user’s answer. However, because this information was not recorded, there is no way of knowing the effect this option may have had.

Beyond these five user-chosen options, the program’s records manager recorded nine other variables which were important for this analysis (see Table 3.2). The first of these variables is the students’ scores on the placement test. The second is the total number of questions the students were asked during the course of the program, and the third is the total number of questions the students answered correctly while working with the program. Fourth is the amount of time the learners spent with the program.

The fifth and sixth variables consist of records of students’ attempts at speech-recognition (SR) activities and the number of times the learners perform correctly on the first attempt of a SR activity. These activities ask the learner a question and the learner constructs the correct response by filling in the blank or unscrambling phrases to make a sentence or question.

Table 3.2: Recorded variables

<b>Variable Name</b>	<b>Explanation</b>
Microphone (record) button	Each time this button was pressed; it allows the user to record her voice for practice
Headphone button	Each time this button was pressed; it allows the user to listen to her own voice (for comparison with the native speakers voices in the program)
Repeat button	Repeats the previous sentence or question
ABC button	Repeats and simultaneously displays the text for the previous sentence or question
Glossary	Displays an English definition and textual examples
Placement test scores	Consolidated to one single score (see pp. 27, 28)
Total questions	The number of questions a student was asked during the course of the entire program
Questions correct	The number of questions a student answered correctly on the first try.
Total time	The total amount of time users spent with the program
SR attempts	The total number of times a student attempted a speech-recognition activity
SR correct	The number of times the computer recognized the learner's first attempt at any speech recognition task
Total phrases	Total number of phrases the learner heard throughout the course of the program
Total phrase length	Total length of the phrases the learner heard (in seconds)
Shuffler level	Adjusts difficulty level based on student performance (see p. 25)

For this option, the records manager only counted actual student utterances during an activity. If a student was using a lesson that had SR as a possibility, but chose to answer via clicking on the correct answer rather than reading it aloud into the microphone, then it was not recorded as an attempt. The records manager also noted each time the program recognized the student's first attempted utterance. If the student was answering a question by making the utterance, and she answered correctly, then it was recorded as both a correct answer and a first-trial recognition of the utterance. If the student did not answer correctly,

but the program recognized the utterance on the first trial, it was recorded as a first-trial recognition, but not as a correct answer. Some activities, such as sentence-reading practice, allowed for students to practice SR that was not counted as a correct answer to a question, but was recorded as a SR attempt.

In other words, all questions posed to the students in the program were counted, all answers students gave which were correct on the first try were counted, and some of those answers were also SR attempts. In the event that the SR attempt was a correct answer to a question on the first try, it was also counted for being recognized on the first try.

The program also recorded the total number (the seventh variable) and length (the eighth variable) of phrases heard within each lesson. The final variable is the *Shuffler Level*. Material in the Records Manager Guide that accompanies the software describes its function:

[The “Shuffler”] operates to open or shut paths within a lesson. As a student answers questions, the *Level* is adjusted up or down, depending on the success rate of the student. The maximum *Level* is 3.0. When the *Level* is in the range of 0-2.0, a lesson is only partially open, restricting a student’s access to various questions and paths within the lesson...As students become comfortable with the initial material and begin responding successfully, the *Level* automatically rises, and additional material, generally with more difficult grammar and vocabulary, is opened up to the students. Should the student begin to have trouble at a higher *Level*, the *Level* will automatically fall, allowing students to focus on a more limited selection of the material before opening up the rest of the lesson. The actual effect of the *Shuffler* varies from lesson to lesson and course to course. (pp. 26-27)

### **Procedure**

After the students took the placement test sometime in late February, depending on which class they were in, they were to interact with the program one day per week from late February to late April of 2000. Upon review of the log-in data, it became evident that there

was significant variation in time on task per student. Through email correspondence with the program's director following this discovery, the researcher was able to ascertain that this was because some of the teachers had the students interact with the program each week, but some teachers were not comfortable with the electronic interface and used the program less often. The director also provided information that, because of a network failure in the third week of the class, some of the teachers abandoned the program at that point rather than continuing for the duration of the course (Mitchell, 2000c).

The average time each student spent using the program over eight weeks was 141 minutes ( $SD = 64$  min 22 sec). The amount of time individual students interacted with the program, though linearly distributed, varied from 30 minutes to just over five hours. However, only 4 students spent less than 55 minutes with the program. 75% of the students used the program between 64 minutes and 216 minutes (3 h., 36 min.).

The director also informed the researcher that the teachers were all given the same instructions for introducing the program to the students, but individual classes may have received more or less instruction in using the software. Students' attendance was checked, but there was no grade involved with the amount of student interaction or performance level while using the program.

In order to increase the speaking practice for each student, students were instructed to listen to the presentations and repeat each sentence that they heard. Using the program was a mandatory section of the course, but students were allowed to work at their own pace and choose which exercises they wanted to do.

The repeat button was introduced to the students as a means of telling the program to repeat the previous statement, and all students were required to use the microphone (record)

button, which records the student's voice for practice. Students who did not use the microphone button in a week's lesson were required by their teachers to use it five times per exercise the following week, repeating each exercise, if necessary, until this had been accomplished. This had significant effects on the study, as much of the data attest. The other options, such as the glossary, and the ABC, fast forward, and rewind buttons were not formally introduced to the students. Students who used these options discovered them on their own (Mitchell, 2000a). These options were not introduced because of the role this class was intended to play in developing the students' listening abilities (Mitchell, 2000d).

Instructions that accompany the software indicate that students who score below 228 on Part 1 of the Placement Test should not take Part 2 (see Table 3.3). Many of the students, however, scored lower than 228 in the original test but higher than 31 in the second test. In such cases, students were scored for placement for this research according to their Part 2 scores. If students scored below 32 on Part 2, their Part 1 score was used for this analysis. Tables 3.3 and 3.4 show the placement test scores and corresponding levels.

Table 3.3: Test Score Part 1 placement levels from *New Dynamic English Placement Test Guide* (12)

Test Score <i>Part 1</i>	Placement Level	New Dynamic English Level
0-31	0.0	1A
32-71	0.2	1A
72-111	0.5	1B
112-180	0.7	1B
181-227	1.0	2A
228-270	1.2 or higher (Take Part 2)	Placement test, Part 2



Table 3.4: Test Score Part 2 placement levels from *New Dynamic English Placement Test Guide* (13)

Test Score <i>Part 1</i>	Placement level	New Dynamic English Level
0-31	1.2 or lower	Use Placement Test, Part 1 result
32-45	1.2	2A
46-87	1.5	2B
88-115	1.7	2B
116-215	2.0	3A; 3B
216-283	2.5	4A
284-300	3.0	4B
301-314	3.2 or higher	

### Analysis

In order to examine how individual students used the program, the data, which consisted of 1811 records of student log-ins, were consolidated into subtotals for each student. These subtotals were then used in the subsequent analysis to reveal the various relationships student performance had with student choice of user options, while controlling for student level. Data collected from three students were discarded as each was recorded to have used the program for less than 15 minutes over the course of the eight weeks (this reduced the number of log-ins represented from 1811 to 1795).

Though the materials that accompany the software instruct the users to take the second part of the placement test only after placing in the “1.2 or higher” of Part 1 (see p. 27), many subjects with scores less than 1.2 nevertheless took Part 2. It was therefore necessary to determine which score and level should be used for each student. If students took the second placement test and scored 31 or below, their level was recorded according to their score in Part 1. However, if students who did poorly on Part 1 took Part 2 and scored

32 or above, the higher score was used to determine those students' levels for the purposes of this research (see tables 3.3 and 3.4).

Because students were not tested comprehensively in their listening and speaking proficiency upon exit, the only measure of performance was taken from student responses that were part of the program itself. This was done by finding the percentage of correct answers to questions within the program. This percentage was then used as the measure of student performance in the Pearson correlation matrix (see pp. 37-38) and as the dependent variable in the multiple regression analyses (where the all the other recorded options were recorded as independent variables, see p. 37 and 39).

Each of the user options within the program was observed to determine its frequency of use by each student. This addressed research question 1 (p. 3), which focused on describing student behavior in using the program. The following were considered options for the user: the ABC button, the repeat button, the glossary, the microphone button, the headphone button, and a speech recognition (SR) task attempt. Hypothesis 1 (H1) was constructed based upon past research by Bland, et al. (1990) and Hsu, Chapelle, and Thompson (1993): this research was expected to reveal that students would use some options more than other options and that this choice would vary from student to student.

Research question 2 (p. 3) dealt with the correlation between students' choice of options, student time on task, other variables (e.g., the Shuffler Level), and student performance (measured by percentage of correct responses while using the program). Placement test scores were also included in the analysis as a means of controlling for the students' level of proficiency upon beginning work with the program. Based upon the past research by Hsu (1994) and Borrás & Lafayette (1994), Hypotheses 2-4 were constructed.

Research was expected to find that the difference in choice of options for each student would have an effect on their performance and that this effect would be observable and statistically significant (H2). H3 postulated that all features of the program would have beneficial effects on student performance, but not all features would benefit student performance to the same degree. I also predicted that (H4) time spent with the program would enhance student performance, or that student performance would improve with practice over time.

To test these hypotheses, all the relevant data<sup>1</sup> were subjected to four different statistical analyses. The data were entered into a Pearson product-moment correlation matrix, a multiple linear regression analysis, a backward elimination stepwise multiple linear regression analysis, and a forward selection multiple linear regression analysis (Bowerman & O'Connell, 1990).

I began by looking for significant relationships between performance and the other individual variables (H2-H4). For this purpose, the Pearson product moment correlation matrix (see pp. 38-39) was used to provide results indicating the tendency for data within each of the variables to be related to each other. For example, if student percentage of correct answers rise alongside repeat button use and fall as repeat button use diminishes, the correlation would show the magnitude of this relationship and provide information (a p-value) about whether or not the two are related.

This analysis, however, only shows relationships between isolated variables without controlling for the other variables. In an interactive environment with so many options and

---

<sup>1</sup> The data included the students' percentage of correct answers; total time on task; total questions heard; questions answered correctly on the first attempt; the shuffler level; the students' placement level; use of the ABC, repeat, microphone, and headphone buttons; glossary usages; speech recognition (SR) attempts; SR correct trials – times the computer recognized the first SR attempt in a question; number of phrases heard; and the total length of phrases heard.

possible correlations, it is very important to control for the other variables. For an example of why this is important, consider the time in minutes and the microphone button. If students who use the program for a longer total time always use the microphone button more than students who use the program for a shorter period of time, then it cannot be determined from the correlation matrix which variable (time or the microphone button) is more important for predicting successful performance; both will show a positive correlation with performance.

To further investigate H2-H4 while controlling for student placement level and the interrelationships of all the variables with one another, multiple linear regression analyses were used to estimate the values of the dependent variable (the percentage of correct answers) as a linear function of the many independent variables (e.g., the repeat or ABC button). The assumptions of a multiple linear regression model are as follows: The errors are distributed normally, the observations are selected randomly and independently of each other, and the model residuals exhibit constant variance. Linear regressions enabled this study to determine the partial contribution of each independent variable toward predicting variation among students in percentage of questions correct (the dependent variable). Through the use of regression analyses, it is possible to predict whether and to what degree use of a specific option has negative or positive effects on student performance (Bowerman & O'Connell, 1990).

Since some of the available options were never used by many of the students, the independent variables did not all exhibit evenly distributed data. This prevented the model from satisfying a key assumption of linear regression models (that is, that errors will be distributed normally). For this reason, some variables (see Appendix C) were re-coded into ranges. This was done to make skewed distributions more consonant with the assumptions of

linear regression, by reducing the influence of the disproportionately large rate of non-usage, and to reduce the effect of large-valued outliers (e.g., a student who had 100 more usages of the headphone button than the next highest student).

In addition to the original multiple linear regression model, two other models were run to gain insight about the most accurate and efficient predictors of the dependent variable. This was done by first running a backward elimination stepwise multiple linear regression model, which discarded variables whose p-values were under 0.10 one at a time. The second model used forward selection stepwise multiple linear regression to accept variables whose p-values were under 0.05 one at a time. These models operated under the same assumptions as the original model, but through discarding inefficient predictor variables became more efficient models that displayed usable results. For this reason, the latter two models were used to show the results of this section of the study (see pp. 37, 39-40, and Appendix C).

Finally, in order to address research question 3 (p. 4), Hypothesis 5 (H5) was constructed. It predicted that there would be observable differences in choice of options between two groups of students: those whose percentage of correct answers ranked in the top 30 of the 91, and those whose percentages ranked in the bottom 30. This part of the analysis helped to demonstrate the general characteristics of students with higher and lower percentages as groups and to find out if there was a noticeable difference between the two groups. It also served as a means for verifying the results of the correlational analyses, providing another way of looking at the data.

## CHAPTER 4. RESULTS

This chapter explains the results of the analyses and addresses the three research questions presented in Chapter One and the hypotheses presented in Chapter Three. Chapter Five relates chapter 4 of the present study to past research in the areas of interaction analysis, modification of input, and video captioning.

### Frequency of Option Selection

The first research question (“Which options do learners use the most often, the least often, repeatedly, or only once?”) was aimed at determining whether learners use what is available to them when presented with an interactive CALL program that has many interactional options. Figure 4.1 gives the mean total usage per student for each option.

The first hypothesis (p. 29) was strongly supported; not only did choice of options vary from student to student, but some students favored certain options almost to the exclusion of others. For example, few students used both the “ABC” (repeat of aural input accompanied by displayed text) and the “repeat” (aural input alone) buttons often.

In order to investigate the degree to which students favored ABC or repeat to the exclusion of the other, analysis of the mean difference between the two options was undertaken. I found the difference between the two choices for each individual student and then found the mean for the entire set. The total use of each option was found for each student, and the mean use of the ABC button was 19.1. Though 46 students (50.5%) did not use the ABC button, those who did use it showed an average of using it more than once (mean = 38.64), and 25 students used it more than 20 times. Although use of the repeat

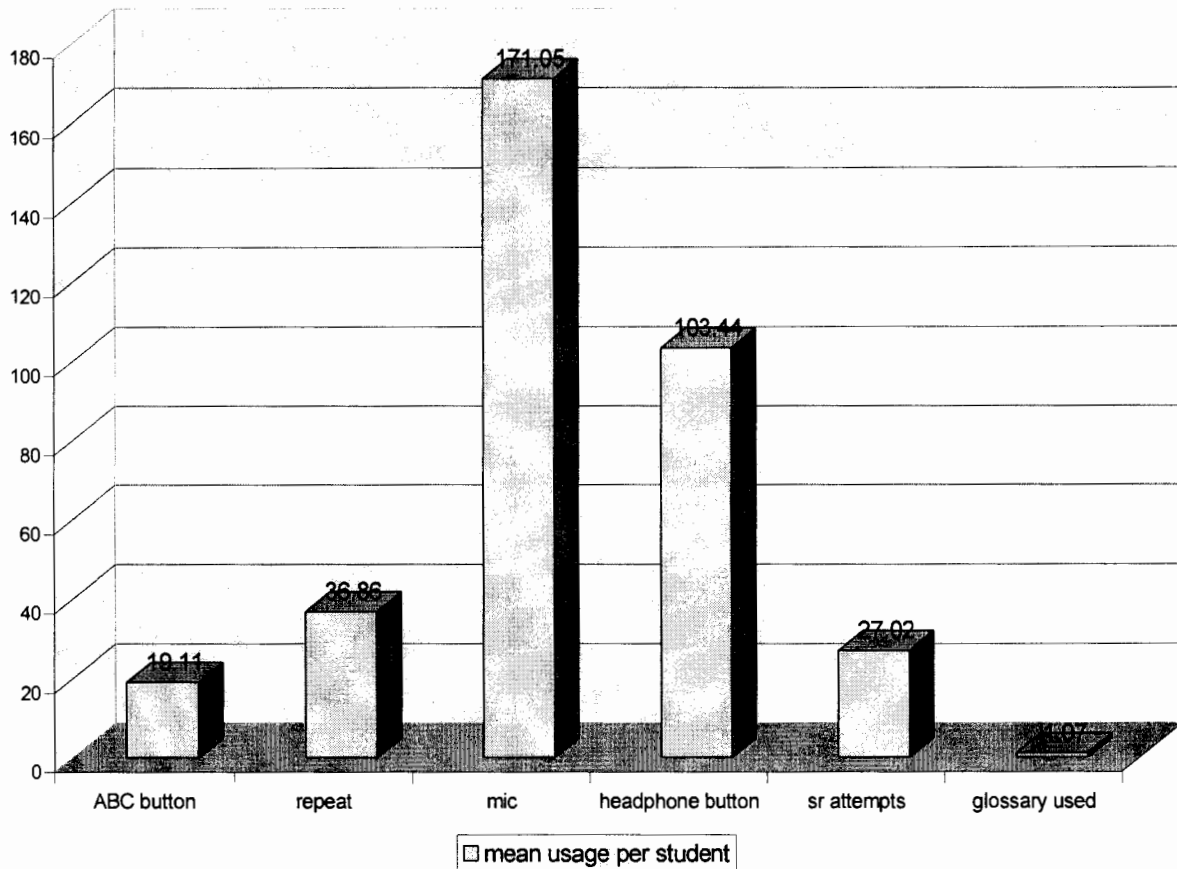


Figure 4.1: Mean usage of options per student

button was much more evenly distributed, 15 students still did not use it at all. The mean for total repeat button use was 36.86. The reason for higher repeat button use can be attributed to the fact that this button was introduced to the users, whereas other options for modification of input (e.g., the ABC button and the glossary) were not (Mitchell, 2000d). The mean difference between the total usage of the ABC button and the repeat button was 48.31. This indicates that students who used the ABC button often tended not to use the repeat button and vice versa. This result may have been completely different had the students been equally introduced to both buttons.

The use of the glossary, on the other hand, was not as apparently related to ABC button or repeat button use. The use of the glossary was quite interesting, because 75 students (82.4%) used it only once in the entire course of their study (see table 4.1). Perhaps this low level of glossary usage is due to the fact that the glossary was not introduced to the students, and because students were required to click on the options menu and find the word in an alphabetical list. This extra step may have deterred students from frequently referring to the glossary. It is likely that students were unaware that clicking on a highlighted word in the text gave immediate access to the glossary entry for that word. This lack of fast access may be the cause for seldom-repeated glossary use. Since most of the dialogue in the program was unlikely to be viewed textually, it may have been difficult for students to access the glossary and find a word in the text without knowing its spelling. Having teachers introduce the highlighted-text-click access to the glossary or adding a glossary button to the menu bar (rather than the pull-down options menu) may effect more frequent glossary use.

Table 4.1 Glossary Use Frequency

Number of Times the glossary was accessed	Frequency (number of students)	Percent (of the students)
0	7	7.7
1	75	82.4
2	6	6.6
3	2	2.2
4	1	1.1
Total	91	100.0



The microphone button, which recorded the student's voice for practice, was used far more frequently than any other option (mean = 171); only two students did not use it, and three students used it more than 500 times. The headphone button, which played back the student's voice, was used less than the microphone button but much more than the other options (mean = 103.4). Prominent use of these two options is attributable to the teachers' requirement that the students use the microphone button and practice saying each sentence that they heard. These options allowed the students to record their voices while they were practicing and then compare them to the native speakers' voices.

It is perhaps unlikely that frequent use of these options would have occurred had students not been required to use the microphone button, because use of these options did not provide scaffolding for answering the questions in the program. They did not represent modifications of input, but allowed the students to evaluate their own output alongside the aural language offered by the program.

Twenty-nine students (31.1%) did not attempt a SR task. Though some of those students did do activities that allowed SR answers, they did not opt to use the SR to answer it. For the students who did attempt SR, the mean number of attempts per student was 39.7. For these students, the mean for recognized first-attempts was 13.7.

### **Statistical Results**

Research Question Two sought to find out if there was a correlation between (a) the learner's choice of options during a lesson, time spent using the program, number and length of phrases heard, and (b) the learner's performance in answering questions within the program. The data regarding these relationships were examined to test H2-H4 (p. 29). H2,

which predicted that the difference in choice of options for each student would be related to their performance and that this relationship would be observable and statistically significant, was supported. Statistical analyses revealed strong relationships between some of the options and student performance.

### **Pearson Product Moment Analysis**

Table 4.2 indicates several significant relationships between options and student performance. Most notably, total time spent with the program, the repeat button, the microphone button, and the total number of phrases heard all show positive relationships with performance, and use of the ABC button shows a negative relationship with performance. The ABC button also appears to be negatively related to student performance on SR tasks, whereas attempts of SR tasks shows a positive relationship with SR performance. It is worth noting that these correlation values only represent relationships between individual variables. Not all the relationships should be considered significant at the same time. After applying a Bonferroni correction to control for the large number of relationships (dividing the significance level of .05 by 66, the number of relationships represented), the 95% protected level of confidence was found to be approximately .00076. With this protection level in place, only the ABC button, the shuffler level, and the placement level were left as simultaneously significantly related to percentage of questions correct.

### **Multiple Regression Analyses**

After preliminary multiple regression analysis, data were recoded to bring the data in the independent variables closer to normality (see Appendix C for a discussion of the

Table 4.2: Pearson Product Moment Correlation Matrix (continued on p. 39)

	% of questions correct	Total Minutes	ABC	Repeat	Glossary
% of questions correct	1.00 .				
Total Minutes	.349*† .001	1.00 .			
ABC	-.360*† .000	-.217* .039	1.00 .		
Repeat	.257*† .014	.496* .000	-.093 .380	1.00 .	
Glossary	-.089 .404	-.089 .400	.053 .617	-.114 .281	1.00 .
Mic	.299*† .004	.715* .000	-.051 .633	.429* .000	-.015 .889
Head-phone	.084 .429	-.217* .039	-.095 .370	.376* .000	-.027 .800
SR attempt	.186 .077	.584* .000	-.281* .007	.195 .063	-.129 .222
SR % Correct	.033 .754	.145 .170	-.225*† .032	.109 .303	-.025 .814
Number of phrases heard	.335*† .001	.922* .000	-.127* .039	.429* .000	-.044 .676
Placement level	.452* .000	-.038 .719	-.133 .210	-.107 .311	-.123 .245
Shuffler level	.695* .000	.373* .000	-.264* .011	.205 .051	.075 .482

Note: The top number in each cell represents the Pearson Correlation; the bottom number represents the 2-tailed significance level. Total minutes = total time students spent with the program; ABC = total times ABC button was pressed; Repeat = total times the repeat button was pressed; Glossary = total number of times the glossary was accessed; Mic = total times the microphone button was pressed; Headphone = total times the headphone button was pressed; SR attempt = total number of speech recognition attempts; SR% correct = percentage of first time trials that the program recognized; Number of phrases heard = total number of phrases heard while using the program; Placement Level = student score on the placement test; Shuffler level = the program's mechanism for adjusting the difficulty of the questions for each student; † = Significant correlation between options and performance.

\* $p < .05$

Table 4.2 (continued)

	Mic	Head- phone	SR attempt	SR % Correct	Number of phrases heard	Placement level	Shuffler level
Mic	1.00 .						
Head-phone	.626* .000	1.00 .					
SR attempt	.318* .002	.277* .008	1.00 .				
SR % Correct	.013 .899	.033 .760	.351*† .001	1.00 .			
Number of phrases heard	.666* .000	.520* .000	.571* .000	.193 .067	1.00 .		
Placement level	-.041 .698	-.175 .097	.057 .590	-.043 .682	-.082 .437	1.00 .	
Shuffler level	.273* .009	.109 .306	.056 .599	.089 .402	.441* .000	.235* .025	1.00 .

preliminary models and recoding of the variables). In the revised regression model ( $F = 11.149$ ), the shuffler level ( $t = 6.632$ ,  $\text{sig.} = .000$ ) and placement level ( $t = 3.566$ ,  $\text{sig.} = .001$ ) were the only significant predictor variables. This indicated that the program's placement test (Appendix A) and shuffler, the mechanism for raising and lowering the difficulty level for the students depending on their performance (p. 25), both accurately predicted student performance on the tasks.

In order to further investigate H2-H4 (p. 29), the data were re-entered into a backward elimination stepwise multiple regression model (which removed variables with significance levels over or equal to 0.10 one at a time) to determine if any of the other variables could be acceptable as predictors. In addition to the shuffler level and placement level (both significant at .000), this model ( $F = 35.458$ ) found the ABC button to be significant at .024

and the repeat button significant at .018. The increased value of the F-distribution indicates that this model was more efficient because it used fewer variables to predict the relationship with the dependent variable. These findings were confirmed further in a forward selection stepwise multiple regression model which accepted variables whose p-values were less than or equal to .05 one at a time, as indicated in Table 4.3.

Table 4.3: Backward Elimination Multiple Linear Regression Analysis Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
(constant)	62.402	2.802		22.269	.000
Shuffler	12.740	1.686	.544	7.554	.000
Placement level	4.496	.971	.321	4.630	.000
Repeat button	1.141	.475	.165	2.402	.018
ABC button	-1.954	.851	-.158	-2.295	.024

H2 (p. 29) was supported by the data; the correlational analyses generally indicated that options students chose did have significant relationships with their performance. Hypothesis 3, which stated that all options would have beneficial effects, but not to the same degree, was partially supported. While most of the options seemed to be weakly correlated with good performance in answering questions, the ABC button was negatively correlated with performance.

Hypothesis 4, which predicted that time on task would be positively correlated with successful performance, was partially supported. In isolation, total time in minutes showed a significant correlation with percentage of correct answers (.349,  $p < .05$ ). However, in the regression models that controlled for the interactional options, time spent with the program was found to have a weak ( $t = .347$ ) and non-significant ( $p = .730$ ) correlation with student performance. This indicates that, while time spent using the program did tend to have a positive relationship with student performance, time spent with the program was not an accurate predictor variable for successful performance. Since students who spent more time with the program tended to use options that aided them in answering questions correctly, use of the options is a more accurate predictor than mere time spent with the program. The fact that controlling for the interactional options available to the students factored out the observable benefit of time spent with the program may indicate that the interactional options were very important in helping the students to interpret the language in the program.

### **Group Ranking Analysis**

Research Question Three was designed to address the differences that could be observed between the group of students who performed at the highest level and the group of students who performed at the lowest level. It also sought to find which options appeared to be used more frequently by students who perform at the highest level. H5 (p. 32) speculated that there would be observable differences in choice of options between those two groups. The highest performing 30 students were compared with the lowest performing 30 students with percentage of correct answers to questions as the only criterion to determine performance level.

Table 4.4 shows the total number of times each option was chosen by the two groups. Most notably, the figures in the table confirm the findings of the regression models, showing that the ABC button was much more frequently used by the lowest-performing group, and that the repeat button was used much more frequently by the highest-performing group. Other options do not appear to have as significant a relationship with high or low performance. The data in this table also confirm the correlational data by showing that, while the students in the higher-performing group did tend to spend more time using the program, the options chosen helped to differentiate the two groups more than time on task. The table helps to explain how time on task was found to have an effect on the learners that was not statistically significant. The students who spent more time with the program inevitably used the options available to them more.

Table 4.4: High and Low Performance Groups

	HIGH GROUP			LOW GROUP		
	(SUM)	Per minute <sup>1</sup>	(AVG)	(SUM)	Per minute <sup>1</sup>	(AVG)
total questions	8434	1.62		5170	1.39	
Questions correct	7398	1.42		3897	1.04	
% of questions correct	2630.9		87.7	2233.7		74.5
<b>Minutes</b>	5155			3468		
<b>total time</b>	(85 hr. 55 min.)			(57 hr. 48 min.)		
ABC button	356	0.071		676	0.349	
Repeat	1688	0.304		587	0.118	
Microphone button	6653	1.280		4142	1.14	
Headphone button	3580	0.642		2729	0.743	
SR attempts	887	0.168		752	0.189	
SR correct 1 <sup>st</sup> time	373	0.068		223	0.054	
%SR correct 1 <sup>st</sup> time	865.336		28.84	491.346		16.38
Glossary used	29	0.006		36	0.014	

<sup>1</sup> per minute values indicate the mean of the number of events per minute for each student in each group, rounded to the nearest thousandth.



## **CHAPTER 5. DISCUSSION AND CONCLUSION**

Results from the present study demonstrate both the benefits and drawbacks of studies in authentic learning environments. Though there were many factors that were not controlled, the study yielded findings that can provide insights for L2 researchers, teachers, and software developers. Stated limitations of the present study may also provide guidelines for L2 research and development. Because of its clear conflict with findings in previous studies, this study has produced questions that seek answers in future authentic learning research environments. This type of environment has great benefits for researchers, teachers, and software developers.

### **Insights for Researchers, Teachers, and Software Developers**

#### **Insights for Researchers**

The data gathered from Research Question One corroborate earlier findings by Hsu, Chapelle, and Thompson (1993) and Bland, et al (1990), which show that learners do not access all the available information when working in an interactive CALL environment. As mentioned earlier, even though the repeat button was the only option for modification of input that was explicitly introduced to the students, 15 of the 91 still did not use it. As Chapelle (1998) stated, "A CALL program can be constructed to allow for beneficial interactions, but that program is only as useful as the interactions that actually occur when particular learners make use of it" (31).

Though Chapelle's statement was addressed at determining the effectiveness of the interaction within CALL programs, it also serves as further justification of this type of

research in an authentic learning environment. Subjects within the classes in this study undoubtedly behaved differently than might students in another study, but they were interacting in an authentic learning environment. Further research using this same program and others like it in different contexts of instruction is likely to offer researchers much more information about which interactional options learners use and do not use; and deciding what options learners are likely to use in a learner-controlled environment is of key importance.

Further research in this area may prove important for showing students' use of options varying across language ability levels, backgrounds, and environments. Advanced students may use these options very differently than the students in this study used them. In the same way, students from various language backgrounds or studying in English-speaking countries may use interactional options differently than the learners in this study.

It is also important for researchers to find out the reasons why students of various situations do not use all the specific options available to them. In the present study, for example, it appears as if the glossary was less accessible than some of the other options. This could be one reason for the learners' lack of frequent glossary use. However, there may be other reasons; the learners in this study were not introduced to the glossary, but they were also mostly low-level learners who may have had limited reading abilities. Had this function been introduced, the students may have felt more comfortable using it, but there is no way of knowing this information in the current study. Researchers may benefit from interacting with teachers and learners in ways that can influence the learners to be more acquainted with all of the interactional options. In addition, other methods of research such as think-aloud protocols and interviews with students may help to reveal why students chose not to use certain options. This relates to Pica's aforementioned (p. 15) assertion that research and

teaching may benefit from research methods that enable the two to complement each other (1997).

### **Insights for Teachers**

In the same way that researchers may benefit from interacting with teachers, teachers may gain valuable insights from interaction with researchers and from teacher-conducted classroom research. The present study yielded results that showed a negative relationship between ABC button use and success in answering questions. In a different classroom, however, teachers or researchers may come to entirely different conclusions. While this is one limitation of authentic learning research, it is also a distinct advantage. This study can be replicated in other classrooms where it may provide classroom-specific or learning environment-specific information about which options or combinations of options offer students the most efficient strategies for improving performance.

The subjects in this research were required to use the record button during each lesson and they were also instructed to repeat each sentence they heard while using the program. They were also instructed in how to use the repeat button, and the program itself at times encourages use of the repeat button, but use of this button was not required. Teachers and researchers may gain important insights from experimenting with requiring the students to use other options or certain combinations of options. Much more research is necessary to determine what forms of options are most beneficial to students of different levels and backgrounds (Robinson, 1991).

The learners in this teaching environment were required to use the program, but were not held accountable for their performance. Though their performance was recorded, it did

not have an effect on their grade. Stakes raising may have an effect on learners' use of the program, and this could be demonstrated through research. This is another factor that may vary across classes or even from student to student, but teachers may benefit from research that identifies the effects of grades being placed on student choice of options and performance while using the program. Classroom-specific follow-up assessment can allow teachers to make decisions about which options are most beneficial for their teaching philosophies and the individual students' needs.

Finally, teachers may benefit from extensive training and introduction to all the options within the program. This may be done through workshops or seminars that provide detailed information about each of the interactional options and current research on the effects those options may have for certain learners. Teachers who acquaint themselves well with the software's user's guide and teacher's guide should have deeper insights into the most effective use of the program. However, because some teachers may be reluctant to read user's manuals, on-line tutorials for both teachers and students may help teachers in obtaining knowledge about the program's most effective use.

### **Insights for Software Developers**

Next to teachers who are committed to explaining the program's use to the students, on-line tutorials may be the most effective way to ensure that students know the options that are available to them. This tutorial may be a mandatory or optional beginning segment of the software's program that could demonstrate the use of each option to the students and provide examples that show the students what to expect when they choose certain options.

After taking the necessary steps to make students aware of all the options available to them, developers can further benefit researchers and teachers by providing mouse-click data for all the options within the program. The present study had a wealth of information about students' interactional choices, but was limited because data about the fast forward, rewind, and pause buttons were unavailable. Though significant results were attained in the current study, lack of knowledge about the other options may cause concern for researchers. More importantly, future research may indicate that use of these options is related to student performance in one way or another, giving teachers further guidance in the use of the program and researchers more knowledge about interactional options.

In this study, the glossary was very seldom used more than once. One way for researchers to find out if this is more related to the glossary's ease of use or the students' introduction to the glossary would be for the developers to experiment with placement of the glossary in the control menu at the bottom of the screen. The current glossary is accessed two ways, through the pull-down options menu, which brings the students to an alphabetical list of all the words in the program, or through clicking on a highlighted word when the text is displayed, which brings the students directly to the entry for that word. There are two alternative possibilities which may warrant experimentation. First, if the glossary was placed in the control bar at the bottom of the screen and made equal (at least in presentation) with the other options, this may effect more frequent use. Also, the glossary could be configured to offer students a scroll menu not of all the words in the lesson, but of only the words from the previous several sentences, this may also give students more incentive to use the glossary.

Finally, the layout of *New Dynamic English* seems very intuitive and is constructed in a way that most students would be familiar with (Fitzgerald, et al. 2000). However, it may

benefit researchers, teachers, and learners if experimentation is done to find out the optimum layout of the screen. Developers may be able to find out whether or not all the options should receive equal prominence and in what order the options should be placed on the screen. Though current research may be conflicting and scarce, through continued and ongoing contact with researchers and teachers, software developers can find the most information about what options to make available to the students and in what layout.

### **Conflict with Past Research**

The second and third research questions served to focus the attention of the study on use of the repeat and ABC buttons. Past research in the areas of modification of input (Hsu, 1994) and video captioning (Borras, 1993; Borras & Lafayette, 1994; Garza, 1991) all indicate positive relationships between simultaneous input of audio-visual information with textual information and student recall and comprehension. Garza explained that he was initially concerned that the learners would be overloaded by trying to interpret both input modes at once, but that the learners with video captioning demonstrated higher comprehension levels than a control group. He did note that perhaps this overload would occur with lower-level students (246).

Though the present study was conducted with lower-level students, which may be one explanation for the negative relationship between the ABC button and successful performance, much more research is necessary in this area. Past research in this area has been conducted with smaller sample sizes (Borras and Lafayette, 44; Garza, 40; Hsu, 15) and shorter, less diverse periods of study. Furthermore, the present study was undertaken in the most authentic learning environment of all of these studies. It is very important to determine

if the conflict is the result of the longer treatment period, the larger sample size, the larger diversity of activities available to the students, the more authentic learning environment, the fact that the subjects in this study were beginners and false beginners, or other factors.

Although this research cannot be generalized to all contexts, strong relationships between the ABC button, the repeat button, and student performance indicate that more research is necessary.

The fact that the ABC button and the glossary were not introduced may have had an important effect on the outcome of this analysis. Students who only try what the teacher instructs them to try and prefer not to explore where teachers have not given them guidance might have been less likely to use the ABC button or glossary. It is possible that in the environment of this study, the students who used the ABC button were not certain about its value. Indeed, Hsu's (1994) research which worked with low-level language learners concluded that replay with text was most likely to enable the students to acquire new words and that the students judged this option as the most helpful. In the current study, it is impossible to determine how much effect introduction of the ABC button would have had. One could speculate that if the students knew their teachers thought the option was important enough to introduce, they might have considered the function of the ABC button as more important and therefore used it more confidently (perhaps even affecting their performance). Clearly, research similar to this study that controls some of the limitations that were present can have important implications for CALL researchers, teachers, and developers.

### **Limitations of the Study**

The current study had many limitations that may have affected the outcome of the results and which may limit the generalizability of the findings. The first limitation of the present study, and perhaps the greatest was, the researcher was not present and did not have video or audio recordings of student interaction, and therefore could not learn when, how, and why students chose the options they chose.

Secondly, there was no exit exam. Therefore, the only measure of student performance was their percentage of questions answered correctly within the program. There were no exit listening exams or oral interviews to determine the students' proficiency before and after the eight-week interaction with the program. This represents a serious problem because the measure of student performance used within this study was not uniform for each student. The shuffler level (p. 25) adjusted the questions given to each student so that students who did well were asked more (and harder) questions. To attain a more accurate measure of student performance, it is important to apply an evaluation method that tests all students with the same scale.

Third, because the study did take place in a real L2 environment, students' time spent with the program varied considerably, and many students were most likely not completely acquainted with the program. Also, students were not equally oriented to the use of all of the program's features. Had the students understood the glossary function, for example, they may have used it more often.

Students were required to use the microphone button and the benefit that resulted from its use is hard to determine in this study because all the students used it frequently.

Furthermore, not all options were recorded. The fast-forward and rewind options may have



affected student performance significantly, and the pause/play button may warrant research. Finally, this extremely authentic learning environment inhibited complete knowledge of all the factors that influenced student performance. It is possible that the statistics gathered are misleading and that future research will contradict these findings, especially in light of past research in the area of simultaneous audio-visual and textual input.

### **Suggestions for Future Research**

The findings in this study warrant further research in many areas. It is very important for language teachers and researchers to understand the exact benefit of video captioning and the effect of the option to replay aural input with accompanying text. It would be beneficial to research this issue by repeating this type of analysis using DynEd's *New Dynamic English* in the following different environments:

1. With students of other L1 backgrounds at the beginning level
2. With students in more advanced levels with the same L1 background as this study (Arabic)
3. With other levels of students with various L1 backgrounds

Such research would be beneficial not only for teachers using this program, but for teachers considering use of video captioning or other video-text input, and for SLA researchers interested in how learners whose L1 is Arabic learn English. This may also have significant implications for beginning learners from other languages or learners with other Semitic languages (such as Hebrew and Amharic) as their L1.

Besides further research pertaining to simultaneous aural and textual input, near-replication of this study would be beneficial for several other reasons. Future research could produce better insight into the benefits of using all of the features of this program. Similar studies could instruct the students equally in the use of all of the features, including both ways to access the glossary. Researchers could also allow students to use the microphone and headphone options without requiring their use in order to determine if students who use them improve in aural proficiency.

Researchers could gain deeper insight by using data collected from the same program over a longer period of time (perhaps a 15-week semester, rather than 8 weeks) with a larger number of subjects. Also, conducting a study in which the students were sure to use the software in class for at least one hour each week would help research.

Researcher contact with the teachers and students throughout the course of the study in order to find out more detailed information about student interaction with the program through interviews, audio or video recordings, and think-aloud protocols may provide more knowledge about SLA and CALL. This would also facilitate the researcher providing detailed instructions for teachers so students working with the program would all have uniform instruction and researchers might attain more accurate results. Researchers (or teachers) could also pretest and posttest students in various discrete areas, such as vocabulary knowledge, to try to determine the relationship between student choices within the program and student acquisition of certain structures or skills.

Finally, although large-scale reproduction of the current study, with minor adjustments, seems to hold very promising potential, smaller-scale studies of the same program may have other benefits. Because this program allows the students many options and provides access

to a real CALL learning environment, important insights could be gained from continuing the current research with a smaller number of students for the following reasons:

1. All the students could be interviewed and thoroughly examined at the beginning and at the end of the study.
2. Researchers and teachers could videotape each individual student's interaction with the program to notice interactional behavior.
3. Teachers and researchers could find out about student attitudes and opinions about the program and gauge the effect they had on learning.

### **Benefits and Drawbacks**

This study in an authentic learning environment has succeeded in producing results that can provide a starting point for much more research. The results were verified by each of the statistical analyses applied to the data. The learners in this study were completely unaware of the research that would take place, and all of the data collection occurred within a real L2 classroom that would have been conducted in the same manner regardless of the research. One of the most obvious benefits of this type of authentic learning research is the abundance of external validity (this research provides suggestions for direct application in the classrooms from which it was taken) and the lack of the researcher's influence on the performance and learning of the students. SLA and CALL researchers may be able to learn more about how learners learn by studying what learners actually do (Hegelheimer and Chapelle, 2000; Chapelle, 1998).

On the other hand, this type of research is limited in its ability to focus on discrete variables and in its level of internal reliability (Hulstijn, 1997). Pre- and Post- test data

about specific areas of student improvement over the course of a semester may be able to provide information about how a program may have helped students, but this information may be hard to distinguish because so many other factors may be involved. Indeed, an EFL setting, rather than an ESL setting is ideal for this type of research, because of the somewhat diminished likelihood of the students being influenced from other means of instruction or interaction with the target language. Still, only in a more controlled environment is it possible to be certain that specific means of instruction are the cause of student gains in specific skill areas.

### **Conclusion**

SLA and CALL researchers working together with teachers and software developers can help each other to achieve the goals of their fields. Researchers interested in finding out about how learners acquire language may gain a great deal by observing actual learning environments. Teachers who use CALL software can work together with researchers to provide these environments in which researchers can have influence in how the data is collected and how the instruction is conducted, all within the realm of a natural L2 classroom. Developers can provide researchers with the tools necessary to collect data in unobtrusive ways and provide the type of interaction that researchers are interested in.

Teachers can learn from current research what options are best for their students. Teachers may also gain insights into ways of assessing student interaction and performance as well as how to conduct classroom research that is specific to the background and level of their students. Researchers can inform teachers in the use of certain programs and their

application in research. Teachers and learners can also gain from the improvements in current technology that may result from empirical research.

Besides the obvious necessity software developers have of contact with teachers and administrators, they can also benefit from ongoing contact with researchers who provide information about options that are the most beneficial and options that may be less beneficial for certain learners. This can guide software developers in the support that they offer teachers and in future development of programs.

Only in the most authentic of research settings are these types of relationships possible. This study, because it includes collaboration with teachers and software developers, has important implications for all three parties. It points to areas that need further research and choices that could be explored both in software development and in L2 teaching.

## APPENDIX A: PLACEMENT TEST DESCRIPTION

[All material (including words, tables, and figures) taken directly from the *Teacher's Guide for DynEd's Placement Tests* (pp. 9-11)]

Table A.1: DynEd Placement Levels (p. 9)

<i>DynEd Placement Level</i>	Appropriate Courses
0.0 Beginner-	NDE 1A
0.5 (TOEIC 250-350)	NDE 1B
1.0 (TOEIC 300-400)	NDE 2A
1.5 (TOEIC 350-450)	NDE 2B
2.0 (TOEIC 400-650)	NDE 3A; 3B
2.5 (TOEIC 600-750)~TOEFL 500	NDE 4A
3.0 (TOEIC 700-850)~TOEFL 550	NDE 4B
3.5 (TOEIC 800-950)	Advanced content studies
4.0 Advanced Non-Native Speaker	
5.0 Educated Native Speaker	

### About the Test

The DynEd Placement Test helps determine where students are in terms of the language progression found in DynEd courses. As such, the test is not a true proficiency test. Nevertheless, the DynEd General Placement Test should correlate reasonably well with most standard tests, and some programs may find the test useful as an indicator of language proficiency, especially at the lower levels.

The Placement Test provides a quick evaluation of language skills, with an emphasis on listening comprehension and grammar. Though some reading is required in the taking of the test itself, there are no test items that focus on extended passages. Rather, the focus is on basic, sentence-level comprehension.

The Placement Test is a computer adaptive, variable length test, which means that it responds to and adjusts to the student's performance. If a student is doing well, the test will ramp up more quickly to items at a higher level. For higher level students, the test will continue until a student has completed 65~75 items for each part of the test, or 140 test items for someone taking both Part 1 & Part 2. For students at a lower language level, the test will move more slowly and will end once a student misses a significant number of items at any stage of the test or if the on-going total score falls below specified cut-off points. For the beginner, the test may stop after only 10 or 15 items, placing the student at the lowest level of the course and reducing unnecessary frustration at having to face a larger number of test items that are too difficult. In such a case, the test may only take two or three minutes.

It is important to note that the question number shown for each test item in a test refers only to its place in the test sequence, and not to the question itself, which will be different for each student. Test items are randomly selected from a database of relevant questions, not all of which are used in every test, to ensure variation for each student and from one test session to another. However, since the number of test items is not large, familiarity with the test will degrade the results, so it is important to use the test only for initial placement purposes and not to use it to test achievement or mastery of course material.

Access to the Placement Test is controlled by the teacher through the Records Manager. Once a student has taken a test, it is automatically locked, which means the student may not take it again unless the teacher specifically unlocks the test.

### **Orienting Students to the Test**

Before giving the Placement Test, teachers should prepare the student or class by providing basic instructions and giving examples of each type of question, in the students' own language if possible. It is also important that the students be able to use the mouse. Otherwise, the Test results will not be valid. Once the Test begins, students should not exit the Test until it is completed and their score is recorded. If a student fails to answer any question, it is counted as an incorrect response.

### **Types of Test Questions**

There are several types of questions in the Placement Test, including multiple choice items that test vocabulary and grammar, listening comprehension, sentence construction, and sentence ordering.

### **Lexical/Grammatical Fill-In**

This multiple choice type of question generally focuses on vocabulary or grammar points, such as which form of a verb to use. Students read a sentence and then indicate which

Hello, my _____ is John.
↘ city
↘ country
↘ language
↘ name

Figure A.2: Lexical/grammatical fill-in (p. 10)



of four choices correctly completes the sentence. These questions are timed, and students have one chance to answer them.

### Listening Comprehension

A second type of question focuses on listening comprehension. For example, the student may hear a question or statement and have to answer or complete it with the correct choice of a word, number, phrase or sentence.

- ✓ An hour.
- ✓ By car.
- ✓ At the office.
- ✓ After work.

<b>24</b>	<b>56</b>	<b>89</b>
<b>32</b>	<b>67</b>	

Figures A.3 & A.4: Listening Comprehension questions (p. 10)

In these types of questions, the student has only one chance to get the correct answer. If a question times out, the question will be repeated once. If the question times out a second time, the program will move on to the next question and students will not have another chance to come back to that question.

### Sentence Construction

In the third type of test item, students are asked to move a set of words or phrases into the correct places within a sentence. In this type of test item, students have one chance. Students are also allowed to time out once before the program moves on to the next item. Please note that each time a student moves a word, the timer begins again, providing the student with enough time to complete that test item, though if the student spends too much time on several such test items, the test itself will eventually reach its time limit and end.

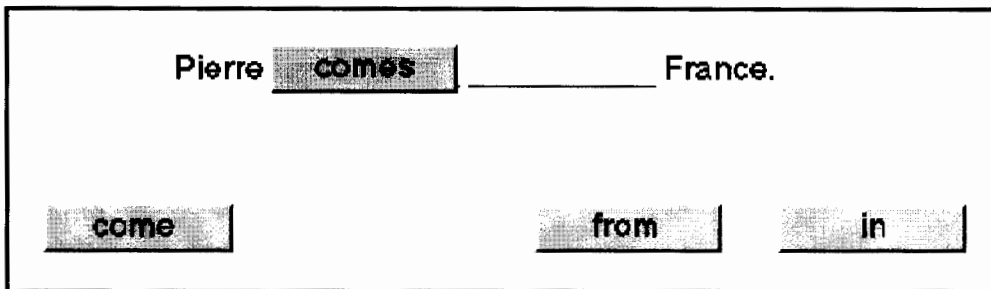


Figure A.5: Sentence construction (p. 11)

### Sentence Ordering

In the latter parts of Part 2 of the Placement Test there are several sentence-ordering items. For these test items, students should read the sentences and then decide the order. In this example, sentence B comes first, followed by C and then A. Students have one chance to get the answer. These items test a student's ability to see how items within one sentence refer to items in other sentences and help to express a sequence.

1 **b** → 2  → 3

**a** Powerful antibiotics have also been developed.

In the field of medical science,  
there have been many important discoveries.

**c** One great success has been the development of vaccines.

**a**

Figure A.6: Sentence ordering (p. 11)

## APPENDIX B: INTERACTION DESCRIPTION

A hypothetical student could begin the program and be given the choices under “Disk 2” of “Level 1” (see Table 3.1, p. 25). The student would then hear “New Dynamic English. Please make a choice.” The student can click on one of the units within that disk, such as “Family Schedule,” (see Table 3.1, p. 25) at which point, the student would be given a choice of all the activities in that unit, and again hear, “Please make a choice.”

The activities in that unit include “The Harris Family,” “Bob and Sandra’s Schedules,” “Collette and John’s schedules,” “Questions,” and “Focus Exercises.” The first three exercises are mostly listening with occasional comprehension checks and the last two quiz the student on the material that was just covered. If the student chose, “The Harris Family,” she would then see the screen displayed in Figure 3.2. There is no writing or speaking involved for the student at this point, only listening and reading. The student is directed by a narrator’s voice (N) to listen or to answer the questions. The narrators are different from question to question however. A student may be directed to answer a question by a man’s voice and congratulated by a woman’s voice or a different man’s voice. Characters in the display may also talk (the characters are minimally animated and their mouths move when they talk). A sample interaction sequence may look like this:

N: This is the Harris Family. Bob and Sandra are married.  
 [continues to describe other family relationships,  
 i.e., their daughter’s name is Collette. Their son’s  
 name is John]

Bob: Sandra is my wife. [border highlights around Sandra]

Sandra: Collette and John are our two children. [borders highlight around the children as names are mentioned]

N: Comprehension check: Who is John's sister?

User: [clicks on image of Collette]

N: Good, you're right. Collette is John's sister. Who is John's father?

User: [clicks on Sandra]

N: No, that's not correct, Sandra is John's mother. Please try again.

User: [clicks on Bob]

N: Yes, that's right. Please make a choice.

User: [clicks on any of the images, in this case, Bob]

N: Look at the pictures. Listen carefully. Use the repeat button to hear each sentence *several* times. Bob works at the post office.

Bob: That's right, I work for the post office.

N: Each morning, Bob gets up at... [describes Bob's morning routine slowly and carefully, with colored illustrations, i.e., a plate with eggs and bacon = breakfast, a garage door with an arrow pointing out = leaves for work, clocks to indicate times].

N: When does he leave for work?

User: [clicks on one of three analog clocks]

N: No, that's not correct. Please try again.

User: [clicks on the correct clock]

N: Good. You're right. He leaves for work at seven forty-five.

N: [explains how Bob goes to work by car, then when he comes home and how long he reads the paper and watches TV].

N: Sandra is a piano teacher. Her schedule changes every day. [describes Sandra's schedule and checks the learner's comprehension occasionally.]

N: Is her schedule the same every day?

User: [clicks "yes" or "no" in a pop-up window]

N: Good choice. That's enough for now. [returns to menu of activities within "Family Schedule"] Please make a choice.

This sequence describes a typical interaction between the learner and the computer.

There are several other types of activities that the learner can participate in. Besides clicking on the correct image or answer to a question, the learner may be asked to drag the appropriate word to the right blank or to read a sentence into the microphone for speech recognition.

There are also occasional multiple choice or fill-in-the-blank exercises. All the questions in the program are posed aurally. Students almost always have the option of choosing to click on the appropriate answer. There are many units and activities within the program which allow the user to answer using speech recognition. However, the user is not often required to produce language.

The program (especially depending on the teachers who may use it) encourages practice producing speech, but this is never a mandatory part of the program. If the user wanted to practice any of the sentences within a unit, he or she could click on the microphone button and repeat the sentence. The student may also listen many times to the sentences within the program, practice saying them, and compare the phonological differences between

her own voice and the voices of the people within the program. *New Dynamic English* therefore provides easy and ample opportunity for students to practice production, but seldom requires it to complete a task. There are units for sentence reading which allow the students to practice reading sentences to see if the computer recognizes. However, these units do not keep score or ask the students questions.

Much of the program does drill the student, but it never requires the student to continue. The pause button allows the student to stop in order to think about or read the text. Also, if a student chooses to practice speaking or comparing her voice to the narrator's or a character's, the program always stops. The exit button is always available to the student throughout each activity. If the student ever wishes to exit the activity, she can press the exit button and return to the previous menu. The same exit button allows the user to exit the program once she has returned to the first menu. If the student completes an exercise, he then hears, "That's enough for now," and, depending on the performance of the student, "You did very well," or "Thank you for your effort."

## APPENDIX C: PRELIMINARY MULTIPLE REGRESSION ANALYSES

The first regression model indicated lack of significant relationships between any of the options and performance, indicating that time on task, the repeat button, the ABC button, and all other options could not be used as predictor variables for student performance. Part of the difficulty with the original model was the lack of normality in the distribution of the errors. The ABC button, for example, had 46 students who did not use it at all. Therefore, the distribution was far from normal.

In order to bring the data close to normality and to deal with some of the outliers, several variables consisting of continuous data were recoded into ranges. The ABC button was recoded into use and non-use, which split the students 45-46, respectively. The repeat button was recoded into four ranges (0, 1-10, 11-99, and >100) because there were 15 students who did not use it. 29 students did not attempt SR, and eight students who attempted SR were never successful, therefore all SR categories (attempts, correct first-trials, and percentage of correct first-trials) were recoded into the three ranges – 0, 1-25, and >26. Finally, the glossary was recoded into the ranges 0, 1, and 2-4.

Student performance, as measured by percentage of questions answered correctly on the first try, was entered as the dependent variable in the revised model. The following independent variables were re-entered for multiple regression analysis: total time in minutes, shuffler level, ABC button, repeat button, microphone button, headphone button, SR attempts, SR correct first trials, % of SR correct, glossary usage, number of phrases heard, total length of phrases heard in seconds, and placement level.



**APPENDIX D: SYSTEM REQUIREMENTS**

For a Macintosh the software requires MacOS 7.0 or later, 8MB of RAM (16MB on Power PC), and a 640 x 480 monitor with thousands of colors. For a Windows system, the software requires Windows 95/98 or Windows NT, a Pentium 90 with 16MB of RAM, and a 640 x 480 monitor with high-color 16-bit display. The program also requires a Quad-speed CD-ROM drive, a 16-bit sound card, microphone, and headphones or speakers.

## BIBLIOGRAPHY

- Bland, S. K., Noblitt, J. S., Armington, S., & Gay, G. (1990). The naive lexical hypothesis: Evidence from computer-assisted language learning. *Modern Language Journal*, 74, 440-450.
- Borras, I. (1993). Developing and assessing "Practicing Spoken French." *Educational Technology Research and Development*, 41 (4), 91-103.
- Borras, I., & Lafayette, R. (1994). Effects of multi-media courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal*, 78(1), 61-75.
- Bowerman, B., and O'Connell, R. (1990). *Linear statistical models: An applied approach*. Boston: PWS-Kent Publishing Company.
- Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, 15, 357-386.
- Chapelle, C. A. (1996). Validity issues in computer-assisted strategy assessment. *Applied Language Learning*, 7(1), 47-60.
- Chapelle, C. A. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning and Technology*, 2(1), 22-34. Retrieved September 12, 2000 from the World Wide Web: <http://llt.msu.edu/vol2num1/article1/index.html>
- Chapelle, C., & Mizuno, S. (1989). Students' strategies with learner-controlled CALL. *CALICO Journal*, 7(2), 25-47.
- Chapelle, C., Jamieson, J., & Park, Y. (1996). Second language classroom research traditions: How does CALL fit? In M. Pennington, (Ed.), *The power of CALL*. Huston, TX: Athelstan.
- Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning*. Cambridge: Cambridge University Press.
- Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, 80, 183-198.
- de Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, 19, 249-276.

- DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, 17, 379-410.
- DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, 19, 195-222.
- Davis, J.N., & Lyman-Hager, M. (1997). Computers and L2 reading: Student performance, student attitudes. *Foreign Language Annals*, 30(1), 58-72.
- Doughty, C. (1987). Relating second-language acquisition theory to CALL research and application. In W. F. Smith (Ed.), *Modern media in foreign language education: Theory and implementation* (pp. 133-167).
- Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition*, 13, 431-469.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.
- Fitzgerald, S., Bower, H., & Morrall, A. (2000). New Dynamic English. Software Review, *CALICO Journal*. 17(2), 308-319.
- Garza, T. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals*, 24 (3), 239-258
- Harless, W., Zier, A., & Duncan, R. (1999). Virtual dialogs with native speakers: The evaluation of an interactive multimedia method. *CALICO Journal* 16(3), 313-336.
- Hatch, E., & Farhady, H. (1982). *Research Design and Statistics for Applied Linguistics*. Rowley, MA: Newbury House.
- Hegelheimer, V. (1998). Effects of textual glosses and sentence-level audio glosses on online reading comprehension and vocabulary recall. Unpublished doctoral dissertation. Department of Educational Psychology, College of Education, University of Illinois, Urbana, IL.
- Hegelheimer, V., & Chappelle, C. A. (2000). Methodological issues in research on learner-computer interactions in CALL. *Language Learning and Technology*, 4 (1), 41-59.
- Hulstijn, J. (1993). When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *The Modern Language Journal*, 77(2), 139-147.

- Hulstijn, J. (1997). Second language acquisition research in the laboratory: Possibilities and limitations. *Studies in Second Language Acquisition*, 19, 131-143.
- Hsu, J. (1994). *Computer assisted language learning (CALL): The effect of ESL students' use of interactional modifications on listening comprehension*. Unpublished doctoral dissertation, Department of Curriculum and Instruction, College of Education, Iowa State University, Ames, IA.
- Hsu, J., Chapelle, C., & Thompson, A. (1993). Exploratory learning environments: What are they and do students explore? *Journal of Educational Computing Research*, 9, 1-15.
- Issidoridies, D. C. (1988). The discovery of a miniature linguistic system: Function words and comprehension of an unfamiliar language. *Journal of Psycholinguistic Research*, 17, 317-339.
- Kolich, E., (1991). Effects of computer-assisted vocabulary training on word knowledge. *Journal of Educational Research*, 84(3), 177-182.
- Koren, S. (1999). Vocabulary instruction through hypertext: Are there advantages over traditional methods of teaching? *Language Learning and Technology*, 4(1), 1-18.
- Levy, M. (1997). *Computer –assisted language learning: Context and Conceptualization*. New York, New York: Clarendon Press.
- Lomicka, L. (1997). To gloss or not to gloss: An investigation of reading comprehension on-line. *Language Learning and Technology*, 1(2), 41-50.
- Long, M. H. (1996). The role of linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia, (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego: Academic Press.
- Meich, E.J., Nave, B., & Mosteller, F. (1996). On CALL: A review of computer-assisted language learning in U.S. colleges and universities. *ERIC 394525*
- Mitchell, Erika <erika\_mitchell@zu.ac.ae> (2000a, August 21). Re: teaching [personal email].
- Mitchell, Erika <erika\_mitchell@zu.ac.ae> (2000b, August 26). Re: information [personal email].
- Mitchell, Erika <erika\_mitchell@zu.ac.ae> (2000c, September 20). Re: class environment [personal email].
- Mitchell, Erika <erika\_mitchell@zu.ac.ae> (2000d, November 21). Re: questions [personal email].

- Nagata, N. (1993). Intelligent computer feedback for second language instruction. *The Modern Language Journal*, 77(3), 330-339.
- New Dynamic English (Version 3.1) [Computer software]. (1999). Burlingame, CA: DynEd International, Inc.
- Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning* 44(3), 493-527.
- Pica, T. (1997). Second language teaching and research relationships: A North American view. *Language Teaching Research*, 1(1), 48-72.
- Records Manager Guide [Software manual]. (1999). Burlingame, CA: DynEd International, Inc.
- Robinson, G. L. (1991). Effective feedback strategies in CALL: learning theory and empirical research. In P. Dunkel (Ed.), *Computer-assisted language learning and testing: Research issues and practice* (pp. 155-167). Rowley, MA: Newbury House Publishers.
- Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced and instructed conditions. *Studies in Second Language Acquisition*. 19, 223-248.
- Teacher's Guide for DynEd's Placement Tests [Software manual]. (1999). Burlingame, CA: DynEd International, Inc.
- User's Guide [Software manual]. (1999). Burlingame, CA: DynEd International, Inc.
- Warschauer, M. (1996). Computer-assisted language learning: An introduction. In S. Fotos (Ed.), (pp. 3-20). Tokyo: Logos International. Retrieved from October 21, 2000 from the World Wide Web: <http://www.logosintl.com/LogosCompassSeries.html>
- Yang, L. & Givón, T. (1997). Benefits and drawbacks of controlled laboratory studies of second language acquisition. *Studies in Second Language Acquisition*. 19, 173-193.

## ACKNOWLEDGEMENTS

First I thank my God and Savior Jesus Christ, who has given me all the ability, people, time, energy, and resources to complete this project. I give credit to Him for providing through other people all the help necessary to write this thesis in a timely fashion.

Thank you to my committee members Volker Hegelheimer, Carol Chapelle, and Mack Shelley, who have taught me much about CALL, SLA, statistics, and thesis writing, and made this paper worthwhile for me and hopefully for many other readers. Thank you also to Carol for the conception of this research and contact with the data.

Thank you to Erika Mitchell and the teachers and students at Zayed University in Abu Dhabi, United Arab Emirates for the data for this research and information about the learning environment.

I would also like to thank Norm Harris and everyone at DynEd International for providing the software and information about its use and for providing contact with Zayed University and the data.

Finally, more than any other person, I thank my wife Nichole, who God has used to encourage, support, and care for me as I've written and edited this thesis.