Thesis for the Degree of Master

Assessing the Effectiveness of
an Online CALL Pronunciation Tool

by

Dean Comeau

Department of TESOL
The Graduate School
Sookmyung Women’s University
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLE OF CONTENTS</td>
<td>i</td>
</tr>
<tr>
<td>LIST OF TABLES AND FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF APPENDICIES</td>
<td>vi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vii</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2 LITERATURE REVIEW</td>
<td>4</td>
</tr>
<tr>
<td>2.1 A Brief History SLA Pronunciation Theory</td>
<td>4</td>
</tr>
<tr>
<td>2.2 SLA Pronunciation Theory and Research</td>
<td>7</td>
</tr>
<tr>
<td>2.2.1 The nativeness principle, the CPH, and adults</td>
<td>7</td>
</tr>
<tr>
<td>2.2.2 Intelligibility</td>
<td>10</td>
</tr>
<tr>
<td>2.2.3 Segmentals vs. suprasegmentals</td>
<td>14</td>
</tr>
<tr>
<td>2.2.3.1 Suprasegmentals</td>
<td>15</td>
</tr>
</tbody>
</table>
2.2.3.2 Segmentals and variation...................................................... 17

2.3 Changing Pronunciation Habits.................................................... 19

2.4 CALL Research ........................................................................... 21

2.5 Literature Review Summary....................................................... 25

CHAPTER 3 METHODOLOGY.......................................................... 27

3.1 Research Questions...................................................................... 27

3.2 Participants................................................................................ 27

3.3 Data Collection and Procedures.................................................. 28

  3.3.1 Pre-test.................................................................................. 28

  3.3.2 Treatment.............................................................................. 29

  3.3.2.1 EnglishCentral ................................................................. 29

  3.3.3 Post Test. ............................................................................. 32

  3.3.4 Surveys................................................................................. 33

  3.3.5 Participant usage reports...................................................... 33

3.4 Data Sources ............................................................................. 34

  3.4.1 Pre and post-tests................................................................. 34

  3.4.2 Rater. .................................................................................. 38
6.1 Implications for Practice ................................................................. 56
6.2 Limitations ..................................................................................... 56
6.3 Future Research ............................................................................ 58
6.4 Conclusion ..................................................................................... 58

REFERENCES ...................................................................................... 60

APPENDICIES ...................................................................................... 64
LIST OF TABLES AND FIGURES

TABLES

Table 3.1
Table of Procedures ................................................................. 33

Table 3.2
Inclusion of Pre-tests and Post-tests ............................................. 42

Table 4.1
HUG- Post-Test / Usage Correlation ........................................... 47

Table 4.2
LUG- Post-Test / Usage Correlation ........................................... 47

FIGURES

Figure 3.1
Summary of steps used in the creation of the pre and post-tests .......... 37

Figure 4.1
Pre-test and post-test scores for all 17 participants ...................... 46

Figure 4.2
Mean usage scores for the HUG and the LUG .................................................. 48

Figure 4.3

Pre-test and post-test comparisons between LUG and HUG ...................... 49

Figure 4.4

Participants’ opinions on using EnglishCentral as a pronunciation training tool ................................................................................................................ 51

Figure 5.1

Pre-test and post-test results for the HUG ...................................................... 54

LIST OF APPENDICES

Appendix A Survey .......................................................................................... 64

Appendix B Sentences used for pre-tests ....................................................... 68

Appendix C Sentences used for post-tests ....................................................... 73
ABSTRACT

Assessing the Effectiveness of an Online CALL Pronunciation Tool

Dean Comeau
The Graduate School of TESOL
Sookmyung Women’s University

Changing EFL student pronunciation errors is often a difficult and time consuming process, which can be compounded by the habit of students reverting to L1 substitutions. This study examines the effects of an online CALL pronunciation tool (EnglishCentral) on beginner-high / intermediate low EFL college students. It seeks to determine if EnglishCentral significantly improves the pronunciation of the participants, if there is a correlation between site usage and improvements in pronunciation, and how the participants feel about this CALL tool. Employing a quasi-experimental design, the research consists of a pre-test, a 9 week treatment, a post-test, analysis of site usage, and participant surveys. Analysis of the data collected indicated that the using the website did not result in a significant improvement in the participants’ pronunciation, however there was a medium correlation between site usage and post-test scores. Participants’ attitudes towards the site were generally favorable. Though this research was able add to the research suggesting that users tend to have favorable
views towards CALL tools, it was unable to produce quantitative data indicating that such tools are effective.
CHAPTER 1 INTRODUCTION

One of the most important aspects of being understood when speaking any language is pronunciation. As Brown (2008) points out, the most often used medium for communication is oral speech. Traditionally, second language acquisition (SLA) theories have tended to focus on grammar or the inner workings of the mind in an attempt to understand how we acquire a second-language. However, all the grammatical knowledge in the world, or a perfect representation of the inner workings of the human brain and how it processes language will be useless if the speaker cannot convey his or her message. Just as someone with illegible writing would struggle to communicate their message in written form, so would a speaker whose pronunciation is so poor that others cannot understand them. It is useless having something to say if you cannot say it in such a way so that others can understand it.

Many teachers feel lost when trying to successfully integrate pronunciation teaching in their classrooms since they are left to rely on their intuitions due to a lack authoritative research on the subject (Derwing & Munro, 2005). Able to get students to produce target sounds in class during focused pronunciation practice, they are disheartened when students revert to ‘old habits’ during general conversation, often substituting sounds from their L1 in place of unfamiliar sounds in the L2. The problem is especially frustrating knowing that it is not a lack of ability that is causing the
pronunciation problem (Lado, 1956). Though some students and teachers believe in the critical period hypothesis (CPH) and thus believe that there is not much point in trying to improve pronunciation, research indicates that adult students believe explicit instruction helps them improve their pronunciation (Nakashima, 2006; Rajadurai, 2001). Though the goal of ‘accentless speech’ is actually a myth, as all speech is accented (Scovel, 1988, as cited in Neri, Cucchiarini, Strik, & Boves, 2002), and although students may never achieve the elusive goal associated with the nativeness principle, they can strive to improve the intelligibility of their pronunciation.

With the advent of the communicative approach to teaching, pronunciation instruction has moved from focusing on attempting to achieve native-like pronunciation to using various models in an attempt to help students achieve intelligible pronunciation. What is needed is some form of sustained, motivating pronunciation practice, outside the classroom, that allows students to change pronunciation patterns over a period of time. However, since students are often unable to self assess pronunciation errors (Dlaska & Krekel, 2008), unguided practice outside the classroom without some sort of feedback is perhaps not the best use of their time. It is in this area that Automatic Speech Recognition (ASR) may be beneficial.

There has been a fair bit of interest in ASR, Computer Assisted
Language Learning (CALL), and Computer Assisted Pronunciation Training (CAPT) (Chun, 2007). Unfortunately, many commercially available CAPT programs are quite expensive, often in excess of $100. However, a free web-based service, www.englishcentral.com, became available in 2008, and is emerging as a potentially useful CAPT system. This website uses mostly authentic videos from YouTube which have been subtitled, to allow users to record their own voices and then have their pronunciation assessed through ASR. Despite indications that the website seems promising, there is no evidence showing it aids pronunciation development.

Though there are indications that ASR-CALL is at least somewhat effective (Engwall & Balter, 2007; Neri, Cucchiarini, & Strik, 2008; Seferoglu, 2005; Sung, 2008), the effectiveness of this tool has yet to be determined. This study seeks to determine if EnglishCentral is a useful tool for practicing pronunciation, if it helps make students' pronunciation more intelligible, and to determine how the students perceive the website. Related theories and research from both the SLA and the CALL perspectives will be examined, followed by an explanation of this research study. The results will be presented and analyzed, followed by a discussion on what the results mean for both students and teachers.
CHAPTER 2 LITERATURE REVIEW

2.1 A Brief History SLA Pronunciation Theory

Views on the teaching of L2 pronunciation have changed numerous times over the past 150 years. Some theories have employed strict guidelines on how pronunciation should be dealt with in the L2 classroom, while others have either ignored pronunciation completely, or given no guidelines on how it should be taught. What follows is a brief history of how pronunciation has been dealt with in different theories, concluding with current views on how pronunciation should be handled in the L2 classroom.

Celce-Murcia, Brinton, and Goodwin (2003) provide an excellent overview of the history of pronunciation theory in SLA in their first chapter. The Direct Method of the 1800s and the early 1900s taught pronunciation through intuition and imitation. Basically, learning an L2 was viewed the same as learning an L1, and thus it was believed learners would simply acquire pronunciation through exposure to good models. Similar claims can be made about naturalistic approaches, where proponents believe that exposure to good models is sufficient for learners to ‘pick up’ pronunciation without having received any explicit pronunciation instruction. In the 1890s with the Reform Movement came the establishment of the International Phonetic Alphabet (IPA). With the belief within the Reform Movement that
the spoken form of a language was primary and therefore of the utmost importance, pronunciation was taught explicitly from the beginning, with the aid of the IPA. The 1940s and 1950s saw the establishment of the Audiolingual Approach and the Oral Approach. In both of these approaches, pronunciation was emphasized and taught explicitly from the beginning, again with the assistance of phonetic information to help demonstrate the articulation of sounds. Importantly, with Audiolingualism came the introduction of a new form of pronunciation strategy; the minimal pair drill, in which words differing by only a single sound are compared for both listening and guided oral production. Up until this point, though the importance of pronunciation was realized, most of the pronunciation instruction consisted of listen and repeat drills.

The 1960s saw the birth of the Cognitive Approach, led by Chomsky. With the belief that all language was rule governed, coupled with the belief that pronunciation was dominated by habit formation and was thus incompatible with the Cognitive Approach since it could not be learned, explicit pronunciation teaching all but disappeared. The Silent Way of the 1970s once again saw pronunciation taught from the beginning, with the aid of a phonetic alphabet. A focus on how words combined to form phrases, along with how stress and intonation shaped the production of an utterance saw the first shift towards teaching the suprasegmental features of
pronunciation. At about the same time, Community Language Learning gained in popularity, in which students recorded their utterances for comparison with the teacher’s model. Thus, after the Cognitive Approach, we can see a return to focusing on pronunciation, but with new innovations such as minimal pair drills, focus on suprasegmentals, and students analyzing self-recordings.

Finally, in the 1980s, the Communicative Approach gained popularity, and while it again acknowledged the importance of pronunciation, there was yet another shift in focus. Rather than having students strive for native-like pronunciation, the pronunciation goal of the Communicative Approach is intelligible pronunciation. Additionally, the focus is not just on the segmental side of pronunciation, but also on the suprasegmental side. However, it is important to note that some believe that the suprasegmental aspects of pronunciation are not easily learnable (Peperkamp and Dupoux, 1992, as cited in Field, 2005; Jenkins, 1998).

As can be seen, over the past 150 years or so, pronunciation has been a part of SLA theory in some form or another, with the exception of the Cognitive Approach of the 1960s. Thus it would appear that both teachers and researchers have been aware of the importance and necessity for teaching pronunciation. However, it seems equally clear that teachers and researchers are uncertain about how to best incorporate pronunciation into
the curriculum, or how to best teach pronunciation to their students. With the rise in popularity of the Communicative Approach, we can see not only a shift with regards to teaching both segmental and suprasegmental aspects of pronunciation, but also a shift in trying to balance fluency and accuracy, with intelligibility rather than native-like perfection being the main goal.

2.2 SLA Pronunciation Theory and Research

2.2.1 The nativeness principle, the CPH, and adults

There have been a number of theories related to the goals of pronunciation teaching in the L2 classroom, as well as factors affecting the attainment of those goals. The main tenet of the nativeness principle is that it is possible for L2 learners to achieve pronunciation on par with native speakers (NSs), and as such, native-like pronunciation should be the goal of pronunciation teaching. In contrast, the CPH puts forth the claim that if the learning of a second language commences after a 'critical period', native-like pronunciation is not possible, regardless of the methods used to improve pronunciation. This section will explore these theories and how they relate to adult L2 learners.

Until about 60 years ago, the dominant principle in SLA pronunciation theory was the nativeness principle. Under this view, not only was it thought
possible for an L2 learner to achieve native-like pronunciation, it was also
desirable (Levis, 2005). Kawai and Hirose (1998) claim that many foreign
language learners still rank achieving native-like pronunciation as their
primary pronunciation goal. However, Lenneberg (1967) and Scovel (1995)
(as cited in Levis, 2005) claim that this view fell out of favor with research
showing that native-like pronunciation may be biologically conditioned to
occur before adulthood. In the debate over the concept of a ‘critical age,’
Long (1990) (as cited in Ellis, 1994, p. 35) claims that acquisition of a native
like accent is not possible after age 6, while Scovel (1998) (as cited in Ellis,
1994, p. 36) claims it is 12. Ellis concludes that regardless of whether a
critical age exists or not, the fact remains that very few adult learners achieve
offers evidence to support this claim that a critical age does exist, but that it
only applies to accent, and not other aspects of language acquisition.

Moyer (1999) unsuccessfully challenged the CPH, but did find that
some variables, like motivation and suprasegmental training, resulted in
more native-like speech. Nakashima (2006) argues that though native-like
pronunciation may be unattainable for adult learners, explicit pronunciation
instruction helps increase students’ relative intelligibility. Rajadurai (2001)
surveyed 74 Malaysian university students studying English, and found that
many of the students believed that adults can improve their pronunciation,
and that pronunciation should be taught explicitly, with the focus being on segmentals rather than suprasegmentals. Thus there seems to be some evidence that explicit pronunciation teaching is both warranted and desired by adults, with the caveat that such instruction may improve pronunciation but is still unlikely to result in native-like pronunciation.

Despite research and evidence showing that native-like pronunciation for L2 learners is rarely attainable, Levis (2005) claims that the nativeness principle still influences pronunciation teaching practices today. Furthermore, the published materials that serve as models for native-like pronunciation are usually General American (GA) or Received Pronunciation (RP). Ironically, only a minority of NSs actually speak using these models (Levis, 2005). It seems paradoxical to hold non-native speakers (NNSs) to higher standards than NSs, though this is what often happens. Davies (1989, as cited in Jenkins, 2006) claims that when NSs use nonstandard forms, they are guilty of performance errors, but when NNSs do the same thing, their nonstandard form is interpreted as a mistake caused by their partial knowledge of the language. Furthermore, Levis (2002) points out that NSs do not make some of the distinctions in prosodic patterns that are often present in pronunciation textbooks. Again, it seems paradoxical to hold NNSs to higher standards than NSs, particularly when those standards are not actually standard.
Though researchers and teachers acknowledge the flaws of the nativeness principle, it appears that both students and teachers are still seeking to attain native-like pronunciation. However, the mere fact that L2 speakers are NNSs means they will always be held to higher standards than NSs, especially when it comes to using non-standard forms of pronunciation. Teachers, students, and researchers need to understand that though it may well be possible to improve the pronunciation of adult students, a critical period very likely does exist. That being the case, seeking to attain a native accent is most likely impossible, and that a more rational goal involves attempting to achieve intelligible pronunciation.

2.2.2 Intelligibility

With the nativeness principle falling out of favor, there has been a shift towards the intelligibility principle, which claims that learners’ utterances simply need to be understood. This is more in line with the concept of communicative competence, in which the ability to communicate is more important than grammatical and other forms of accuracy. The focus is not on imitation or attainment of illusory ideal models, but on allowing accepted variation, in as much as it is still intelligible to the listener. This is much more in line with beliefs that adults can change their pronunciation, while not holding them to an unattainable standard. Though the term 'intelligible'
appears to be easily understood, attempting to define and set guidelines for
'intelligible pronunciation' proves a much more difficult task.

As outlined in Silveira (2002), Bowen (1972) proposed three goals for
pronunciation teaching: the ability to easily and efficiently communicate
orally, the ability to produce the basic contrasts of the target language sound
system, and the ability to understand fluent speech from NSs. This approach
was a shift away from the nativeness principle, but still clung to the idea of
using native-like models. Acton (1984) proposed a model designed to help
fluent NNSs of English improve their pronunciation. However, his approach
put too much of the onus on students completing work outside of the
classroom, and as Silveira (2002) points out, due to real world constraints,
most students and language programs do not have the ability to maintain
such a rigorous regime.

Despite all this talk of ‘intelligibility’, there still is not a clear
definition of what intelligibility actually is. Abercrombie (1949, as cited in
Munro & Derwing, 1999) argued that the average language learner simply
needed what he referred to as ‘comfortably intelligible pronunciation.’ More
recently, Pickering (2006) has stated that there is still no universal definition
nor agreed upon means of measuring intelligibility in relation to
pronunciation. She summarizes Smith and Nelson's (1985, as cited in
Pickering 2006) definition, which is divided into 3 parts from the viewpoint
of the listener: intelligibility (individual words and utterances), comprehensibility (meaning), and interpretability (the speaker's intended meaning). However, Field points out that Smith and Nelson (1985, as cited in Field, 2005) are aware that 'intelligibility' and 'comprehensibility' are often used interchangeably. As such, they suggest limiting 'intelligibility' to words and utterances, while assigning meaning to the term 'comprehensibility.' Thus, it would seem prudent to limit the concept of intelligible pronunciation to the ability of the listener to understand the individual words or utterances at a phonetic level, not to their ability to understand or comprehend the intended message. Besides, other factors such as lexical discourse markers and syntactic relationships can also influence the comprehensibility of an utterance (Derwing & Munro, 1997). Thus, an utterance may be intelligible, yet incomprehensible, or it may be incomprehensible simply due to problems with intelligibility.

Having defined intelligibility of pronunciation as the ability for a listener to understand individual words as spoken by a speaker, another problem arises; namely to whom should speakers be intelligible. Minematsu’s (2004) view is that defining intelligible pronunciation is difficult because it is dependent upon the listener. As such, people from different backgrounds and languages will define intelligible pronunciation differently. This belief is supported by Taylor (1991), who claims that
though many researchers and educators agree that we should strive to help our learners achieve intelligible pronunciation, it is not clear to whom they should be intelligible. With more interactions occurring between NNSs, defining guidelines for intelligible pronunciation becomes much more difficult. He goes on to argue that NS models may not always be the best models, depending on the context, a claim echoed by Deterding and Kirpatrick (2006). Taylor (1991) (citing Gass and Varonis, 1984 and Smith and Bisazza, 1982) found that familiarity with different accents and varieties actually aids comprehension. Combined with Jenkins’ (2006) claim that English is a world language, often used for communication between NNSs from differing L1s, it would appear that whom speakers should be intelligible to is just as complicated as what intelligible pronunciation is.

We can see that the idea of intelligible pronunciation is perhaps a more realistic goal than native-like pronunciation. However, the complicating factor is that ‘intelligible’ is not an easily defined term, as what is intelligible for one person is not necessarily so for another. As is often the case with language, we are reminded that context plays a very important role. However, by providing a number of different pronunciation models to learners, perhaps we can give them an idea of acceptable variations in pronunciation and help them understand that there is no one perfect model that they should strive to attain.
2.2.3 Segmentals vs. suprasegmentals

Any discussion of pronunciation needs to examine the two main aspects of pronunciation, namely segmentals and suprasegmentals. Segmentals are the phonemic sounds of language (the vowels and consonants), while suprasegmentals are features of speech such as stress, rhythm, and intonation (Morley, 1991). Both can affect pronunciation and perceptions of intelligibility. However, just as many SLA theories debated how to teach pronunciation, many theories on pronunciation teaching have debated the importance of segmentals and suprasegmentals, especially with regard to which feature greatest affects pronunciation.

Initially, the rise in popularity of the intelligibility principle brought about a shift in focus from the segmental aspect of pronunciation to the suprasegmental aspect of pronunciation. The "long-standing belief that instruction should focus on suprasegmentals (e.g., Avery & Ehrlich, 1992) assumes that focus on these features leads to better and quicker speaker intelligibility than a focus on segmentals" (Levis, 2005, p. 371). Surveying teacher reference books, articles in journals, and student texts related to pronunciation from the past several years, Morley (1991) noticed a shift emphasizing the importance of the suprasegmental aspects of pronunciation, specifically stress, rhythm, and intonation. However, recent research has shown that a more balanced approach including both segmental and
suprasegmentals is more desirable. Jenkins (1998) agrees that for NSs, suprasegmentals may affect intelligibility to a greater degree, but feels that in the context of English as an International Language (EIL), a balance between teaching segmentals and suprasegmentals is more in line with the learner's goals.

Though it would appear that a balanced approach is necessary, it is important to keep in mind that intelligibility can be relative, dependent upon the listener, who more often than not is a NNS, and as such, perhaps the emphasis placed on suprasegmentals is unwarranted. Varonis and Gass (1982, as cited in Derwing & Munro, 1997, p. 3), determined that both grammar and pronunciation errors affected intelligibility, while Jenkins (2002) found that segmental errors caused more problems in NNS – NNS interactions.

2.2.3.1 Suprasegmentals
Though some argue for a balanced approach, including the teaching of both segmental and suprasegmentals, some, like Field (2005), still feel suprasegmentals are more important. However, when teaching the suprasegmental features of pronunciation, Jenkins (1998) states that many aspects of phonology are not necessary in most EIL contexts, nor are they easily learnable. Peperkamp and Dupoux (1992, as cited in Field, 2005) suggest that some language learners, such as those whose L1 is a fixed-stress
language, may not even include lexical stress as part of their phonological representation of words. Regardless, Field (2005) reminds us that while identifying the key features of what constitutes natural speech and intelligibility, be they segmental or suprasegmental, is still one of the biggest challenges for pronunciation specialists, it is necessary if useful materials are to be developed. Thus features such as word stress, where the rules are highly complex and variable even within generally accepted models such as RP, should perhaps not be over emphasized. However, remembering that most interactions are of the NNS-NNS variety, Deterding and Kirkpatrick (2006) found that while unexpected emphasis and failure to de-accent weak syllables may seem odd to NSs, it may not pose too many problems with intelligibility, especially between NNSs. Furthermore, they determined that the stressing of pronouns, which commonly occurs with L2 speakers of English when their L1 only includes pronouns when they are necessary (i.e. important), also did not seem to affect communication. Citing a number of studies, Field (2005) claims NS listeners rely more on stressed syllables than on unstressed ones and when asked to shadow (repeat the utterance they have just heard) are 3 times more likely to notice and reproduce misplaced stress than mispronounced phonemes.

Thus again we are reminded that context is of great importance, and that who students will be interacting with in English should determine what
aspects of pronunciation are taught. Jenkins (1998) feels that features of connected speech, such as linking and weak forms, are not necessary for NNSs, as she feels it is possible to highlight some syllables without the need to reduce others. Jenkins (2004) feels time may be better spent using "chunking" to teach lexical phrases and "intonational idioms." Furthermore, Flores (1997) claims that rhythm is one of the most difficult pronunciation features to teach because of the difficulty involved in both perceiving and concentrating on the rhythmic pattern as a chunk.

2.2.3.2 Segmentals and variation

With respect to segmentals, Jenkins (1998) feels that the crucial difference between NNSs and NSs is that NNSs produce deviant utterances with relation to the core sounds of English. These core sounds apply mainly to consonants, long and short vowel sounds, and consonant clusters. Furthermore, these deviant utterances often occur because of negative L1 transfer. As a result, she suggests that pronunciation instruction should focus on certain segmentals, nuclear stress, and the effective use of articulatory settings such that it supports the first two areas. Thus the focus should be on helping students produce acceptable variations of core sounds and focusing on nuclear stress to assist learners with the reception of the spoken language. However, some sort of model is necessary, and Deterding and Kirkpatrick
(2006) feel that inner-circle pronunciation, or NS models, are not always the easiest nor most appropriate models for NNS learners, citing the example of British style triphongs. In their study, they attempted to identify shared features of pronunciation and their affect on intelligibility in relation to NNSs of English. They found that widespread pronunciation features appear to not disrupt communication and that some amount of variation is accepted. However, uncommon, non-standard features, which are not shared by speakers from a variety of countries, may occasionally cause problems. They also found that the use of reduced vowels (schwa) did not interfere with intelligibility, but rather enhanced it for NNSs.

Jenkins (2002) found that segmental pronunciation errors caused more problems in NNS-NNS interactions for the simple fact that learners below bilingual proficiency are simply not well enough equipped to deal with and process contextual cues that NSs would be able to use to interpret their interlocutor’s errors in pronunciation. Upon analyzing empirical data collected through personal observations, Jenkins (2002) attempted to identify the segmental factors that impact intelligibility the most as she attempted to develop a Lingua Franca Core (LFC). For example, she arrived at a similar conclusion as Deterding and Kirkpatrick (2006), that substitutions of /θ/ and /ð/ are acceptable, while omissions of sounds in word-initial clusters are not. Though her study is a starting point, it may be difficult to develop a LFC
which accounts for all L2 English learners and their variations in pronunciation.

Thus we can see that it appears a more balanced approach of teaching both segmentals and suprasegmentals is currently in favor, rather than an approach that favors one aspect over the other. Though it would appear that suprasegmental errors cause greater problems for NSs, segmental errors cause more errors in exchanges involving only NNSs. Therefore, it is important to consider our students' context and with whom they will be interacting when providing pronunciation instruction, allowing for accepted variations with regard to segmentals, while continuing to offer instruction on the suprasegmental aspects of language.

2.3 Changing Pronunciation Habits

It has been well observed that changing a student’s pronunciation is a difficult task. Though able to accurately produce target utterances during guided pronunciation exercises, students often revert to old pronunciation habits and reliance on their L1 phonology when not focusing on pronunciation. Lado (1956) believes that learners revert to their L1 habits as they are channeling their cognitive power to the task of constructing meaning. Thus, it seems obvious that changing these ‘habits’ will take time. Orion
(1987) echoes this sentiment in the introduction of her book, warning students that attaining good pronunciation and improving pronunciation takes time.

This being the case, the responsibility for change clearly lies with the student. Acton (1984) agrees with this concept, indicating that practice and change must be effected outside the classroom, and as such, the onus is on the student. However, students must have guided practice, with feedback, as Dlaska and Krekeler (2008) reported that students often have difficulty self-assessing inaccurate pronunciation. In their study of 46 adult advanced learners of German, in which the subjects were asked to self-assess their recorded utterances of specific words by comparing them to NS recordings, they found that learners were only able to recognize less than half of their inaccurate utterances. Thus, it would appear that some form of structured, guided, effective feedback is necessary during pronunciation training.

It is quite apparent that there has been a shift away from trying to achieve native-like pronunciation, to striving for intelligible pronunciation. Even when limited to the understanding of an utterance at the phonemic / word level, intelligibility is a difficult concept to define, and is dependent upon a number of factors, of which the most important is arguably the listener. It would also appear that a balanced approach highlighting both the
segmental and suprasegmental features would most benefit students. Finally, the impetus must be placed on the student to maintain sustained practice outside the classroom, over a period of time, in order to effect change.

2.4 CALL Research

A number of studies have indicated that CAPT is both motivating and an effective means of providing feedback. Though there is limited research indicating that CAPT effectively improves pronunciation, there is research supporting the view that CAPT is viewed positively by students. What follows is an examination of some recent research regarding CAPT and the findings.

Neri, Cucchiarini, and Strik (2008) compared the effectiveness of feedback from an ASR-based CAPT system with a version of a commercial CAPT system that was restricted so that it could not provide feedback. A third group did not use either CAPT system. The group using the unrestricted version of the software were informed of pronunciation errors with the letters corresponding to the mispronounced phoneme colored red and underlined in the onscreen transcription, accompanied by a disappointed red smiley, and a prompt for the subjects to attempt the utterance again. Pre and post-tests consisted of the subjects reading 2 phonetically rich sentences,
which were then evaluated at the segmental level by 6 Dutch L2 expert raters. After pre and post-test scores were analyzed, it was determined that the different types of stimuli did not affect segmental quality in a significant way, which may have been a result of the relatively short timeframe (4 weeks). However, qualitative data collected through anonymous surveys indicated that the subjects enjoyed using the CAPT system, commenting that it was an interesting addition to their regular courses, while also believing in its usefulness, indicating that it helped make them aware of specific pronunciation problems as well as Dutch spelling pronunciation issues.

Sung (2008) set out to determine the effectiveness of another CAPT system, Dr. Speaking. This CAPT package is capable of showing learners pictures of the vocal organs to assist with learning how to pronounce sounds, comparing recorded utterances with those of NSs, and comparing spectrograms of subject utterances with those of NSs. Eleven native speaking Korean university graduates used the Dr. Speaking program daily for 30-60 minutes over a two week period. The pre and post-tests involved subjects recording 90 sentences, each containing target words, which were words with consonants in the coda position, consonant clusters in the onset position, and consonant clusters in the coda position. Analysis of the data revealed that the participants made significantly fewer errors related to vowel insertion, consonant deletion, and sound change. Qualitative data gathered
from a questionnaire indicated that the subjects had positive reactions to the CAPT system, considering it both interesting and useful.

Engwall and Balter (2007) studied 37 adult subjects who used the CAPT system, ARticulation TUtoR (Artur) to practice the unique Swedish phoneme [ ] in a 6-word sentence, in which each word started with the target phoneme. Subjects listened to the sample sentence, and then attempted to repeat it. Artur provided graphical feedback through articulatory animations, the ability to hear individual words repeated at a slower rate, the ability to hear previous attempts, and a graphical representation to show the difference between the recorded utterance and the model one. Subjects perceived the system as effective, and indicated that the practice was both fun and engaging.

Seferoglu (2005) employed a quasi-experimental, pre-test / post-test design to determine the effectiveness of the CAPT system Pronunciation Power. The program allowed subjects to hear sounds produced by NSs, view the waveform of those sounds, record their own voices and compare the waveforms, as well as offered them views of how sounds were articulated through detailed front and side views of the human mouth. The study involved 40 university students studying to become teachers in Turkey. The experimental group used Pronunciation Power for 3 weeks, while the control group used ‘traditional’ (audio, pen, paper) instruction for 3 weeks. During
an interactive presentation designed to stimulate discussion on assigned topics, subjects were rated using a 5-point scale on individual sounds, diphthongs and clusters, linking sounds, stress patterns, sentence stress and rhythm, and intonation and pitch. Analysis of the results showed that the experimental group significantly outperformed the control group, suggesting that the software was useful in improving students’ pronunciation.

Hardison (2004) studied the effect of CALL assisted prosody training using computerized visual displays of pitch contours. Sixteen female native English speakers enrolled in the first semester of second year French at an American university recorded sentences and compared the pitch contours of their utterance with those of a native French speaking model for 13 40-minute sessions over a 13 week period. It was determined that computer assisted training had a significant effect on L2 prosody, as well as on segmental accuracy, indicating that the feedback provided by CAPT systems helps improve pronunciation.

Weinberg and Knoerr (2003) compared the effectiveness of using audio cassettes compared to CDROMs for pronunciation training on 61 low-intermediate French learners. Though they found there was no significant difference between the use of audio cassettes and CDROMS, they found that the subjects using the CDROMs were more satisfied with the instruction they received than the subjects who only used audio cassettes.
Precoda, Halverson, and Franco (2000) investigated the effect of receiving feedback in the form of a numerical score from speech recognition software (FreshTalk) on pronunciation. Forty five subjects studying Spanish at a community college were divided into 3 groups (Group A - software with feedback, Group B - software with no feedback, Group C - control group), with Groups A and B using the software for 3 half-hour sessions a week over a 3 week period. While Group A received feedback in the form of a numerical pronunciation score, Group B did not. There was no significant difference in the post-test scores, though the subjects in Group A reported they appreciated receiving automatic feedback.

2.5 Literature Review Summary

Though evidence showing that CAPT is an effective means of improving pronunciation is scarce (seemingly due to a lack of longitudinal studies), there is an abundance of research showing that CAPT is viewed positively by students. They seem to feel that it is an effective means of practicing pronunciation while also claiming that it is both enjoyable and motivating. Furthermore, CAPT provides the feedback that is lacking in self-study pronunciation practice, and appears to be a crucial aspect of pronunciation practice.
Though the EnglishCentral website appears to fit the bill for an effective CAPT system, there is a lack of research showing its effectiveness. In addition, despite documentation showing that CAPT systems are well received by students and learners, and that they believe in their effectiveness, there is a lack of research showing that such systems effectively improve the intelligibility of their pronunciation. Therefore, this study seeks to determine if EnglishCentral has an effect on the intelligibility of the participants’ pronunciation, as well as to determine their opinion on EnglishCentral as a self-study pronunciation tool.
CHAPTER 3 METHODOLOGY

3.1 Research Questions

This research is focused on the use of a CAPT website, www.englishcentral.com, which was used by a group of college students over the course of a semester. The purpose of this study is to determine: 1) Does pronunciation training using the website outside of the classroom help improve the intelligibility of the participants’ pronunciation? 2) Is there a correlation between site usage and increases in intelligibility? 3) How do the participants perceive the website as a tool for practicing and improving pronunciation?

3.2 Participants

The research was conducted with 79 participants (31 males and 48 females) enrolled in 2 different classes. They were in their second semester of a 2-year English program at a junior college in Seoul, South Korea. The participants were enrolled in an ‘English Conversation’ class, which met twice a week, once for 2 hours and once for 1 hour during a 16 week semester. The proficiency level of the participants was beginner-high to intermediate-low, and the aim of the course was for the students to be able to
hold a 4-minute conversation with a random partner by the end of the semester. As such, class time focused on pair and group work, in an attempt to give students time to practice their speaking skills using various grammar points and thematic discussions. The majority of participants were in their early 20s, though a few participants were returning students who had taken time off from their studies for a variety of reasons. Being educated in South Korea, all participants had received at least 6 years of English education in the Korean public school system, 3 years in middle school and 3 years in high school. Some participants had also studied English at private academies or with private tutors, and a few had studied English abroad for short times.

3.3 Data Collection and Procedures

This study consisted of 5 parts: a pre-test, the treatment, a post-test, a survey, and usage reports downloaded from the website. This section will outline the procedures for collecting data at each step.

3.3.1 Pre-test

First, before administration of the pre-tests, during the first two weeks of the semester, the participants created accounts on the EnglishCentral website,
and were given an in-class demonstration of the website and its functions. The participants were shown how to view videos and how to record their own utterances, as well as other functions available, such as the ability to compare their own utterances to those provided or to look up the meaning and pronunciation of individual words. The participants were then given 2 weeks to test out the website at home and report any problems or technical difficulties they encountered. After it was determined that all students were successfully registered and could use the website, the pre-tests, which entailed recording the participants reading 5 phonetically rich sentences, were conducted during the 6th week of the semester.

### 3.3.2 Treatment

Second, the students used the website to practice their pronunciation outside of the classroom. Students were encouraged to visit the website several times a week so as to achieve the weekly point target in small increments, but there was no way to enforce this procedure.

#### 3.3.2.1 EnglishCentral

The website www.englishcentral.com was used by the participants for a 9-
week period after the pre-test had been administered. This is a website which has a collection of videos which have been taken from the popular YouTube (www.youtube.com) website and had subtitles added. When users record a line, their utterance is assigned a grade, ranging from A+ to F, along with an accompanying numerical score. According to the website, EnglishCentral is capable of detecting dropped words, dropped phonemes, inserted phonemes, substituted phonemes, and inappropriate pauses between words. Each individual error per line deducts half a grade from the line score. Thus, a recorded utterance with no errors would receive an A+, while a recorded utterance with an inserted phoneme, a substituted phoneme, and a dropped word would receive a B.

Videos are arranged by category (daily life, music, sports, technology, etc.) and by difficulty (easy, medium, and hard). Each video is assigned a difficulty level and numerical point value based on a modified version of the Flesch-Kincaid readability formula. Thus, ‘easy’ videos are usually shorter and contain easier vocabulary, while ‘difficult’ videos usually contain more words of a more advanced variety.

Participants were given an in-class lesson on how to use the website. Then on their own, participants selected videos of interest from +800 available mostly authentic videos and watched them with the transcript at the
bottom of the screen. Participants were free to choose videos from any
category or difficulty level, but were encouraged to pick videos that were of
particular interest to them. The participants watched the video in its entirety
and then listened again, with the video pausing automatically after each line
to allow the participant to record the line. Participants were then given
immediate feedback in the form of a numerical score, with any problem
words being highlighted. Participants were free to rerecord their utterances
if they desired, but attaining a perfect score was not necessary to proceed.
Participants were also able to click on words to get definitions, as well as
compare their recorded utterances to the models provided, but were not
required to do so. The participants repeated this procedure, recording as
many lines from as many videos as was necessary for them to achieve at
least 2,000 points each week. The target of 2,000 points was decided upon
based on feedback from a preliminary study done with different participants
in the previous semester. Based on data obtained from a survey at the end of
the semester, those participants indicated that achieving 2,000 points on the
website took about 30 minutes on average. The researcher determined that
requiring the participants to accumulate more than 2,000 points per week
may have placed too much of a burden timewise on the participants.

It is theorized that this website will improve the intelligibility of the
participants’ pronunciation for a number of reasons. First, it is a means of providing sustained pronunciation practice outside of the classroom in what is believed to be an entertaining and motivating way (partially based on feedback from a previous pilot study). Second, the participants are provided with immediate feedback in the form of a numerical score, as well as notification of specific pronunciation ‘errors.’ The participants are also able to compare their utterances with the model utterances in an attempt to correct their pronunciation. Third, the videos come from a variety of sources, employing a number of accents and model pronunciations, thus exposing participants to accepted variations in English pronunciation.

Immediately after the pre-test, the participants were informed that they were required to achieve a total of 2,000 points each week over a 9-week period. The participants were informed that failure to meet the weekly target would result in a 1% deduction from their homework grade, which was worth 30% of their final grade. On a weekly basis, the researcher downloaded and maintained usage reports generated by the website.

3.3.3 Post-test

Third, in the 15th week of the semester, the post-tests were conducted in an
identical fashion to the pre-tests.

3.3.4 Surveys

Fourth, after the post-tests were conducted, the participants completed surveys about their experiences with the website.

3.3.5 Participant usage reports

Fifth, EnglishCentral generated usage reports, showing how many points each participant achieved each week, along with the number of visits, the number of videos watched, lines recorded, and total accumulated points. These usage reports were downloaded by the researcher on a weekly basis. Table 3.1 outlines the steps used for data collection in this study.

Table 3.1
Table of Procedures

<table>
<thead>
<tr>
<th>Step</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Participants completed the pre-test (Oct. 6th)</td>
</tr>
<tr>
<td>2</td>
<td>Participants used the website to practice pronunciation outside of class time for 9 weeks (Oct. 6th – Dec. 7th)</td>
</tr>
</tbody>
</table>
Participants completed the post-test (Dec. 9th)

Participants completed surveys about their experience using the website. (Dec. 9th)

Researcher downloaded subject usage reports (Weekly, every Wed.)

3.4 Data Sources

There are 4 data sources for this study. First the pre-tests and post-tests will be described, followed by the usage reports, and finally the qualitative survey.

3.4.1 Pre and post-tests

It was decided that participants would read sentences for the pre and post-tests to ensure the measure of their intelligibility was on pronunciation. Thus, other factors that could affect intelligibility, like grammar, discourse markers, and long pauses while speaking were minimized, since the definition of intelligible pronunciation for this study is limited to the ability of the listener (in this case, the rater) to understand the individual words of the participants’ utterances at the phonetic level. Furthermore, since EnglishCentral involves participants reading subtitles while recording, it was determined that this
method of data collection would suitably mimic what the participants were doing while using the website. Furthermore, Cucchiarini, Strik, and Boves (1997; 1998), and Neri, Cucchiarini, and Strik (2006; 2008) have used the reading of phonetically rich sentences in their ASR based CAPT studies, as this method ensures that all the sounds of the target language are produced and that intelligibility ratings are not affected by participants avoiding problematic sounds.

The pre-test and post-test involved participants reading 5 phonetically rich sentences, which are sentences “containing approximately uniform phoneme frequency distributions” (Gibbon, Mertins, & Moore, 2000, p. 455). The sentences came from the SCRIBE (Spoken Corpus of British English), an anglicized version of the US TIMIT 460 sentence set, which was created by replacing decidedly American words and phrases with British ones. The sentences were rated for difficulty of pronunciation for the PF-STAR British English Children’s Speech Corpus. To ensure that each subject would read a different sentence for the pre-test and the post-test (so the rater would not hear the same sentence more than once), the researcher chose to use the 200 SCRIBE sentences that were rated the most difficult, those for children 10-14 years old. If the rater were to hear the same sentences repeatedly, it would be impossible to determine if he was able to
understand the participant’s utterance due to intelligibility or due to familiarity with hearing the same sentence repeatedly. As the participants were all college students at the beginner-high to intermediate-low level, it was expected that they would not have too much trouble producing these phonetically rich sentences rated for 10-14 year old native speaking children.

To avoid having the post-tests scored harsher due to more complex or difficult phonetically rich sentences, the 200 sentences were randomized. First, the 100 SCRIBE sentences for 10-12 year olds were randomized and divided in half, 50 sentences becoming part of the pre-test and 50 becoming part of the post-test. The same thing was done with the 100 SCRIBE sentences for 13-14 year olds. Then, the 1st group of 50 sentences for 10-12 year olds was combined with the 1st group of 50 sentences for 13-14 year olds, and all 100 sentences were randomized to create the pre-test. The same procedure was repeated for the post-test. Figure 3.1 shows the steps involved in creating the pre-tests and post-tests.
Due to time constraints, it was not possible to record the pre-tests and post-tests on an individual, one-on-one basis. The recordings were made using “Nurinet-AOD” digital recorders in a language lab. The participants were randomly divided into groups of 7 or 8 in the language lab, with all participants recording into their personal devices at the same time, being spaced far enough apart so as to avoid background noise or interference from other participants. In the recording room, the participants determined at which of the 8 recording stations they would sit. The participants on the left side of the room recorded the same pre-test consisting of 5 sentences, while participants on the right side of the room recorded a different pre-test consisting of 5 different sentences. The 100 sentences for the pre-test had been divided into groups of 5 for recording purposes. The same 5 sentences were given to each group of 3-4 participants to record as a time saving measure. The same procedure was used for the post-test.
Prior to recording, after the participants had been instructed on how to operate the recording device, the researcher read the sentences to be recorded so that the participants’ reading ability had a minimal effect on the recording. The researcher modeled each sentence once, reading the first sentence for the group on the left, then the first sentence for the group on the right, to ensure that one group did not have more time to silently practice before recording began. The participants did not listen and repeat. By reading all the sentences for the participants in a staggered fashion, it was theorized that enough time passed so as to minimize any parroting. Flege, Munro, and MacKay (1995) used the same reasoning in their study, theorizing that “the delay between the model and its repetition, as well as the intervening speech material, probably prevented direct imitations from sensory memory” (p. 3127). Again, the same procedure was followed for the post-tests.

3.4.2 Rater

One native English speaking rater was employed to transcribe the randomized sentences to the best of his ability. The English speaking rater was an ESL/EFL teacher with 2.5 years experience in New Zealand, and 2.5 years experience in South Korea. An .mp3 file, consisting of randomized
sentences from both the pre-tests and post-tests was played to the rater. The rater was instructed to transcribe each utterance to the best of the ability. He was also informed that partial transcriptions were acceptable and that if he was only able to transcribe one word, that was also acceptable. Each sentence was played once and then the file was paused to give the rater sufficient time to attempt to transcribe the sentence he had heard. The procedure was repeated until all the sentences had been played and transcribed.

3.4.3 Usage reports

The website allows teachers to monitor the participants’ progress in the form of downloadable usage reports. It was possible to manually set the desired time frame and download the data for that particular time period. The data allowed the researcher to monitor which participants were meeting the weekly 2,000 point target. The usage reports also included data on total points accumulated, which were analyzed for correlations related to improvements in intelligibility scores between the pre and post-tests.
3.4.4 Surveys

Upon completion of the post-test, to determine the participants’ perception of EnglishCentral as a learning tool, the participants were given a survey consisting of both multiple choice questions and open-ended questions (see Appendix A). The survey was constructed to determine the participants’ opinions on the website in line with other current CALL research (Lomicka & Lord, 2009; Neri, et al., 2008; Sung, 2008; Weinberg & Knoerr, 2003). The survey was conducted online, using www.statcrunch.com. The multiple choice questions used a 5 point Likert scale, and were written in English but translated into Korean for administration to ensure that the participants understood the questions clearly. The participants were asked to rate their experience with the website in areas such as how accurate they believed the scoring system to be, how effective they thought the site was as a pronunciation tool, as well as how often they visited the site and how long it took them to complete their weekly goal. The participants were free to write their answers to open-ended questions in Korean or English, with answers written in Korean being translated into English for analysis. The surveys were done anonymously in an attempt to get the participants to answer honestly about their experience with the website.
3.5 Data Analysis

The following section will explain how sentences from the pre and post-tests were selected for analysis, and how they were analyzed. It will also explain how the surveys and usage reports were analyzed.

3.5.1 Pre / post-tests

There were 20 versions of the pre-test and 20 versions of the post-test, each of which was read by either 3 or 4 participants. Both the pre-tests and post-test were assigned to participants randomly. Thus, the group of students that read pre-test “1” were not the same students that read post-test “1”. For rating purposes, it was necessary to choose only one version of each pre-test and post-test. The rationale behind this was to ensure that the rater did not hear the same sentence more than once, which could have affected the results. To select which participant’s recordings would be analyzed it was necessary to correlate their pre-tests and post-tests to ensure that there was no overlap. Website usage was analyzed and the participants were ranked in order from highest usage to lowest usage. Then, the researcher chose the highest usage participant and selected that participant’s pre-test and post-test, noting which
tests had been accounted for. Then, the researcher chose the lowest usage participant and included that participant’s pre-test and post-test, again noting which tests had been accounted for. If a participant’s pre-test or post-test had already been accounted for, their data was not included. For example, the student with the highest usage completed pre-test #19 and post-test #20. Thus, both of those tests were accounted for. Next, the participant with the lowest usage was determined to have completed pre-test #2 and post-test #6, thus striking these tests of the list. This process was repeated, alternating between high usage and low usage students, until the complete list of participants had been sorted through, and when completed, seventeen pre-tests and seventeen post-tests were available for rating. It was impossible to include pre-tests 3, 9, and 18, and post-tests 2, 8, and 14. Table 3.2 shows the participants, and their corresponding pre-tests and post-tests that were included for rating.

Table 3.2

*Inclusion of Pre-tests and Post-tests*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Pre-Test</th>
<th>Post-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>17</td>
</tr>
</tbody>
</table>
Using Audacity, the 5-sentence pre-tests and post-tests were randomized and combined into one .mp3 file. Half a second before each utterance, a brief tone was inserted to indicate to the rater when the next utterance would begin. Ten seconds of silence were inserted between each utterance to give the rater time to attempt to transcribe the utterance he had just heard. The researcher played the .mp3 file for the rater who was wearing a pair of headphones. If ten seconds was not long enough for the rater to transcribe the utterance, the researcher paused the file to allow the rater time to complete his attempted transcription.

The researcher then analyzed the attempted transcriptions, counting the number of correctly transcribed words for each utterance. A word was considered correctly transcribed if it matched the script that had originally been provided to the participants. In some cases, the participants misread
the utterances from the pre-test or post-test, which involved adding or dropping phonemes from the end of words (‘remain’ instead of ‘remained’) or in some cases substituting entirely different words (‘describe’ for ‘prescribe’). In these cases, the word was considered correctly transcribed if it matched the audio rather than the written version of the pre-test or post-test. In cases where the rater only partially transcribed a word, half a point was awarded (‘contain’ instead of ‘contained’). These values were converted into percentages since the number of words in each utterance was different.

The percentage scores for each subject’s pre-test and post-test were calculated and then analyzed using t-tests and Pearson correlation product coefficients to determine if there was an increase in the intelligibility of their pronunciation between the pre-test and the post-test, as well as if there was any correlation between increases in intelligibility and amount of usage of the English website (determined by usage reports generated by the site).

### 3.5.2 Usage reports

Data from the usage reports were analyzed to determine if there was any correlation between increases in intelligibility and participant use of the website. Participants were given a minimum weekly point total to achieve, but several participants greatly surpassed their minimum weekly goal, while
others barely achieved their weekly goal. Thus the researcher attempted to determine if the amount of effort expended on pronunciation practice corresponded with the participants’ intelligibility scores. This was achieved by taking the mean score of the participants and then dividing them into a high usage group (HUG) and a low usage group (LUG) and then comparing the data from the HUG and the LUG.

3.5.3 Surveys

The participants’ surveys were then analyzed to determine their perceptions of EnglishCentral. The multiple choice questions were tallied and converted to percentages, while the open-ended questions were analyzed for recurring trends. The researcher examined the answers and grouped them into similar categories, which were then tallied in order to determine the most common answers to the open-ended questions.
CHAPTER 4 RESULTS

4.1 Pre Tests and Post-tests

The pre and post-test scores were analyzed using a t-Test, however it was determined that there was no significant increase between the pre-tests and post-tests ($p=0.07$, $t=-0.155$) for the participants as a whole group. Figure 4.1 shows the pre-test and post-test scores for the 17 participants included in the study.

![Figure 4.1. Pre-test and post-test scores for all 17 participants.]

4.2 Site Usage and Post-tests

The EnglishCentral website generated usage reports based on student activity and usage. A comparison of pre-tests for the HUG (n=9) and LUG (n=9) resulted in values of $p=0.17$ and $t=+0.97$, indicating that initially there was
no significant difference between the 2 groups.

Their post-test scores were then compared with their total points accumulated during the 9-week treatment. Table 4.1 shows the post-test scores and the total points accumulated on the EnglishCentral website for the HUG, while Table 4.2 shows the post-test scores and total points accumulated for the LUG.

**Table 4.1**

*HUG- Post-Test / Usage Correlation*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Post-test</th>
<th>Usage (total points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>92.38</td>
<td>76795</td>
</tr>
<tr>
<td>2</td>
<td>92.62</td>
<td>48366</td>
</tr>
<tr>
<td>3</td>
<td>52.98</td>
<td>31914</td>
</tr>
<tr>
<td>4</td>
<td>57.96</td>
<td>28700</td>
</tr>
<tr>
<td>5</td>
<td>72.77</td>
<td>27982</td>
</tr>
<tr>
<td>6</td>
<td>96.67</td>
<td>26302</td>
</tr>
<tr>
<td>7</td>
<td>65.38</td>
<td>24823</td>
</tr>
<tr>
<td>8</td>
<td>88.71</td>
<td>24791</td>
</tr>
<tr>
<td>9</td>
<td>80.57</td>
<td>24183</td>
</tr>
</tbody>
</table>

**Table 4.2**

*LUG- Post-Test / Usage Correlation*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Post-test</th>
<th>Usage (total points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>92.50</td>
<td>22800</td>
</tr>
<tr>
<td>11</td>
<td>45.56</td>
<td>22392</td>
</tr>
<tr>
<td>12</td>
<td>46.58</td>
<td>22078</td>
</tr>
<tr>
<td>13</td>
<td>81.00</td>
<td>21886</td>
</tr>
<tr>
<td>14</td>
<td>94.29</td>
<td>21832</td>
</tr>
</tbody>
</table>
Figure 4.2 is a comparison of the mean usage scores for the HUG and the LUG.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>84.32</td>
<td>21389</td>
</tr>
<tr>
<td>16</td>
<td>45.50</td>
<td>20839</td>
</tr>
<tr>
<td>17</td>
<td>54.75</td>
<td>16982</td>
</tr>
</tbody>
</table>

*Figure 4.2. Mean usage scores for the HUG and the LUG.*

The Pearson correlation coefficient was calculated for each group, returning values of $r=0.38$ and $p=0.15$ for the HUG, and values of $r=0.31$ and $p=0.23$ for the LUG, indicating a medium correlation, but not a significant relationship, between post-test scores and site usage for both groups.

Next, to further explore the relationship between usage and post performance, the pre-test scores for both groups were compared using a t-Test, yielding values of $p=0.17$ and $t=+0.97$, indicating that there was no significant difference on the pre-tests between the 2 groups. The post-test
scores for both groups were then analyzed to determine if there was any significant difference between them. A t-Test indicated that there was no significant difference, with values of $p=0.16$ and $t=+1.05$.

Finally, the pre-tests and post-tests for each group were compared using a t-Test to determine if there was a significant increase between the pre-test and the post-test. There was no significant increase for the LUG ($p=0.25$, $t=-0.7$), and while the increase for the HUG was not significant ($p=0.06$, $t=-1.76$), it is approaching significance. Figure 4.3 shows a comparison of the pre and post-tests for the low usage and high usage groups.

![LUG / HUG Pre-Test / Post-Test Comparisons](chart.png)

Figure 4.3. Pre-test and post-test comparisons between LUG and HUG.

4.3 Surveys

Of the 79 original participants, 65 answered the anonymous online survey. The majority of participants had been studying English for between 7 and 11
years (65%), though 67% of them had not supplemented their English classes outside of their college studies, and 95% of them had never studied English in a country outside of Korea. While 61% of the participants found the website somewhat helpful for practicing pronunciation, 32% of them found it very helpful. In addition, 20% of the students found it very effective, with 49% of them finding it quite effective for practicing pronunciation. Furthermore, 26% of the students found the assignments very interesting, while 46% of them found it quite interesting. Judgments on the accuracy of the rating system were split with 25% finding it pretty accurate, 43% finding it somewhat accurate, and 26% finding it not very accurate. A majority of participants felt the website improved their pronunciation, with 34% indicating it probably helped, and 49% responding that it might have helped. Finally, 79% of the participants spent 30 minutes or less reaching the 2000 point weekly target, with 53% visiting the website on an average of once a week, and 42% visiting 2 to 3 times a week. Figure 4.4 shows participants’ opinions on using the website as a pronunciation training tool.
Figure 4.4. Participants’ opinions on using EnglishCentral as a pronunciation training tool.

When asked if they had encountered any problems using the website, the most common answer (21%) from participants was that no problems had been encountered. The most commonly reported problems where that the system did not recognize pronunciation accurately (18%), followed by complaints about the inaccuracy of the scoring system (11%).

The most common suggestion was for more interesting/diverse videos (22%). While 20% of the respondents felt the website helped them or was useful, 11% wanted more accurate recognition of their utterances.
CHAPTER 5 DISCUSSION

5.1 Research Question 1

With regard to the 1st research question, there is no conclusive evidence that the use of this website outside of the classroom helped improve the intelligibility of the participants’ pronunciation. The intelligibility scores for 2 members of the high usage group and 3 members of the low usage group were actually lower on the post-test than on the pre-test. Though 3 of the participants saw increases of more than 20% between their pre-test and post-test, the average increase was only 4.87%.

This finding is in line with a number of other research studies conducted on CAPT, such as Neri, et al. (2008), Weinberg & Knoerr (2003), and Precoda, et al. (2000), which have failed to find conclusive proof that pronunciation training with CAPT is effective. There are a number of possible explanations for these results.

As stated earlier, frequent, motivated practice over a long period of time is necessary to change pronunciation habits. However survey data indicated that 78.78% of the participants spent 30 minutes or less a week achieving their weekly target and that more than 50% of the participants only used the site once a week to achieve their weekly point total. Only 5.5% of
the participants used the site 4 or more times a week. This suggests that perhaps the participants were not using the website frequently enough to exact any change in the intelligibility of their pronunciation. This view was echoed by one participant, who commented that “The amount of homework is not enough to improve my skill. It only took 10 minutes to get to 2,000 points. The target score should be over 20,000 points.”

Furthermore, the relatively short time frame of 9 weeks may not have been long enough to see an improvement in overall intelligibility. Perhaps if the participants had been instructed to concentrate on only a few phonemes it would have been possible to notice an increase in the intelligibility of those phonemes.

5.2 Research Question 2

Despite a lack of definitive proof that using EnglishCentral improved participants’ intelligibility, there does seem to be some correlation between site usage and increases in intelligibility scores. Analysis of both the HUG and LUG showed there was a medium correlation between post-test scores and site usage (HUG – \( r=0.383 \), LUG – \( r=0.306 \)). These results seem to indicate that regardless of the site’s effectiveness, focused pronunciation
practice does seem to yield benefits. Figure 5.1 shows the results for the pre-tests and post-tests for the HUG.

![HUG Pre vs Post Test](image)

*Figure 5.1. Pre-test and post-test results for the HUG.*

### 5.3 Research Question 3

The survey data collected to answer the third research question seems to support previous studies related to participants’ perceptions of using such CALL tools for practicing pronunciation. Despite a lack of qualitative data indicating that the website actually improved their intelligibility, the majority of students found the website helpful, effective, and interesting. Comments such as “I was able to practice my pronunciation while enjoying diverse subjects,” and “It was fun to choose the field that I want to improve and it has helped my pronunciation immensely,” support findings by Neri, et al (2008), Sung (2008), Engwall & Balter (2007) and Weinberg & Knoerr (2003), who all found that students had positive reactions to CAPT systems, deeming them both interesting and useful. Despite the fact that 30% of the
participants rated the EnglishCentral scoring system as ‘not very accurate’ or ‘not accurate at all,’ (“Even when I pronounced it incorrectly, I got good scores or vice versa.”) more than a third of the participants (34%) felt the site ‘probably helped’ improve their pronunciation, while 3.3% felt the site ‘definitely helped’ (“It was very helpful and I am going to continue using this site,” “It has improved my pronunciation a lot. I want to recommend this site to other people.”). This seems to indicate that the participants believe in the effectiveness of this tool, regardless of how accurately it actually assesses users’ pronunciation.
CHAPTER 6 CONCLUSION

6.1 Implications for Practice

The findings from this study seem to indicate that using EnglishCentral outside of the classroom is a fun, motivating, and useful way for students to practice pronunciation. Though there is no definitive proof that this site improves overall intelligibility, students seem to enjoy using the website and believe that it is an effective tool. The website has recently undergone an upgrade, which includes a paid part of the site, which allows students to track progress on individual phonemes. The functionality that was used in this study is still available for free; however the new paid version seems to provide for more individualized monitoring and a more focused approach, combined with the ability to practice vocabulary at the same time. As such, it would appear that this website would be useful to most students studying ESL/EFL, at least on some level.

6.2 Limitations

There were a number of limitations to this study. One of the biggest limitations was that even though the original group consisted of 79 participants, due to the way the pre-tests and post-tests were administered,
the data from only 17 participants were available for quantitative analysis. As such, non-parametric tests may have been better suited to the small sample size.

The use of several raters from varying backgrounds may have perhaps been better able to assess the participants’ overall intelligibility. As acknowledged in the literature review section, intelligibility is subjective, and thus employing a number of native English-speaking and non-native English-speaking raters would have been a better yardstick for measuring intelligibility, rather than the single rater used in this study.

Though the way the participants practiced pronunciation using EnglishCentral was done by reading subtitles, obtaining speech data by having participants read phonetically rich sentences was perhaps not the best way to get a natural speech sample. The choice to use sentences from the SCRIBE corpus rated for 10 to 14-year-olds also seems to have been somewhat above some participants’ level. There were a number of instances of participants stumbling over unfamiliar words despite having the researcher read the sentences out loud before recording. Thus, devising a means of obtaining a more natural speech sample may have yielded more accurate judgments about overall intelligibility.
6.3 Future Research

As stated earlier, there is now a new paid version of the EnglishCentral website available. This new paid version allows users to focus on individual sounds, especially those that may be particularly troublesome for individual users. Paid users are also able to view videos showing basic sound formation and videos providing advanced tips and tricks for pronunciation. Users may also track their progress on the improvement of individual phonemes. This being the case, future research based on the paid version of the website may yield the quantitative proof missing from this study by allowing users and the researcher to focus on particular problematic phonemes rather than overall intelligibility. Furthermore, future research should include a higher overall weekly point target to ensure that users are practicing pronunciation more frequently. Rather than a large weekly point total, perhaps a smaller daily point total would yield quantitative data showing an increase in intelligibility.

6.4 Conclusion

This research sought to determine if EnglishCentral is an effective means of improving the intelligibility of the participants’ pronunciation.
Though unable to provide definitive proof that using EnglishCentral outside of the classroom improved the intelligibility of the participants’ pronunciation, this study supports research showing that participants find CAPT systems motivating, enjoyable, and believe in their effectiveness. The medium correlation between post-tests scores and usage scores seems to indicate that using EnglishCentral did indeed have some effect on the participants’ pronunciation. As such, it would appear that EnglishCentral is a useful tool for both teachers and students who are looking for an enjoyable way to practice pronunciation, removing the monotony of ‘typical’ listen and repeat pronunciation drills. Though the participants did seem to enjoy using the website, and believed in its usefulness, further research is necessary to provide quantitative data supporting the theory that such CAPT systems are an effective means of improving the intelligibility of pronunciation. Future studies on EnglishCentral and other CAPTs should seek to adopt a more focused approach to determine if such tools are effective when used to combat a set of targeted problem sounds. The new features of the paid service at EnglishCentral could also be studied to determine if the paid version is more effective than the free version.
References:


Minematsu, N. (2004). *Pronunciation assessment based upon the compatibility between a learner's pronunciation structure and the target language's lexical structure*. Paper presented at the


APPENDIX A: SURVEY

Please complete this survey as honestly as you can. This survey is anonymous and will not affect your grade, although I will deduct 1 point from your homework score if you don’t do it by 11:59 pm, Friday, December 10th. I will check with your class captain to see who hasn’t completed the surveys. **Please remember, your answers will not affect your grade in any way.**

1. My survey number is (given to me by my class captain): _____

2. I have been studying English for about ____ years. ( )
   a. less than 5 years (5 )
   b. between 5 – 7 years (5 7 )
   c. between 7-9 years (7 9 )
   d. between 9 – 11 years (9 11 )
   e. more than 12 years (12 )

3. I have taken supplementary English classes for about ____ a week over the past 4 months. ( 4 ?)
   a. 0 hours ( )
   b. 1 hours (1)
   c. 2 hours (2)
   d. 3 hours (3)
   e. 4 hours or more (4 )

4. I have lived in or studied English in a country other than Korea for _______. ( ?)
   a. 0 months ( )
   b. 1-6 months (1 6 )
   c. 6 months – 1 year (6 1 )
   d. 1 – 2 years (1 2 )
   e. more than 2 years (2 )

5. I found the English Central homework assignments__________ for
practicing pronunciation. (______________________.)

a. very helpful ( )
b. somewhat helpful ( )
c. not very helpful ( )
d. not helpful at all ( )
e. I don't know (didn't do the assignments) ( [ ])

6. I found the English Central homework assignments________ for practicing pronunciation. (______________________.)

a. very effective ( )
b. quite effective ( )
c. somewhat effective ( )
d. only a little effective ( )
e. not effective at all ( )

7. I found the English Central homework assignments _______. (______________________.)

a. very interesting ( )
b. quite interesting ( )
c. somewhat interesting ( )
d. only a little interesting ( )
e. not interesting at all ( )

8. I ________ continue to use English Central on my own. ( )

a. will ( )
b. will probably ( 70% )
c. might ( 50% )
d. probably won’t ( 20% )
e. won’t ( )

9. How would you rate the English Central scoring system? ( }
66

a. very accurate ( )
b. pretty accurate ( )
c. somewhat accurate ( )
d. not very accurate ( )
e. not accurate at all ( )

10. I think practicing with English Central _______ improve my pronunciation. ( )
   a. definitely helped ( )
b. probably helped ( )
c. might have helped ( )
d. probably didn’t help ( )
e. definitely didn’t help ( )

11. How much time a week did you spend reaching the 2,000 point weekly target? ( 2,000 )
   a. less than 15 minutes (15 )
b. 15-30 minutes (15 30 )
c. 30-45 minutes (30 45 )
d. 45-60 minutes (45 60 )
e. more than 60 minutes (60 )

12. How many times a week (on average) did you visit the English Central site to complete your homework? ( )
   a. once a week (1 )
b. 2-3 times a week (2~3 )
c. 4-5 times a week (4~5 )
d. 5-6 times a week (5~6 )
e. every day ( )

13. Did you have any problems using the English Central website? If so,
please explain briefly (you can answer in English or Korean).
([   ,
     ])

14. Do you have any other comments or suggestions related to English Central? (You can answer in English or Korean).
    ? [   ,   ]

APPENDIX B: PRE TESTS
Pre-test - 1
Are holiday vouchers available to us?
Challenge each general’s intelligence.
Which church do the smiths worship in?
Shipbuilding is a most fascinating process.
In the long run, it pays to buy quality clothing.

Pre-test - 2
The clumsy customer spilled some expensive perfume.
Michael coloured the bedroom wall with crayons.
Children can consume many fruit tarts in one sitting.
Rob sat by the pond and sketched the stray geese.
The annoying rabbits slipped into Phil’s garden every night.

Pre-test - 3
Rock-and-roll music has a great rhythm.
Would you allow acts of violence?
A young mouse scampered across the field and disappeared.
Withdraw only as much money as you need.
They own a big house in the remote countryside.

Pre-test - 4
A leather handbag would be a suitable gift.
Women may never become completely equal to men.
Draw every outer line first, then fill in the interior.
The best way to learn is to solve extra problems.
George seldom watches daytime television.

Pre-test - 5
Laugh, dance and sing if fortune smiles upon you.
Each stag surely finds a big fawn.
The proof that you are seeking is not available in books.
If people were more generous there would be no need for welfare.
The bonnet of the jeep was steaming in the hot sun.

Pre-test - 6
They remained lifelong friends and companions.
‘Gremlins’ is yet another exciting movie by Stephen Spielberg.
My ideal morning begins with hot coffee.
First add milk to the shredded spinach and grated cheese.
The eastern coast is a place for pure pleasure and excitement.

Pre-test - 7
Clasp the screw in your left hand.
Gregory and Tom chose to watch cartoons in the afternoon.
Butterscotch fudge goes well with vanilla ice cream.
Jeff’s toy go-cart never worked!
Even I occasionally get the Monday blues!

Pre-test - 8
Kindergarten children decorate their classrooms for all holidays.
Young people participate in athletic activities.
My instructions desperately need updating.
Bob found more clams at the rock-pool’s edge.
Would a tomboy often play outdoors?

Pre-test - 9
A huge power cut rarely occurs.
Bob bandaged both wounds with the skill of a doctor.
Should giraffes be kept in small zoos?
Only the best players enjoy popularity.
We experience distress and frustration working for our degrees.

Pre-test - 10
It’s illegal to postdate a cheque.
The avalanche triggered a minor earthquake.
Gwen grows green beans in her vegetable garden.
Count the number of teaspoons of soy sauce that you add.
When peeling an orange, it is hard not to spray juice.

Pre-test - 11
Kangaroo point overlooked the ocean.
Thick glue oozed out of the tube.
Call an ambulance for medical assistance.
Who authorised the unlimited expense account?
Where were you while we were away?

Pre-test - 12
Good service should be rewarded by big tips.
Don’t look for unwanted valuables in the bank vault.
The hallway opens into a huge chamber.
Traffic jams occur on many relief trunk roads.
The morning dew on the spider’s web glistened in the sun.

Pre-test - 13
The giant redwoods shimmered in the glistening sun.
Nothing is as offensive as innocence.
Be careful not to plough over the flower beds.
That diagram makes sense only after much study.
Special task forces rescue hostages from kidnappers.

Pre-test - 14
These exclusive documents must be locked up at all times.
Penguins live near the icy Antarctic.
Cement is measured in cubic yards.
Gary attacked the project with extra determination.
The schoolgirl couldn’t disguise her feelings about the emergency conditions.

Pre-test - 15
Tim takes Sheila to see American movies twice a week.
It’s healthier to cook without sugar.
His sudden departure shocked the cast.
Ralph prepared red salmon with fresh lemon sauce for dinner.
Please sing just the ‘Cotton Club’ theme.

Pre-test - 16
Our experiment’s positive outcome was unexpected.
A lawyer was appointed to execute her will.
Military personnel are expected to obey government orders.
John’s brother repainted the garage door.
Addition and subtraction are learned skills.

Pre-test - 17
Every cab needs repainting often.
The Boston Ballet overcame their funding shortage.
The gorgeous butterfly ate a lot of nectar.
The carpet cleaners shampooed our oriental rug.
Young children should avoid exposure to contagious diseases.

Pre-test - 18
Combine all the ingredients in a large bowl.
His failure to open the store by eight cost him his job.
Once you finish greasing your chain, be sure to wash thoroughly.
The cow wandered from the farmland and became lost.
The system may break down soon so save your files frequently.

Pre-test - 19
The high security prison was surrounded by barbed wire.
The Thinker is a famous sculpture.
Smash lightbulbs and their cash value will diminish to nothing.
I’d tried the tube, but I hadn’t enough change.
The emperor had a mean temper.

Pre-test - 20
A chosen few will become generals.
Trespassing is forbidden and subject to penalty.
Growing well kept gardens is very time consuming.
Iris thinks this zoo has eleven Spanish zebras.
Keep the thermometer under your tongue!

APPENDIX C: POST-TESTS
Approach your interview with steady composure.
The two artists exchanged autographs.
Critical equipment needs proper maintenance.
Herbert’s birthday frequently occurs on Thanksgiving.
A moth zig-zagged along the path through Oliver’s garden.

Withdraw only as much money as you need.
Those answers will be straightforward if you think them through carefully first.
Serve the coleslaw after I add the oil.
Basketball can be an entertaining sport.
She wore warm, fleecy woolen overalls.

Shaving foam is a popular item on Hallowe’en.
Doctors prescribe drugs too freely.
Do you have the yellow ointment ready?
Project development was proceeding too slowly.
Cliff’s display was misplaced on the screen.

Destroy every file related to my audits.
A huge tapestry hung in her hallway.
He will allow a rare lie.
Russ saw pine trees and redwoods on his walk through the National Forest.
She always jokes about too much garlic in his food.
How good in your endurance?
Only the best players enjoy popularity.
Youngsters love candy floss as a treat.
‘Highway’ and ‘freeway’ mean the same thing in American English.
If Carol comes tomorrow, arrange for her to have a meeting at two.

Post-test - 6
Even a simple vocabulary contains symbols.
Just drop notes in any suggestion box.
Which theatre shows ‘Mother Goose’?
May I order an apricot mousse after I have eaten dinner?
Rationalise all errors.

Post-test - 7
She slipped and sprained her ankle on the steep slope.
Cottage cheese with chives is delicious.
The cranberry bog gets very pretty in autumn.
This brochure is particularly informative for a prospective buyer.
Of course you can have another tuna fish sandwich.

Post-test - 8
How oily do you like your salad dressing?
Do you hear the sleigh bells ringing?
The news agency hired a great journalist.
Her wardrobe consists of only skirts and blouses.
The water contained too much chlorine and stung in his eyes.

Post-test - 9
Chocolate and roses never fail as a romantic gift.
That noise problem grows more annoying each day.
Thomas thinks a larger clamp solves the problem.
How much allowance do you get?
The prowler wore a ski mask for a disguise.

Post-test - 10
They enjoy it when I audition.
His scalp was blistered from today’s hot sun.
Fill that can and spray the screen with fresh spring water.
Please take this dirty table cloth to the cleaners for me.
The singer’s finger had a splinter.

Post-test - 11
Cheap stockings lather the first time they’re worn.
The fog prevented them from arriving on time.
A doctor was in the ambulance with the patient.
Last year’s glass shortage caused steep price increases.
I itemise all accounts in my agency.

Post-test - 12
How permanent are their records?
Regular attendance is seldom required.
The fifth jar contains big, juicy peaches.
Spherical gifts are difficult to wrap.
How do oysters make pearls?

Post-test - 13
Put the butcher’s block in the garage.
The bungalow was pleasantly situated near the shore.
Remove the splinter with a pair of tweezers.
I’d rather not buy these shoes than be over charged.
Move the hostage nearer to the large window.

Post-test - 14
Steve wore a bright red cashmere sweater.
Eat your raisins outside on the porch steps.
Bright sunshine shimmers on the ocean.
Elderly people are often excluded.
She encouraged her children to make their own Hallowe’en costumes.

Post-test - 15
Flying standby can be practical if you want to save money.
Irish youngsters eat fresh kippers for breakfast.
Drop five forms in the box before you go out.
It’s hard to tell an original from a forgery.
Call an ambulance for medical assistance.

Post-test - 16
Ralph controlled the stopwatch from the back seat.
Those thieves stole thirty jewels.
The frightened child was gently subdued by his big brother.
The drunkard is a social outcast.
Only lawyers love millionaires.

Post-test - 17
Valley lodge yearly celebrates the first calf born.  
Get a tabby cat to keep the rodents away.  
They all enjoy ice cream sundaes.  
Too much curiosity can get you into trouble.  
How would you evaluate this algebraic expression?

Post-test - 18
As a precaution, the outlaws bought gunpowder for their stronghold.  
Reading in poor light gives you eyestrain.  
They assume no burglar will ever enter here.  
Non-profit making organisations often have frequent fundraising events.  
December and January are nice months to spend in Miami.

Post-test - 19
Remember to allow identical twins to enter freely.  
Steve collects rare and novel coins.  
The hedgehog clearly saw his shadow, but stayed out only a moment.  
Draw each graph on a new axis.  
I honour my mum.

Post-test - 20
Are you looking for employment in April?  
Any contributions will be greatly appreciated.  
A large household needs lots of appliances.  
Bob’s gold bracelet was a graduation present.  
Almost all colleges are now co-educational.

Comments:

I just read through the thesis. It is interesting that the author used the same measure of intelligibility that we are considering. Unfortunately, the methodology was weak in certain respects, some of which he recognized. I
suspect that, taken together, these weaknesses account for his negative findings.

1. He threw out almost 80% of his data, leaving only 17 subjects -- a very small sample.
2. He had only one rater (transcriber), which was why he discarded so much data--he very reasonably didn't want the rater to transcribe the same sentence more than once. He should have just used multiple transcribers. Since he has the recordings, he could actually still pursue this.
3. His rater was an ESL teacher -- arguably such a person would be unusually skilled at this task and perhaps too effective at compensating for poor pronunciations. Perhaps he should have used naive raters.
4. His statistical analysis was incomplete and somewhat inconsistent. It could be improved and might well then show a more positive result. For one thing, I would try using a nonparametric test like a signed rank test instead of a t test. Also, he correlated site usage with post test transcription accuracy instead of with pre to post test improvement.
5. The pre and post tests were too small -- 5 sentences is an awfully small sample for assessing intelligibility.

- 70 students

- Took pretest sentences from TIMMIT corpus/SCRIBE Corpus; contains phonetically rich sentences (including all 36 sounds); asked students to use EnglishCentral for 9 weeks

  200 sentences leveled; 11-14 years old level;
  phonetically rich random sentences

- Had a native rater rate after 9 weeks, by simple transcription;
  listen to utterance just once; measured performance by number of correctly transcribed words; converted to percentage and then tallied results

- No significant improvement between pre and post test, although some more positive results on students who were very active.