TRAINING SPANISH SPEAKERS IN THE PERCEPTION AND PRODUCTION OF ENGLISH VOWELS

by

Grisel María García Pérez

B.A., Language Institute Máximo Gorki, 1977
B.Ed., Language Institute Pablo Lafargue, 1982
M.A., Simon Fraser University, 1992

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department

of

Linguistics

© Grisel María García Pérez 2003

SIMON FRASER UNIVERSITY

July 2003

APPROVAL

Name: Grisel María García Pérez

Degree: Doctor of Philosophy

Title of thesis: Training Spanish Speakers in the Perception
                                             and Production of English Vowels

Examining Committee: Dr. Trude Heift
                                             Chair

                                             Dr. Juan Manuel Sosa
                                             Senior Supervisor
                                             Associate Professor

                                             Dr. Murray Munro
                                             Supervisor
                                             Associate Professor

                                             Dr. Teresa Kirschner
                                             Supervisor
                                             Professor Emeritus

                                             Dr. Zita McRobbie
                                             Internal Examiner
                                             Associate Professor

                                             Dr. Bernard Schulz-Cruz
                                             External Examiner
                                             Associate Professor

Date Approved: July 24, 2003

PARTIAL COPYRIGHT LICENCE

Title of Thesis/Project/Extended Essay:

TRAINING SPANISH SPEAKERS IN THE PERCEPTION AND PRODUCTION OF ENGLISH VOWELS

Author:

(Signature)

Grisel María García Pérez
(Name)

July 24, 2003
(Date)

ABSTRACT

This thesis investigates the effects of training native speakers of Spanish in the perception and production of the English pairs of vowels /i/-/ɪ/, /u/-/ʊ/, and /ɑ/-/ʌ/ in a regular ESL classroom setting. Thirty-two adult native Spanish speakers, sixteen in the control group and sixteen in the experimental group, participated in the study.

The experimental design included a pretest-posttest procedure in order to compare the subject's performance before and after training. Perception was tested using a minimal pair forced choice task including multiple samples of the three pairs of vowels. Production data was provided by reading 15 sentences and a paragraph, all of which contained the target vowel contrasts.

Over a three-week training period, the subjects in the experimental group were given instruction on how to identify and produce the English pairs of vowels /i/-/ɪ/, /u/-/ʊ/, and /ɑ/-/ʌ/. Their progress was tested through quizzes at the end of each week. No recordings were used during the lessons.

The effect of training on perception was demonstrated through a direct comparison between the scores on the pretest and posttest. Analysis reveals a significant improvement (from 60.1% to 83.3%) in the subjects' performance as an effect of training. A mixed design ANOVA with 1 between factor (group) and 2 within factors (vowel and time) shows a significant group x time interaction $(p<.05)$.

The effect of training on production was assessed through a category goodness test and a comprehensibility test. Overall, there was no significant improvement as an effect of training.

The study provides evidence that Spanish speaking adults who learn ESL can be taught to perceive certain vowels with more accuracy in a regular classroom environment. Accuracy gained for perception in the training was not transferable to production. The study suggest that exposure to the language might have been an important factor affecting these results in production.

For Sergio and Daniel

ACKNOWLEDGEMENTS

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

Chapter II

Chapter III

CHAPTER I

TEACHING PRONUNCIATION

1.1 INTRODUCTION

When students are presented with phones that are not used in their own language, they typically show performance that is not as good as a native speaker of the language from which the phones were selected (Bradlow and Pisoni, 1997; Munro, Flege, and Mackay, 1996; Schmidt, 1996; Rochet, 1995; Munro, 1993; Werker, 1989). From a Second Language Teaching (SLT) point of view, this issue has both practical and theoretical implications. In practice, the teacher can foresee the difficulties the students may experience distinguishing and producing a specific phonetic contrast in the second language. Theoretically, the teacher can look for the conditions that facilitate the development of the second language student's ability to differentiate and produce the new phonetic categories.

In the present study, training in the perception and production of non-native vowel contrasts was given to the participants in a regular classroom setting. Over a three week training period, thirty-two adult native Spanish speakers were trained to identify and produce the English pairs of vowels /i/-/ɪ/, /u/-/ʊ/, and /ɑ/-/ʌ/.

The information in the thesis has been organized in five Chapters. Chapter I presents a brief history of how the teaching of pronunciation has been approached by the

different methods in second language teaching (SLT). Following this, there is a review of current trends in the teaching of pronunciation, along with some discussions on research related to perceptual training. The end of the chapter presents the contrasting features between English and Spanish vowel sound systems, the research questions formulated for this thesis, and the possible answers to each of these questions.

Chapter II describes in detail how the study was conducted. It starts with a description of all the participants. Subsequently, it describes all the materials used in the experiment and concludes with a sequence of the steps taken to carry out the study.

Chapters III and IV present the results of the study. Chapter III lays out the data related to the perceptual training of the English vowel contrasts /i/-/ɪ/, /u/-/ʊ/, and /ɑ/-/ʌ/. The second part provides a detailed analysis of the results. The data related to production are presented and analyzed in Chapter IV.

Chapter V concludes with an overview of the study establishing a comparison between perception and production. Then it discusses the factors which may have contributed to the results of the study. The chapter concludes with the strengths and weaknesses of the experiment, and future directions are addressed in the final thoughts.

1.2 BRIEF HISTORY OF APPROACHES AND METHODS IN SECOND LANGUAGE TEACHING

Pronunciation is one of the language phenomena that distinguishes native speakers from non-native speakers. However, it was for some time considered largely irrelevant in Second Language Teaching (SLT) since more emphasis was given to vocabulary and grammar (Goodwin, 2001; Nunan, 2000; Kelly, 1969). Interest in

pronunciation has gradually increased and this can be seen through a review of the history of some approaches and methods in language teaching.

The 19th century gave birth to the Grammar Translation Method which ignored the teaching of pronunciation. The study of a language was carried out through a detailed analysis of rules of grammar and the application of this knowledge to the task of translating sentences into and out of the target language. Reading and writing were given a crucial role without systematic attention to speaking or listening. (Richards & Rodgers, 1986).

In 1886, the International Phonetic Association was formed, and its members proposed and developed the International Phonetic Alphabet. Important followers of Phonetics[1] in this period were Paul Passy and Henry Sweet. Teachers benefited a great deal from this alphabet because it was the first time they had a consistent system to refer to when they wanted the students to have a one-to-one correspondence between the written symbol and the sound it represented.

Along with the appearance of the International Phonetic Alphabet, a number of scholars started pursuing phonetic spelling. Leopold Bahlsen, for example, thought that languages had to be represented phonetically while Brigham Young had many local books and records transcribed into a locally developed phonetic script called the Deseret Alphabet. An increase of representations of linguistic sounds brought about publications like *The History of English Sounds* and *a Handbook of Phonetics* by Henry Sweet. Around this time, the Phonetic Method appeared. The method rapidly lost popularity because of the

---

[1] The branch of Linguistics that examines the inventory and structure of the sounds of a language (O'Grady and Dobrovolsky, 1996).

lack of teaching philosophy, intellectual focus in the language programs, and the heavy requirements for linguistic expertise on the part of the teachers.

By the end of the 19th century, reformers like Gottlieb Heness and Lambert Sauveur proposed a series of principles emphasizing the way children learnt a language –that is, they advocated naturalistic principles of language learning. These principles provided the foundation for what was known as the Direct Method, the official language teaching method in France and Germany in 1902, and the method used in American schools in 1920 (Bowen, Madsen, & Hilferty, 1985). The Direct Method paid close attention to correct pronunciation through intuition, imitation, and repetition. The students were asked to imitate a model with accuracy, and grammar was taught inductively.

Between 1920 to 1930, in an attempt to develop a more scientific basis for an oral approach, a group of British applied linguists developed The Oral Approach and Situational Language Teaching. The leaders of this movement were Harold E. Palmer and A. S. Hornby. The fundamental characteristic of this approach was that the language was first presented orally and then in the written form. The students practiced controlled sentence patterns linked to situations in which the structures could be used. These approaches advocated the correct formation of speech habits through drills while accuracy in pronunciation and grammar was crucial.

In 1939, the University of Michigan created the first English Language Institute in the United States. Charles Fries was the director of the institute and an advocate of structural linguistics. Since structural linguistics proposed that the primary medium of language is oral, it was assumed that speech had a priority in language teaching. In order to learn a language, the students had to master the elements of that language. Part of this

included learning how a language was combined, from the level of the phoneme[2] to morpheme, word to phrase, and phrase to sentence. Based on these assumptions, language teaching focused first on speech, whereas the written component of the language was left for a later stage in the learning process.

When the United States became involved in the Second World War, its government suddenly needed people who were fluent in different languages to work as interpreters or translators. An extensive and intensive training was needed for this massive operation, but there was a limited amount of time. The army reached a formal agreement with some universities in the States to provide special language training designed by army linguists. This enterprise originated what is known as the Army Specialized Training Program (ASTP).

The program was welcomed by many linguists who started calling it the Army Method. In the first year, the teachers trained 15 000 students in 55 colleges and universities throughout the country. Students' fast acquisition of the foreign language was widely seen as an enormous success of the army's language programs. The method reflected the conviction of linguists like Bloomfield, who argued that the teaching of speech had to be given a primary importance.

With a focus on the sounds of language, the main characteristic of the ASTP was emphasis on oral work. The students had to imitate native-speaking drill-masters, memorize dialogues, repeat them in chorus and individually, and answer questions.

---

[2] A phoneme is the phonological unit into which predictable variants of non-contrastive segments are grouped. For example in English [l] and [ɫ] belong to the phoneme /l/ (Celce Murcia et. al, 1996).

Pronunciation practice included minimal-pair drills[3], phonemic analysis, and transcription to dictation (Celce-Murcia, Brinton, and Goodwin, 1996).

Technology introduced new recording devices, so terms like "audiolingual" entered the lexicon. Audiolingualism had a great success, part of which can be attributed to the "Army Method". The Audiolingual Method provided the teachers with carefully prepared materials and well-developed teaching programs. The method was skill-oriented with a practical oral emphasis, aspects which made it gain great prestige from the beginning. Its popularity was supported not only by research scholars of the time, but also by teachers and the U.S. Army.

However, in 1957, the notion of transformational-generative grammar shifted the attention of the linguistic community to viewing language as rule-governed (Chomsky, 1957). This theory was a reaction to the basic belief that language learning was based on imitation and habit formation. Chomsky said that children were born with the ability to discover for themselves the underlying rules of a language system, and that children's acquisition of these grammatical rules was guided by principles of an innate Universal Grammar which could apply to all languages.

While some linguists embraced Chomsky's ideas, others continued using the Audiolingual Method. Alternative approaches were also derived giving birth to methods like the Total Physical Response (TPR), proposed by Asher. According to Asher, all structures could be taught in the imperative mood. In TPR, the students were asked to follow orders given by the teacher so that they could internalize these commands as whole or chunks rather than single lexical items (Asher, 1977). The method had a pure

---

[3] This is a type of drill where the students are exposed to words that differ in meaning. The words are transcribed exactly the same except for one sound segment which occurs in the same place either at the beginning of the word (bill and pill), in the middle (seat and sit), or at the end (bed and bet).

communicative focus with emphasis on comprehension. The idea was that the initial focus on listening without pressure to speak gave the students the opportunity to internalize the target sound system.

In the 1970s, Lozanov proposed a method called Suggestopedia which was used in Canada and the Soviet Union. This method was popular because learning was made pleasurable through a relaxed, non-threatening atmosphere which provided the assurance of progress in the new language, physical comfort, and aesthetic satisfaction. Music, carpeted floors, and comfortable chairs were introduced for relaxation facilitating the acquisition of a language (Lozanov, 1978). Emphasis was on communication and the learning of an extensive amount of words per day, but pronunciation was only taught when the teacher thought that some sounds were very difficult for the student to acquire. While the students reached good fluency and were able to communicate their thoughts, they still had difficulties with grammar and phonology.

A couple of years latter, Cattegno suggested another alternative approach to language teaching through what is now known as the Silent Way (Cattegno, 1972). This approach took a structural view to the organization of the language to be taught. Here, language was seen as a group of sounds arbitrarily related to specific meanings and organized into bits of meaningful units by grammar rules, placing a great emphasis on vocabulary, grammar, and pronunciation. Classes were learner-centered and the students had to be focused on the sound system of the language they were learning. The teacher's silence was said to be the best vehicle for learning, as the students could be concentrated on any task the teacher asked them to do. The teacher spoke as little as possible and used the following resources:

a) a sound-color chart containing all vowel and consonant sounds of the target language

b) a Fidel wall chart containing all possible spelling patterns for each sound in the target language

c) a wall chart, containing common words in the target language; and

d) sets of rods of different colors and sizes. These rods were used to demonstrate and build intonation patterns and to contrast pronunciation of morphological endings.

The fact that the method used the sound-color correspondence as a basis for teaching indicates that it placed special emphasis on pronunciation.

Around this time, a specialist in counseling, and professor of Psychology at Loyola University, created a method called Community Language Learning. This specialist saw the teacher as a counselor and the students as the clients. Learning took place in a community manner. Students were grouped in a circle all facing one another, with a tape recorder in the middle of the circle. Students were prompted to talk about topics they wanted to discuss. The teacher stood outside the circle immediately behind the student who asked the first question. The teacher then translated the question into the target language and had the student repeat that sentence until clear pronunciation was achieved. The utterance was then recorded on tape, and the activity continued until all the students had asked or answered a question (Curran, 1972). The teacher was like a "human computer." As it is described, the students had total control over the content of the lesson and the extent to which repetition took place. This method emphasized paying attention to the sensitivity of the students regarding their desire to communicate.

During these years, other ideas related to a more susceptible period for language acquisition appeared. In 1967, Lennenberg argued that the language acquisition device worked only when stimulated at the right time (Lightbown and Spada, 1999). This time is called the critical period, which represents a biologically specific period of life during which maximal conditions for language acquisition exist. The Critical Period Hypothesis constituted a point of discussion in many subsequent studies related to pronunciation (Munro, Flege, and MacKay, 1996; Singleton, 1995; Ioup, 1995; Flege, 1981; Krashen, 1973; Scovel, 1969).

Some researchers do not see enough empirical evidence to back up the idea of a critical period in language learning. Flege states that "neither physiological maturation nor neurological reorganization renders an adult incapable of speaking a foreign language without an accent" (Flege, 1981). The idea that adult second language learners cannot achieve native like levels of proficiency is very debatable, as there are adult second language learners who have proved quite the opposite with their performance (Lightbown and Spada, 1999). Whether or not the critical period hypothesis is supported by current research, the problems of acquiring the sounds of a second language present a great challenge to any theory of second language acquisition.

The Communicative Approach to language teaching started taking over around the mid to late 1970's (Widdowson, 1978). Language was seen as interaction and communication, and its structure was said to reflect its functional and communicative uses. The primary units of language were not only its grammatical and structural features, but categories of functional and communicative meaning as seen in discourse too. The types of activities related to the communicative approach were unlimited. Listening, speaking, reading, and writing were emphasized. Because the main objective of the

approach was communication, a tendency to teach pronunciation in language classrooms started to increase. This approach is currently dominant in language teaching.

Contrastive analysis emerged when researchers started discovering how the first language (L1) influenced the acquisition of the second language (L2). Points of similarities and differences between native and target languages were identified by systematically comparing two languages. The results of such studies convinced linguists that the differences between languages could be used to predict language problems in second language acquisition (SLA). This conviction soon started to decay when such predictions were challenged by a number of empirical tests. The results of these tests demonstrated that not all problems were predicted, and some of those anticipated failed to materialize (Wardhaugh, 1970).

The potential utility of contrasting the students' native and target languages posed interesting questions attributed to first language interference. Error analysis then became popular (Corder, 1974). This analysis was based on errors that occurred in the learner's interlanguage system. This was an excellent means of discovering the strategies learners adopted when learning a second language. Richards created three categories of errors in 1971, which include:

a) interlingual: errors caused by negative transfer from the learner's first language

b) intralingual: errors rooted in a marked feature of the target language and being made by L2 learners regardless of their L1

c) developmental: errors which would be of the same type a child would make when acquiring his first language

In 1972, Selinker introduced the notion of interlanguage. This is the language system the learners have been able to construct out of the linguistic input to which they

have been exposed that is, the system that the L2 learners are using at a particular period in their acquisition of a second language. Research then was focused on how the learners' interlanguage developed and on the patterns of sound acquisition across age and language groups. When interlanguage stops changing, it is said to have fossilized (Ioup and Weinberger, 1987).

As we have seen, interest on the teaching of pronunciation has increased with years. At the level of classroom action, efforts to teach pronunciation have resulted in practices using minimal pair exercises, grapheme-phoneme correspondence, rhythmic chants, and even exercises related to intonation and stress. Textbooks like *Well Said, Pronunciation Pairs, Pronunciation Plus,* and *Clear Speech*, among others, are very popular in our ESL classrooms nowadays to teach pronunciation.

Generally speaking, Spanish speaking learners regard English as a language difficult to understand and to pronounce, causing second language teachers (SLT) to look for ways to meet the demands English language learners have. Numerous research actions have been taken to address their concerns.

1.3 CURRENT TRENDS AND RESEARCH

Language teachers are primarily interested in the patterns of sounds that make up speech, that is phonology. They are also interested in phonetics, because if they analyze the sounds and describe them phonetically, they are in a better position to carry out research on how the individual sounds pattern in a given language. For example, given the fact that a minimal pair can be defined as a pair of words differing in lexical meaning based on a difference in sound, one can say that there is a phonemic difference between the pronunciation of the words *late-rate* or *sheep-ship*. However, this phonemic

distinction between the consonant sounds /l/ and /r/ or /i/ and /ɪ/ does not exist in all languages.

The fact that not all sounds in one language are necessarily present in another language has inspired a number of researchers to carry out studies where the perception and production of L2 sounds has been addressed. Some are interested in the ability second language learners have to discriminate contrastive consonant or vowel sounds, others have addressed the underlying factors leading to production, while others are interested in seeing the relationship between perception and production in children and adults.

1.3.1 Infant and adult perception studies

Categorical perception is a phenomenon in which labeling limits discrimination. This allows listeners to segment the words they hear according to the phonemic categories of their language and ignore unessential variations within a category (Pisoni, 1978).

According to a general universal theory, there should be a perceptual ability present at birth. This statement was corroborated by Eimas in 1975, and by Streeter in 1976. The results of Eima's studies revealed that infants do not need to learn about phonemic categories to be able to group speech stimuli, they can do this soon after birth. In addition, Streeter's studies concluded that infants were able to distinguish many sounds that were not used in their native language. However, other researchers have found out that in order to maintain the integrity of this ability, there should be specific types of early experience, and that the absence of this early experience may result in a partial or complete loss of the perceptual ability. The loss might be irreversible even when this experience is acquired at a later point in language development (Pisoni, Lively

and Logan, 1994). This last statement raised a very serious implication for second language teachers because, if the loss was in fact irreversible, then training students to identify phonetic distinctions was not possible. A series of attempts to use laboratory-training procedures turned out to be unsuccessful (Strange and Jenkins, 1978).

In 1989, Werker carried out a study in which she traced how speech perception changed during development. She tested groups at different ages (adults, twelve, eight, and four year olds, and eight months old), concluding that there was strong evidence that the decline in universal phonetic sensitivity was occurring during the second half of their first year. In her study, she argues that recovery of sensitivity is very difficult in adults, and that after several hundred trials, English speaking adults were unable to discriminate the most difficult Hindi contrasts.

In a study carried to see how adult Korean listeners perceived the English consonants (Schmidt, 1996), 100% of the listeners labeled English /l/ and /r/ as Korean /l/. The sound /l/ is included among the Korean consonants, but it never occurs in initial position. However, the sound /r/ is not included in the Korean language. The lack of the sound /r/ in Korean resulted in the incorrect discrimination of /l/ and /r/ in minimal pairs contrasting these two sounds in English.

In addition, performance of Korean students who were presented with a /r/ and /l/ discrimination task is not uniform. Several studies carried out with Korean and Japanese listeners on the perception of /l/ and /r/ have reported that performance is dependent on the phonetic context of the sound and that training is possible (García-Pérez, 1997; Logan, Lively and Pisoni, 1991; Sheldon and Strange, 1982). More recent studies have shown that when adults are trained to discriminate novel phonetic contrasts that are not distinctive in their native language, their performance identifying the correct sounds can

improve (Wang, 2002; Flege, Mackay and Meadow, 1999; Wang, 1997; Bradlow and Pisoni, 1997; Logan, Lively and Pisoni, 1991; Flege and Wang, 1989).

1.3.2 Speech Perception Models

In cross-language perception, two influential models explain how foreign vowels are assimilated to native phoneme categories: the Perceptual Assimilation Model (Best, 1994); and the Speech Learning Model (Flege, 1993, 1991a, 1988, 1987, and 1981).

The Perceptual Assimilation Model (PAM) and the Speech Learning Model (SLM) state that many L2 production errors have a perceptual origin. Table 1.1 summarizes the most important characteristics of these two models.

The PAM suggests that infants' self-perception of speech sounds plays an important role in establishing a link between the perception and production of novel sounds. If this is so, we may predict that adult learners could modify gestural representations until auditory feedback indicates to them that the L2 sound has been met with success. The PAM predicts that there will be maximum discrimination of two non-native sounds when each is assimilated to a different phoneme category. Moderate discrimination will occur when both sounds are assimilated to a single phonemic category, especially if one sound is a closer phonetic match to that category than the other. If two non-native sounds are equally good candidates for a single category, then the pair could be assimilated to a single native phoneme equally well or poorly.

Table 1-1. Characteristics of the PAM and SLM speech perception models

Model	Origin of error production	Predictions	Cross-language perception objects
PAM	perceptual	L2 vowels may be discriminated	gestures used to form sounds
SLM	perceptual	L2 vowels may be accurately produced	vowel and consonant segments

The SLM introduces two concepts: "new" and "similar". The concept "new" refers to the ability L2 learners have to create a "new" phonetic category once they perceive that a sound differs from a corresponding sound in the L1. The greater the differences between the L1 and L2 are, the more possibilities there will be for a creation of such 'new' phonetic category; this is to say, the more likely it is that learners perceive the differences between the sounds. The word "similar" refers to sounds that are slightly different from L1 phones. The L2 sounds are usually transcribed with the same symbol as that of the L1.

The SLM establishes a link between perception and production. It claims that once the learners have established a category for representing a novel sound, their production of that sound will be as good as that of a native speaker of the L2, provided their phonetic categories have been accurately represented.

1.3.3 Relationship between perception and production

Little empirical evidence exists on the relationship between perception and production. Advocates of the motor theory would say that speech perception is dependent on speech production. Furthermore, studies have proved that production can in fact precede perception in adults. (Goto, 1971; Sheldon and Strange, 1982; Liberman, et al., 1967).

Studies with children provide evidence that perception precedes production (Williams, 1979). Whether perception precedes production or production precedes perception may be determined by earlier development of speech perception in children.

So far, it seems that speech perception develops according to the language environment of the individual. But according to the following studies, the L2 learner's

perceptual system can be modified through linguistic experience and through training (Derwing, Munro, and Wiebe, 1998).

In 1989, Perlmutter carried out a study in which adult ESL learners were given language instruction with a special emphasis on pronunciation. The findings showed that the students' intelligibility improved. The findings of Perlmutter's study were corroborated in 1998, when Derwing, Munro, and Wiebe showed that long-term ESL individuals' intelligibility could improve significantly in a 12-week program emphasizing global production skills.

Other studies have assessed the effects of English language experience on non-native speakers' production and perception of English vowels (Flege, Bohn, and Jang, 1997; Ioup, 1995; Rochet, 1995). According to these studies, if adults are given sufficient native speaker input, they will be able to produce and perceive certain L2 vowels more accurately.

Most of the studies mentioned above have been carried out in sound treated rooms and the subjects have been tested in cubicles equipped with headphones and a workstation. In these studies, the stimuli have been recorded in a sound-attenuated booth, filtered and digitized with 16-bit resolution for presentation on the workstations.

The positive results in these studies led to the idea that similar procedures could be modified for use in L2 classrooms. In practice, the majority of our ESL institutions cannot provide the teachers or the students with digitized sounds and workstations to do this type of training, although it is a fact that larger colleges are provided with computer labs these days.

The present study is primarily aimed at finding the effects of training ESL Spanish speaking students to differentiate and produce contrasting novel vowel sounds in a regular ESL classroom condition.

1.4 CONTRASTING FEATURES BETWEEN SPANISH AND ENGLISH VOWEL SYSTEMS

When two languages are compared, there might be differences in the number of contrastive vowels they have, in the phonetic realization of the vowels shared by the two systems, and in contextually induced variation. Learners usually perceive (L2) vowels as instances of the closest (L1) vowel, and produce them accordingly. If teachers know what the perceptual relationship of L1 and L2 vowels is, they can predict how inexperienced students will produce L2 vowels (Flege, Bohn, and Jang, 1997).

In 1990, Odisho carried out a study where he concluded that teaching vowels to Hispanic students was far more difficult than teaching them the consonantal system of the English language. He pointed out three main differences:

a) The vowel systems of both languages are drastically different from the qualitative and the quantitative point of view,

b) The dynamics that controls the qualitative and quantitative variation of the vowels is diametrically opposed in both systems, and

c) There is a one-to-one grapheme-phoneme correspondence of vowels in Spanish, but a highly inconsistent grapheme-phoneme correspondence in English. For example, in Spanish the letter a tends to retain its phonemic value very consistently (gato, almohada, arriba, abajo, allá). In all these cases the pronunciation of the letter a is /a/. In English, however, the same letter a can

render a wide variety of phonemic values such as, /ə/ as in **above**; /æ/ as in **apple**; /ɑ/ as in **ball**; and /ɛ/ as in **any**.

Vowel quality is regarded as the acoustic impression of the vowel on the ear. It results from the differences in tongue and lip configurations and other movements required for the production of a vowel. The four basic parameters for vowel description in English are front-back, high-low, lip position (rounded-unrounded), and the added dimension tense-lax (Rogers, 1991).

According to these parameters, English[4] has 14 basic vowel sounds which are:

/i/	beat	high	front	unrounded	tense
/ɪ/	bit	high	front	unrounded	lax
/ɛ/	bet	mid	front	unrounded	lax
/e/	bait	mid	front	unrounded	tense
/æ/	hat	low	front	unrounded	lax
/ɑ/	bought	low	back	unrounded	tense
/o/	boat	mid	back	rounded	tense
/ʊ/	pull	high	back	rounded	lax
/u/	pool	high	back	rounded	tense
/ʌ/	cup	mid	central	unrounded	lax
/aj/	bite	low	front	unrounded	tense
/aw/	cow	low	central	rounded	tense

---

[4] In our study, we refer to Canadian English only, as it was the variety of English used in the study.

/ɔj/	toy	mid	back	rounded	tense
/ə/	banana	mid	central	unrounded	lax

Spanish, regardless of dialect, has five pure vowels /i, e, a, o, u/ as in piso, peso, paso, pozo, and puso. Spanish vowels may be classified according to following three parameters which are front-back, high-low, and lip position (rounded-unrounded). Notice that the Spanish vowel system does not have the tense-lax distinction. According to this, the Spanish vowels are classified as follows:

/a/	ya	low	central	unrounded
/e/	té	mid	front	unrounded
/i/	sí	high	front	unrounded
/o/	yo	mid	back	rounded
/u/	tú	high	back	rounded

Vowel quantity, length, or duration is another feature that differentiates Spanish and English vowel sounds. For example, the vowels /i/-/ɪ/ are apparently not distinguished from one another by duration differences by Spanish speakers. Auditory evaluation reports that Spanish /i/ is closer to English /i/ than /ɪ/ (Stockwell and Bowen, 1965). Moreover, Flege conducted a study in which Spanish monolinguals were asked to use the leters ⟨a, e, i, o, u⟩ to label the vowels in naturally produced English words. Spanish speakers classified the realizations of English /i/ and /ɪ/ as instances of Spanish /i/ (Flege, 1991b).

Three pairs of English vowels will be used in our study: /i/-/ɪ/, /u/-/ʊ/, and /ɑ/-/ʌ/. The pair /i/-/ɪ/ poses interesting differences for Spanish speakers. The English vowel /i/ is

very similar to the Spanish vowel /i/. The English sound is a bit higher than the Spanish sound, and the English /i/ is a bit longer than the Spanish /i/. For example, in order to pronounce the English word *see* correctly, a Spanish student is asked to think of this vowel as an equivalent to two Spanish /i/ sounds, as the emphatic pronunciation of the Spanish sí /sii/. The English vowel /ɪ/ is also high and front but is a much more relaxed sound than the vowel /i/. For Spanish speakers this would be an intermediate vowel, lower than his /i/ and higher than his /e/. This intermediate quality makes it very difficult for them to distinguish and reproduce.

The pair of English vowels /u/-/ʊ/ are two high vowels which the Spanish speakers relate to their single high back vowel /u/. The English sound /u/ is less tense than the Spanish /u/ but similar to it in most respects. The English vowel sound /ʊ/ is the one giving the most trouble to Spanish speakers. It is a more relaxed vowel, so no much muscular effort is needed to pronounce it. Because this sound stands between Spanish /u/ and /o/, most Spanish speaking learners will pronounce the high, tense /u/.

The pair of English vowels /ɑ/-/ʌ/ are new sounds for the Spanish speakers. The sound /ɑ/ is usually substituted by the Spanish /o/, a mid-back vowel. Because of the frequent spelling of /ɑ/ with two letters —*ou* as in *thought, au* as in *taught, aw* as in *saw*— many students tend to turn this sound into a diphthong by adding a /u/ sound. So, ESL teachers usually give a thorough oral practice before the Spanish speaking students are allowed to see these words in print. Once they see it, emphasis is made on the correspondence between the sound and spelling.

1.5 RESEARCH QUESTIONS

Spanish speakers learning English find it difficult to perceive the difference between English vowels in the contrasts presented in the previous section. Once we became aware of these contrastive features, and of the results of previous studies related to perception and production of vowels, we were in a position to define the specific problem of our investigation. This research had a distinct characteristic from the rest of the studies carried out so far. Spanish speakers received training on the pairs of contrasting Canadian English vowels /i/-/ɪ/, /u/-/ʊ/, and /ɑ/-/ʌ/ in a regular ESL classroom setting. The effects of this training could corroborate or refute previous theories in speech perception and production. The following are the research questions addressed in this thesis:

a) Would there be a significant improvement in the performance of ESL Spanish speaking students when they are trained to differentiate and produce contrasting vowel sounds in a regular ESL classroom setting?

b) Would there be a correlation between the performance in the perception of the contrasting pairs of vowels and the performance in their production?

c) Will the students show an improved performance in the perception and production of a specific pair of vowels as compared with the other two pairs?

d) Will the comprehensibility test show an improvement in the global production of these vowels?

Based on the results of previous research we predict that, although the conditions in which the study takes place differ, and given the fact that training is carried out in a regular ESL classroom setting, there could be a significant improvement in the perception of novel contrasting sounds.

According to the SLM, the students should show an improved performance in the perception and production of the specific pair of vowels /ɑ/-/ʌ/, as they are totally new sounds for Spanish speakers. We might not see a correlation between the students' performance in the perception of the contrasting pairs of vowels and the performance in their production. We do, however, think that individual students might show improvement in pronunciation as an effect of training.

CHAPTER II

METHOD

2.1 INTRODUCTION

This chapter describes in detail how the study was conducted. At the beginning, there is a specific identification of all the research participants.

After identifying the participants, the chapter describes the materials and the reasons for which they were used. Specific information regarding the stimuli preparation is included, and brief information about the computer programs used is also given. The chapter concludes with an explanation on each of the steps taken to carry out the research.

2.2 PARTICIPANTS

Before carrying out the experiment, a request for ethical approval of research was completed. After the request was approved by the Simon Fraser Research Ethics Board (Appendix A), a group of students from Van West College[5] was recruited with the permission from the school administrator. This was done using posters (Appendix B) and word of mouth. During the first meeting, the group was informed that the purpose of the research was to gain a better understanding of how second language learners learn to

---

[5] This international college, situated at 200-1215 West Broadway, Vancouver, offers a wide range of ESL programs. It is registered with the Private Post-Secondary Education Commission of the Province of British Columbia.

perceive and produce speech sounds in a classroom setting. They were also told that the results of the proposed experiment would lead to conclusions about theoretical models of second language acquisition, and that we hoped our findings would have practical applications in the field of second language teaching. The group was also told that they were going to be asked to complete a language background questionnaire (Appendix C), to participate in a 12 hour training period, and to take a test before and after the training period on the perception and production of English vowels.

Following this information session, the subjects who wanted to participate signed a consent form (Appendix D). Initially, 51 students signed this form. From these 51 students, only 42 completed the language background questionnaire, and because of schedule conflicts, family problems in their homelands, and work related problems, ten students withdrew from the study.

2.2.1 Subjects

Thirty-two native speakers of Spanish (18 females and 14 males) at the intermediate proficiency level in a full-time ESL program participated in the whole study. They ranged in age from 18 to 32 with a mean age of 22. There were 21 from Colombia, 9 from Mexico, 1 from Argentina, and 1 from Guatemala. Their length of residence in Canada ranged from 1.5 weeks to 19 months at the time of initial testing (only one student had come to Canada a month prior to the time the students took the first test), with a mean length of residence of 3 months. All subjects reported normal hearing. They started studying English at a mean age of 15. On a scale from 1 to 7, where 1 was *never* and 7 was *very often*, the 32 subjects roughly estimated that 62% of their time, they speak

English to a native speaker of English, and 49% of the time to a non-native speaker of English[6]. These results are summarized in Table 2.1.

Table 2-1 Language background information for the experimental and control groups

GROUP	COUNTRY				Gender		Age	AOA	LOR	AOL	% USE NS	% USE NNS
	C	M	A	G	F	M						
Exp.	11	3	1	1	12	4	23	22	3,3	16	60.0%	53.4%
Control	10	6	-	-	6	10	21	21	2.6	14	64.0%	44.0%
Total	21	9	1	1	18	14	22	21.5	3.0	15	62.0%	49.0%

COUNTRY: C= Colombia, M= Mexico, A= Argentina, G= Guatemala
AOA= Age of arrival in Canada
LOR= Length of residence in Canada (months)
AOL= Age when subjects started studying English in home country
% Use NS= Subject's estimated % of English language use with a native speaker
% Use NNS= Subject's estimated % of English language use with a non native speaker

After the introductory meeting, all the subjects were very interested in the training sessions, but they had to be sorted into two groups: the experimental group, and the control group. In order to be fair, the 16 members of the experimental group were selected through a random draw.

The participants in the experimental group ranged in age from 18 to 32 with a mean age of 23. Eleven were from Colombia, 3 from Mexico, 1 from Argentina, and 1 from Guatemala. Their length of residence in Canada was 2 weeks to 19 months, with a mean length of residence in Canada of 3.3 months at the time of initial testing. The subjects in the test group started to study English at a mean age of 16. On a scale from 1

---

[6] On this 7-point scale, number 7 was equivalent to 100%. Because the students estimated how often they spoke English to a NS or to a NNS, sometimes they circled number 7 in both questions (the sum would then be 200%). This is why the percentages representing these questions are so high.

25

to 7, where 1 was *never* and 7 was *very often*, the subjects estimated that 60% of the time, they speak English to a native speaker (NS) of English, and 53,4% of the time to a non-native speaker (NNS) of English. Detailed results are provided in Table 2-2.

The students in the control group ranged in age from 18 to 28 with a mean age of 21. Ten were from Colombia and 6 from Mexico. Their length of residence in Canada was 1.5 week to 7 months, with a mean length of residence in Canada of 2.6 months at the time of initial testing. These students started to study English at a mean age of 14. On a scale from 1 to 7, where 1 was *never* and 7 was *very often*, the subjects estimated that 64% of the time, they speak English to a NS, and 44% of the time to a NNS. Detailed results are provided in Table 2-3.

2.2.2 Teacher

Before the course started, the school administrator from Van West College and one teacher were approached and asked if they were willing to participate in the study. Once the nature of the research was explained, the school and a particular instructor showed an immense interest in the project. This teacher was a male NS who was born and raised in Vancouver.

The content and the procedures of the course were negotiated, and it was agreed that the researcher would introduce each unit, as the selected instructor was not very comfortable with the linguistics terminology. There were weekly meetings with the instructor to ensure that everything was going as planned. He was very enthusiastic about everything and said that the students were enjoying the course.

2.2.3 Speakers

A total of five Canadian English NSs, two males and three females, were recruited from the Department of Linguistics at Simon Fraser University. All had grown

Table 2-2. Language background information for the experimental group

Code	Gender	Country	Age	AOA	LOR	AOL	% use NS	% use NNS
E-2	M	Colombia	30	29	19	29	66.0%	66.0%
E-15	F	Guatemala	18	18	3.5	8	100%	33.0%
E-17	F	Colombia	18	18	2	13	33.0%	49.5%
E18	M	Argentina	26	26	0.5	26	50.0%	50.0%
E-14	F	Colombia	18	18	3	10	66.0%	66.0%
E-20	F	Colombia	18	18	0.5	14	66.0%	49.5%
E-24	F	Mexico	25	25	1.5	6	49.5%	33.0%
E-25	F	Colombia	32	31	1.5	28	66.0%	33.0%
E-26	F	Colombia	27	26	1.5	25	16.5%	16.5%
E-33	F	Colombia	19	19	0.5	9	49.5%	82.5%
E-36	F	Mexico	21	21	7.5	18	33.0%	100%
E-37	F	Colombia	29	29	6	14	66.0%	49.5%
E-39	F	Colombia	24	24	1.5	14	100%	49.5%
E-40	F	Mexico	22	21	2	16	49.5%	100
E-45	M	Colombia	19	18	2	14	100%	49.5%
E-46	M	Colombia	18	18	1	11	49.5%	33.0%
Mean			**23**	**22**	**3.3**	**16**	**60.0%**	**53.4%**

AOA= Age of arrival in Canada
LOR= Length of residence in Canada (months)
AOL= Age when subjects started studying English in home country
% Use NS= Subject's estimated % of English language use with a native speaker
% Use NNS= Subject's estimated % of English language use with a non native speaker

27

Table 2-3. Language background information for the control group

Code	Gender	Country	Age	AOA	LOR	AOL	% use NS	% use NNS
C4	M	Mexico	18	18	0.3	6	100%	82.5%
C6	M	Colombia	25	25	2.5	14	49.5%	49.5%
C7	M	Colombia	25	25	2	25	49.5%	33.0%
C9	F	Colombia	18	18	2	14	49.5%	33.0%
C10	F	Colombia	18	18	2.5	12	33.0%	49.5%
C11	M	Colombia	24	24	3.5	19	66.0%	82.5%
C12	M	Colombia	18	18	3.5	12	100%	66.0%
C19	F	Colombia	28	27	7	14	66.0%	66.0%
C22	M	Mexico	25	24	2	13	100%	49.5%
C23	M	Colombia	18	18	2	14	16.5%	49.5%
C30	M	Mexico	23	23	1	21	16.5%	16.5%
C41	F	Mexico	18	18	2	8	16.5%	16.5%
C42	F	Colombia	23	23	1	22	82.5%	49.5%
C44	M	Colombia	20	19	2	14	100%	33.0%
C48	F	Mexico	18	18	5	6	100%	33.0%
C50	M	Mexico	23	22	3	14	82.5%	33.0%
Mean			**21**	**21**	**2.6**	**14**	**64%**	**44%**

AOA= Age of arrival in Canada
LOR= Length of residence in Canada (months)
AOL= Age when subjects started studying English in home country
% Use NS= Subject's estimated % of English language use with a native speaker
% Use NNS= Subject's estimated % of English language use with a non native speaker

up in Canada west of Quebec, and all had taught English as a second language (ESL).

One of these speakers, a female NS, provided the recordings for the stimuli used in the pre-post perception test. The other four NSs (two males and two females) provided recordings that were used as a check for the rating scales included in the production tests.

2.2.4. Listener-Raters

A group of seven listener-raters Canadian English NSs (one male and six females) were recruited from the Department of Linguistics at Simon Fraser University. Just as the speakers, these listener-raters had grown up in Canada west of Quebec. Three of them were undergraduate students, three graduate, and one phonetically trained professor. The seven of them reported normal hearing. Five of these listener-raters had experience teaching ESL, two did not.

2.3 MATERIALS

2.3.1 Written Materials

A language background questionnaire was used to obtain language background information from all the participants. Although the native speakers did not have to fill out all the information, the data they provided at the beginning of the questionnaire was crucial for the experiment (normal hearing, born and raised in Canada, etc.)

A pretest-posttest evaluation procedure was used in the experiment. The same test was used before and after training to evaluate changes in the subjects' performance attributable to training. The first part of the test consisted of a forced-choice identification task[7]. The participants were given a list, with 24 different words (twelve minimal pairs)

---

[7] Identification tasks have been used to investigate cross-language phenomena in short-and long-term training studies (Pisoni et. al., 1982). They have proved to be an effective means of training listeners to perceive non-native phonetic sounds.

which were grouped by vowel sound (Appendix E-1). The words were organized in such a way that each vowel contrast was represented in a separate group.

a) First group: /i/-/ɪ/ contrast

b) Second group: /ɑ/-/ʌ/ contrast

c) Third group: /u/-/ʊ/ contrast

d) Fourth group: /ɑ/-/ʌ/ contrast

e) Fifth group: /u/-/ʊ/ contrast

f) Sixth group: /i/-/ɪ/ contrast

The first and sixth groups were represented by the /i/-/ɪ/ contrast, and constituted a block which was referred to as "front". The second and fourth groups were represented by the /ɑ/-/ʌ/ contrast, and formed another block. The term "mid" was used for convenience because we know that /ɑ/ is a back vowel; and the third and fifth groups included /u/-/ʊ/ contrast and composed the last block, which we called "back".

For example, minimal pair *cheap-chip* was in the first and sixth groups. In the first group the word recorded was *cheap*, and in the sixth group the word recorded was *chip*. The students' task was to identify minimal pairs with the target vowel contrasts.

The pairs of contrasting vowel sounds used in our study were selected based on two different criteria. The first is related to the SLM which labels sounds as "new" and "similar". "New" sounds in our study would be the contrasting pair /ɑ/-/ʌ/, and "similar" sounds would be represented by the /i/-/ɪ/ and /u/-/ʊ/ contrasts. We predicted that performance in general would be better if the sound was "new" than if it was similar.

The second criterion refers to functional load. This is based on the number of pairs of words in the English lexicon that are distinguished by the contrast (Brown, 1988). For example, the contrast between English vowel /i/ in sheep, and /ɪ/ in ship serves to distinguish many pairs of words, for example peak-pick, peel-pill, etc.

In 1987, Catford argued that the opposition /i/-/ɪ/ has a high functional load (95%) and that the opposition /u/-/ʊ/ has a low functional load (7%). Our study selected one pair with high functional load /i/-/ɪ/, one with middle functional load /ɑ/-/ʌ/, as in cut-caught (65%); and one with low functional load /u/-/ʊ/.

The second part of the test (Appendix E-2) consisted of fifteen sentences, which the students had to read, each containing a single clause with high frequency lexical items (e.g. "The pool is too cool" or "She seems to be extremely pleased"). Right afterwards they read a paragraph with sentences that had the same characteristics. At the end they orally answered a question specifically asking them if they thought they would speak a lot of English when they went back to their countries. This provided more information related to the students' personal motivation and served as an extemporaneous speech sample.

The teacher in the experimental group used *Training Spanish Speakers in the Perception and Production of English Vowels* (García-Pérez, 1999), a booklet specifically designed for this project (Appendix F). The student's progress was evaluated through a quiz at the end of each unit. No recordings were used in the training.

2.3.2 Speech Materials

Stimuli used for the forced-choice identification task in the pretest and posttest were recorded by a Canadian English female speaker in a quiet room, in a home

environment, with a SONY tape recorder Model TCM-454 VK and a SONY microphone Model DR-30PCAMP. The speaker was given a printed paper with a list of 24 different words, and she was asked to read each word out loud.

After the stimulus list was recorded, it was tested with two Canadian English listeners (one male and one female). They were simply asked to circle the word they heard, and they did the task with 100% accuracy.

The second set of speech samples was collected from all ESL subjects who participated at two different periods of time. Before the training, all the subjects recorded the fifteen sentences, the paragraph and the answer to a question on a tape recorder (the same tape recorder and microphone were used in the whole study). Once the training was over, the recording procedure was done exactly the same as the first time.

As mentioned before, recordings of four Canadian English NSs were used as a check on individual listeners' use of the rating scales for the comprehensibility and category goodness test. The expectations were that the listener-raters would agree in assigning good scores to the NSs. Failure to do so might indicate either a misunderstanding of the instructions or a poor design of the rating task (for example, too difficult).

2.4 PROCEDURE

2.4.1 Language Background Questionnaire

Forty-two students completed the language background questionnaire. It was given to the students in the first session where the general objectives of the study were laid out. This session was conducted in Spanish to avoid misunderstandings. The students were asked to read the questionnaire and feel free to ask any questions. Some students did

not understand a few of the questions and the Spanish translation was immediately provided.

2.4.2 Selection of Experimental and Control Groups

As explained before, subjects were randomly assigned to two separate groups (the control or comparison group, and the experimental or treatment group). This was done in order to investigate the effect of the treatment administered to the experimental, but withheld from the control group.

2.4.3 Pretest

For the perception part, participants in the control and experimental groups were presented with the recorded stimuli just once. The students were given printed paper with clear instructions. They were presented with minimal pairs and their task was to identify the word they heard by circling it.

On a different day, the students recorded the sentences. They were given the opportunity to go through the sentences and ask questions about unfamiliar words. Then the students read the 15 sentences out loud onto a tape. The same procedure was followed with the paragraph.

2.4.4 Training

The objective of an experiment is to compare the effect of one condition on one group with the effect of a different condition on a second group. Hence, the manipulation of independent variables began. The term manipulation is used in the sense that one group received instruction and the other one did not.

Classes started in the experimental group, and feedback was received on a daily basis from the instructor. The control group received no specific pronunciation instruction, while the experimental group attended 40 minutes of pronunciation classes,

33

three times a week for three weeks. The exercises, which included articulatory awareness, listening practice (discrimination tasks), oral repetition, reading out loud, dictation, spelling awareness, and conversation practice, varied among sessions. In all cases, the students were presented with very similar tasks to the ones included in the pretest and the posttest. However, they did not practice the exact material presented in the test. All the students in the control and experimental groups attended ESL classes, 20 hours per week, and their regular program included emphasis on pronunciation.

2.4.5 Posttest

After a three-week training period, both groups were tested again with the same measurement used in the pretest. The same procedure followed in the pretest was followed in the posttest.

2.4.6 Stimuli Preparation

Three sets of stimuli were elicited from three different sources. The first source was elicited from a Canadian female speaker. These were the stimuli presented to the participants for the forced-choice identification task used in the first part of both pre-and posttests.

The second set of stimuli consisted of recordings elicited from each participant before and after training. This took place in a regular ESL classroom. The room had one station with a tape recorder and a microphone. Individual students were provided with written material consisting of:

a) fifteen sentences

b) a paragraph

c) a question

34

They were given the opportunity to go over the sentences and the paragraph and ask questions about unfamiliar words. Then, they read the sentences and the paragraph out loud onto a tape. They also recorded the answer to the question. The entire task took two to four minutes for each participant.

The third set of stimuli was drawn from four Canadian NSs. The native speakers were given the same written material provided to the students and they were asked to do exactly the same task under the same conditions, using the same recording equipment. The objective of doing recordings was very simple. These stimuli were going to be added to the stimuli provided by the students. Once both stimuli were randomized, it was expected that the raters would give very high (very good) rating scores to the native speakers. The reliability of the scores could then be determined.

As there were 32 participants, and each provided us with recordings before and after training, we ended up having 64 speech samples (32 participants x 2 times = 64) plus 4 samples from the NSs, which made a total of 68 speech samples. These were digitally re-recorded on an IBM computer using *Gold Wave* shareware version 4.6. The files were saved on a CD as wave files at 22 kHz with 16-bit resolution.

We selected 12 words from the 15 sentences which were individually digitized using the editing feature of the *Gold Wave* program. The words representing the contrasting vowels were:

1 pull-pool   /ʊ/-/u/

2 took-too   /ʊ/-/u/

3 caught-cut   /ɑ/-/ʌ/

4 boss-bus   /ɑ/-/ʌ/

5	seems-Tim	/i/-/ɪ/

6	weeks-mix	/i/-/ɪ/

As a result, the final set of individual words included in the sentences consisted of 768 speech samples (32 participants x 2 times x 12 words = 768) elicited from the subjects, and 48 speech samples (4 NSs x 12 words = 48) elicited from the NSs, for a total of 816 (768 + 48) speech samples. These samples would serve as the stimuli for the category goodness test.

We mentioned before that in the pretest and the posttest the students had to read a paragraph onto a tape. We needed these recordings to see if there was an improvement in the global production of these vowels. Three sentences from this paragraph with lexical items containing the vowel sounds included in the study were also digitized with the same procedure used for the words. The sentences are:

a) She is always sick.

b) Luke is strong.

a) I think he is a fool.

As a result, these final stimuli consisted of 192 speech samples (32 participants x 3 sentences = 192) elicited from the subjects, and 12 speech samples (4 NSs x 3 sentences = 12) elicited from the NSs, for a total of 204 (192 + 12) speech samples. These samples served as the stimuli for the comprehensibility test.

Once the samples were all digitized and recorded onto a CD, they were processed in the Phonetics lab in the linguistics department. Using Sound Edit™ 16 V.2, we translated the audio file format (from wave to sound files). This was done to be able to use another program that would randomize the data for the listening task.

After the files were translated, we used the Audio Experiment Generator 4.0 (Munro, 1999) to set up the listening task for each of the 12 words and three sentences. The program allowed us to select among other aspects, the number of times the data could be randomized, the number of stimuli, and the point scale we wanted to use. For both, the comprehensibility test and the category goodness test, we used a 9-point scale. The ratings "1" (very poor) to "9" (very good) were assigned to the category goodness test, and the ratings "1" (poor) to "9" (native like) were assigned to the comprehensibility test.

The Audio Stimulus Playback 4 (Munro, 1999) was then used to play each stimulus. The program allowed the raters to select the audio file they wanted to hear (the audio files were named after the word representing the vowel, for example: mix, weeks, boss, bus, etc.). The raters had to write their names, and then provide the program with an ID (the ID was assigned by the person conducting the experiment). The program had 50 possible IDs, which meant that there were 50 different ways of randomizing the data. After saving the name and ID, the program was ready to start.

2.4.7 Rating Task

The raters completed the experiment in several listening sessions held over a period of one month. In order to reduce the possibility of fatigue, we asked the raters to carry out the task on different days. They were provided with professional headphones and were given a very short training prior to the beginning of the actual task.

The raters controlled the time they spent rating the samples. They just had to click the number representing their rating (Figure 2-1), and the computer recorded their judgements in a data file. Once they were ready, they rated the stimuli by clicking a number from 1 to 9, and once they did so, the computer provided them with the next

37

stimulus. For example, the raters heard the word *boss* 68 times. Each time, they had to rate how well the word was pronounced.

Figure 2-1. Rater's task for the category goodness test.

| very poor | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | | very good |

Click to continue
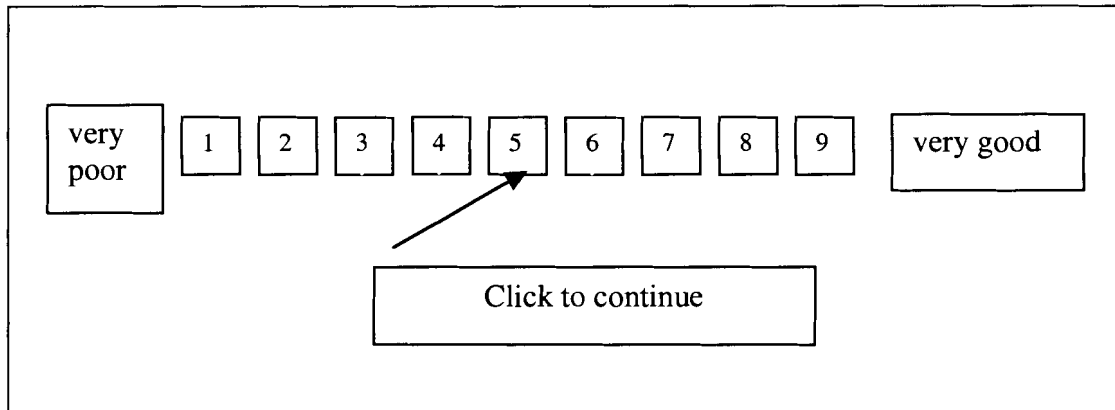
The same procedure was followed with the sentences. Once the raters finished their task, the results were obtained from the data file and copied onto an Excel file to be analyzed. These results are presented in Chapters III and IV.

CHAPTER III

RESULTS AND ANALYSIS ON PERCEPTION

3.1 INTRODUCTION

This chapter presents the results related to perception. It starts with an explanation of the key terms used in the statistical analysis. It continues laying out the results of the tests given on perception of all vowels, followed by the results on individual pairs of vowels. The end of the chapter presents the data drawn from the factorial analysis and concludes with an analysis of the possible reasons for the findings.

3.2 DEFINITION OF TERMS

3.2.1 Related to the experimental study

Among the different types of research that can be carried out to obtain data we have the experimental research. An **experimental study** assigns subjects randomly to two or more separate groups (the control or comparison group, and the experimental or treatment group) to investigate the effect of some treatment which is administered to the experimental group but withheld from the control group.

Statistical equivalence of subjects (usually achieved by random assignment of subjects in the different groups) should be achieved because the internal validity threat of selection should not be a factor in interpreting the results. That is to say, the groups being

compared should be equivalent so that any differences in their performance is not attributed to differences between the groups.

An experimental design needs to have at least two comparable groups or **conditions**. The objective of an experiment is to compare the effect of one condition on one group (in this study, this is the effect of pronunciation training in the experimental group) with the effect of a different condition on a second group (no specific pronunciation training in the control group).

When the researcher controls the group of subjects that will receive a particular treatment or condition, then the term **manipulation of independent variables** is used. An independent variable is a variable that researchers hypothesize occurred before and had an effect on another variable called the dependent variable.

3.2.2 Related to inferential statistics

Inferential statistics is used to make probability statements about the results. It is important because the researcher should be able to generalize the findings to similar groups or to the population of subjects. This type of statistics allows us to make such types of generalizations.

Probability in a scientific way means the degree of confidence we have in predicting something. If we refer to probability in terms of measurement error, the statement used is the **null hypothesis**. This hypothesis presumes that the population means are the same. That is to say, in this study this hypothesis would state that there are no differences between the means of the experimental and control groups. Although the results might clearly say that in fact there are differences, inferential statistics uses a test to prove that the null hypothesis is false, and then there is a high probability of stating that there is an existing difference between the groups.

The phrase **level of significance** or probability (p level) indicates what the chances are that we are wrong in rejecting the null hypothesis. This p level is expressed as a decimal and basically tells us how often we would expect no real differences, even if we rejected the null hypothesis. The lower the p level is, the more confidence we have to reject the null hypothesis. So, a level of $p < 0.05$ is usually considered better than $p < 0.1$ (Gall, et. al., 1999).

Factorial analysis of variance is used when we analyze two or more independent variables together. This study has three independent variables: time, vowel, and group. To compare these three variables, we used **ANOVA** (analysis of variance).

In ANOVA the statistical formula uses the variances of the groups to calculate a value that reflects the degree of differences in the means. This value is reported as F **ratio** or F **value**. Each **ANOVA** reports three F ratios. In our analysis, the first F ratio refers to group, the second refers to time, and the last one refers to the interaction between group and time. These values reported by ANOVA are of great importance because they tell us the possibility of generalizing the difference between two or more sample means to the populations from which the samples were drawn.

3.3 RESULTS

3.3.1 Test results on perception of all pairs of vowel contrasts /i/-/ɪ/, /u/-/ʊ/, and /ɑ/-/ʌ/.

Overall, the students in the experimental and control groups showed very similar results on the pretest (mean percentage of correct responses for the experimental group 60.1%, and 59.9% for the students in the control group). It shows that they just differed by 0.2% before training and indicates that the two groups were very similar before training.

41

From the administration of the pretest (Time 1) to the posttest (Time 2) all students in the experimental group showed an improvement (mean percentage of correct responses 83.3%) in the perception of vowel contrasts /i/-/ɪ/, /u/-/ʊ/, and /ɑ/-/ʌ/. The mean difference between Time 1 and Time 2 for the experimental group was 23.4%. Only 6 students in the control group improved slightly, 3 remained the same, and 7 decreased. Overall the control group decreased in its mean percentage from Time 1 (59.9%) to Time 2 (55,7%) with a mean difference of –4.2%. Table 3-1 summarizes these results. The individual analysis of the results of the different pairs of vowels might explain the causes of this decrease in the control group.

3.3.2. Results on individual pairs of vowel contrasts

3.3.2.1 Results on vowel contrast /i/-/ɪ/

The contrasting pairs /i/-/ɪ/ showed an improvement in the experimental group of 23.4%. That is, the experimental group scored 60.9% of correct responses at Time 1, and 84.4% at Time 2 (Figure 3-1).

Individually, only one student from the experimental group decreased her score from Time 1 (75%) to Time 2 (50%). No specific causes were attributed to this as this student attended all training sessions. Three students remained the same, and 12 students improved. In contrast, 8 students in the control group showed a decrease in their percentages of ID scores, 4 remained the same, and only 3 out of 16 improved slightly.

The experimental and control groups had very close percentage scores for the /i/-/ɪ/ contrast at Time 1 (experimental: 60.9%, control: 57.8%). The difference between the two groups was of 3.1%, and this was considered appropriate. However, there was actually a decrease (Time 2) in percentages of correct responses for the /i/-/ɪ/ sounds in

Table 3-1 Total % ID scores for vowel contrasts /i/-/ɪ/, /u/-/ʊ/, and /ɑ/-/ʌ/ in the pre-post perception tests.

Group	Code	Pretest	Posttest	Difference
E1	2	75,0%	87,5%	12,5%
E2	15	54,1%	70,8%	16,7%
E3	17	54,1%	79,1%	25,0%
E4	18	54,1%	70,8%	16,7%
E5	14	66,6%	79,1%	12,5%
E6	20	58,3%	70,8%	12,5%
E7	24	58,3%	95,8%	37,5%
E8	25	50,0%	95,8%	45,8%
E9	26	54,1%	70,8%	16,7%
E10	33	58,3%	87,5%	29,2%
E11	36	50,0%	75,0%	25,0%
E12	37	66,6%	87,5%	20,9%
E13	39	66,6%	95,8%	29,2%
E14	40	58,3%	91,6%	33,3%
E15	45	54,1%	83,3%	29,2%
E16	46	83,3%	91,6%	8,3%
C1	6	58,3%	58,3%	0,0%
C2	4	58,3%	50,0%	-8,3%
C3	7	62,5%	37,5%	-25,0%
C4	9	66,6%	70,8%	4,2%
C5	10	62,5%	62,5%	0,0%
C6	11	66,6%	70,8%	4,2%
C7	12	37,5%	58,3%	20,8%
C8	19	79,1%	54,1%	-25,0%
C9	23	50,0%	45,8%	-4,2%
C10	22	66,6%	50,0%	-16,6%
C11	30	62,5%	75,0%	12,5%
C12	41	83,3%	45,8%	-37,5%
C13	42	58,3%	66,6%	8,3%
C14	44	54,1%	58,3%	4,2%
C15	48	54,1%	50,0%	-4,1%
C16	50	37,5%	37,5%	0,0%
Mean E		60,1%	83,3%	23,2%
Mean C		59,9%	55,7%	-4,2%

Figure 3-1. Mean % of correct responses in the experimental and control groups for the contrasting pair of vowels /i/ and /ɪ/.
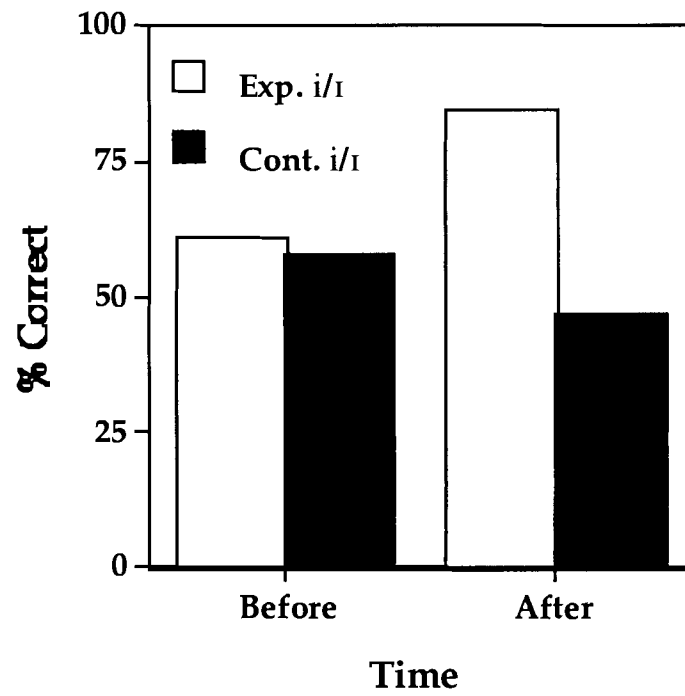
the control group. The percentage reached was of 46.88% with a difference between Time 1 and Time 2 of –10.9%. Table 3-2 summarizes the results.

So far, it appeared that training had had an effect on the perception of the perception of the /i/-/ɪ/ vowel contrast. This was very encouraging, because according to Flege's Speech Learning Model perceptual categories could be established if listeners could perceive the differences between the target vowel contrasts, and we really tried our best to make these differences clear when we introduced each training lesson. It was now the time to find out if we had been as successful with the other two pairs of vowel contrasts.

3.3.2.2 Results on vowel contrast /u/-/ʊ/

The /u/-/ʊ/ vowel contrast revealed a higher improvement than the /i/-/ɪ/ pair of vowel sounds. The percentage of ID scores for the experimental group at Time 1 was 47.7% and 78.9% at Time 2, with a mean difference of 31.3 %. (Figure 3-2).

From Time 1 to Time 2 15 students out of 16 increased their percentages of ID scores in the experimental group. Only one student from this group decreased her percentage of ID scores from Time 1 (87.5%) to Time 2 (75.0%) with a mean difference of –12.5%. However she kept a high percentage of ID scores in both tests.
This student was a different one from the one exhibiting the decrease in the /i/-/ɪ/ vowel contrast.

The results of the experimental group could be attributable to training. The experimental and control groups had very similar percentage scores for the /u/-/ʊ/ contrast at Time 1 (experimental: 47.7%, control: 46.1%) with a mean difference between the two groups of 1.6%. This was considered a minimal difference.

45

Table 3.2 Total % ID scores for the vowel minimal pair /i/-/ɪ/ in the pre-post perception tests.

Group	Code	Pretest	Posttest	Difference
E1	2	62,5%	75,0%	12,5%
E2	15	25,0%	62,5%	37,5%
E3	17	62,5%	75,0%	12,5%
E4	18	62,5%	62,5%	0,0%
E5	14	75,0%	50,0%	-25,0%
E6	20	50,0%	75,0%	25,0%
E7	24	87,5%	87,5%	0,0%
E8	25	50,0%	100,0%	50,0%
E9	26	75,0%	100,0%	25,0%
E10	33	62,5%	100,0%	37,5%
E11	36	50,0%	87,5%	37,5%
E12	37	37,5%	100,0%	62,5%
E13	39	62,5%	100,0%	37,5%
E14	40	75,0%	100,0%	25,0%
E15	45	37,5%	75,0%	37,5%
E16	46	100,0%	100,0%	0,0%
C1	6	50,0%	50,0%	0,0%
C2	4	62,5%	37,5%	-25,0%
C3	7	75,0%	25,0%	-50,0%
C4	9	100,0%	87,5%	-12,5%
C5	10	50,0%	50,0%	0,0%
C6	11	50,0%	75,0%	25,0%
C7	12	25,0%	37,5%	12,5%
C8	19	62,5%	25,0%	-37,5%
C9	23	37,5%	37,5%	0,0%
C10	22	62,5%	62,5%	0,0%
C11	30	87,5%	62,5%	-25,0%
C12	41	100,0%	50,0%	-50,0%
C13	42	50,0%	62,5%	12,5%
C14	44	37,5%	37,5%	0,0%
C15	48	50,0%	37,5%	-12,5%
C16	50	25,0%	12,5%	-12,5%
Mean E		60,9%	84,4%	23,4%
Mean C		57,8%	46,9%	-10,9%

Figure 3-2. Mean % of correct responses in the experimental and control groups for the contrasting pair of vowels /u/ and /ʊ/.
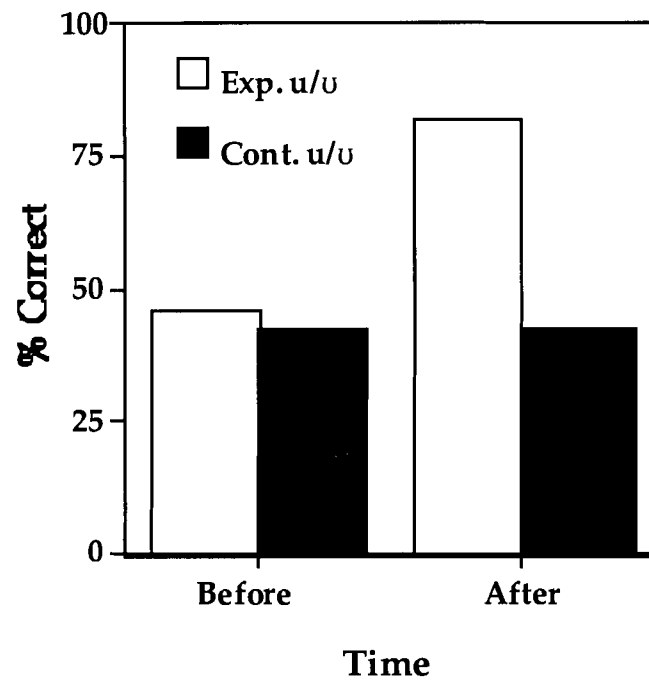
However, the control group as a whole remained exactly the same at both Time 1 and Time 2 (46.1% of ID scores). Six students showed a decrease in their ID scores from Time 1 to Time 2, three students remained the same and 6 increased. The results are summarized in Table 3-3.

3.2.2.3 Results on vowel contrast /ɑ/-/ʌ/

The /ɑ/-/ʌ/ contrast revealed the following results. To begin with, the scores in the pretest were very high for both the experimental (71.9%) and the control groups (72.7%) with a mean difference of 0.8% in favor of the control group. Both groups showed an increase in correct responses in the posttest. The percentage of ID scores for the experimental group increased from 71.9% at Time 1, to 86.7% at Time 2, with a mean difference of 14.8%. In the control group the percentage of ID scores increased from 72.7% to 74.2% showing a mean difference of 1.6% (Figure 3-3).

Individual scores show that 3 students from the experimental group decreased their percentage of ID scores, 1 remained the same, and 12 students showed an increased percentage of ID scores. In the control group, 5 students decreased, 5 remained with the same performance as Time 1, and 6 increased their scores. Table 3-4 summarizes these results.

3.4 STATISTICAL ANALYSIS

Although a simple comparison between the experimental and control groups reveals the differences in their performance, we used a mixed design (ANOVA) with one between factor –group (experimental and control) and two within factors –vowel (/i/-/ɪ/, /u/-/ʊ/, and /ɑ/-/ʌ/) and time (before and after). The results showed a significant group x time interaction [$F(1, 30) = 31.3, p < .05$]. Although there were significant main effects

Table 3.3 Total % ID scores for the vowel minimal pair /u/-/ʊ/ in the pre-post perception tests

Group	Code	Pretest	Posttest	Difference
E1	2	62,5%	100,0%	37,5%
E2	15	62,5%	75,0%	12,5%
E3	17	50,0%	75,0%	25,0%
E4	18	37,5%	75,0%	37,5%
E5	14	50,0%	87,5%	37,5%
E6	20	37,5%	62,5%	25,0%
E7	24	50,0%	100,0%	50,0%
E8	25	37,5%	100,0%	62,5%
E9	26	25,0%	37,5%	12,5%
E10	33	50,0%	75,0%	25,0%
E11	36	37,5%	62,5%	25,0%
E12	37	87,5%	75,0%	-12,5%
E13	39	50,0%	87,5%	37,5%
E14	40	25,0%	75,0%	50,0%
E15	45	37,5%	75,0%	37,5%
E16	46	62,5%	100,0%	37,5%
C1	6	62,5%	37,5%	-25,0%
C2	4	25,0%	50,0%	25,0%
C3	7	37,5%	50,0%	12,5%
C4	9	37,5%	37,5%	0,0%
C5	10	50,0%	37,5%	-12,5%
C6	11	50,0%	50,0%	0,0%
C7	12	37,5%	87,5%	50,0%
C8	19	87,5%	50,0%	-37,5%
C9	23	50,0%	12,5%	-37,5%
C10	22	37,5%	12,5%	-25,0%
C11	30	12,5%	62,5%	50,0%
C12	41	62,5%	25,0%	-37,5%
C13	42	37,5%	62,5%	25,0%
C14	44	75,0%	87,5%	12,5%
C15	48	50,0%	50,0%	0,0%
C16	50	25,0%	25,0%	0,0%
Mean E		47,7%	78,9%	31,3%
Mean C		46,1%	46,1%	0,0%

Figure 3-3. Mean % of correct responses in the experimental and control groups for the contrasting pair of vowels /ɑ/ and /ʌ/.
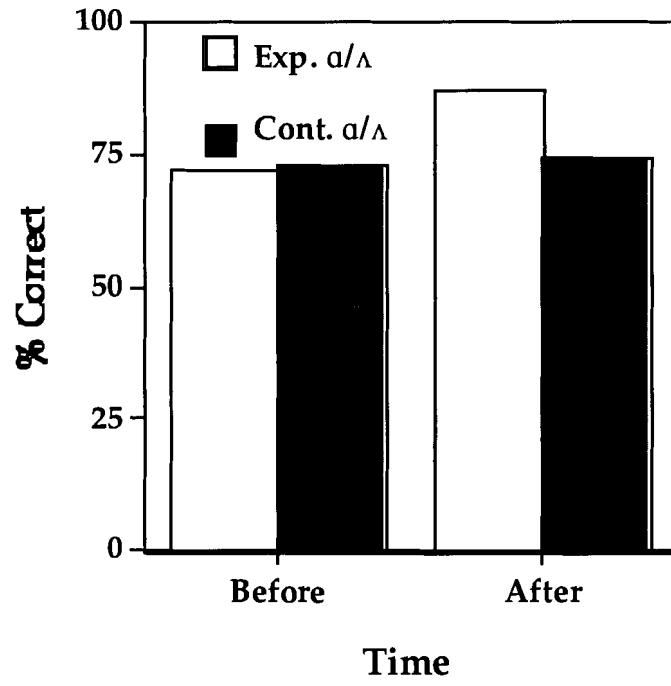
Table 3.4 Total % ID scores for the vowel minimal pair /ɑ/-/ʌ/ in the pre-post perception tests.

Group	Code	Pretest	Posttest	Difference
E1	2	100,0%	87,5%	-12,5%
E2	15	75,0%	75,0%	0,0%
E3	17	50,0%	87,5%	37,5%
E4	18	62,5%	75,0%	12,5%
E5	14	75,0%	100,0%	25,0%
E6	20	87,5%	75,0%	-12,5%
E7	24	37,5%	100,0%	62,5%
E8	25	62,5%	87,5%	25,0%
E9	26	62,5%	75,0%	12,5%
E10	33	62,5%	87,5%	25,0%
E11	36	62,5%	75,0%	12,5%
E12	37	75,0%	87,5%	12,5%
E13	39	87,5%	100,0%	12,5%
E14	40	75,0%	100,0%	25,0%
E15	45	87,5%	100,0%	12,5%
E16	46	87,5%	75,0%	-12,5%
C1	6	62,5%	87,5%	25,0%
C2	4	87,5%	62,5%	-25,0%
C3	7	75,0%	37,5%	-37,5%
C4	9	62,5%	87,5%	25,0%
C5	10	87,5%	100,0%	12,5%
C6	11	100,0%	87,5%	-12,5%
C7	12	50,0%	50,0%	0,0%
C8	19	87,5%	87,5%	0,0%
C9	23	62,5%	87,5%	25,0%
C10	22	75,0%	75,0%	0,0%
C11	30	62,5%	100,0%	37,5%
C12	41	87,5%	62,5%	-25,0%
C13	42	87,5%	75,0%	-12,5%
C14	44	50,0%	50,0%	0,0%
C15	48	62,5%	62,5%	0,0%
C16	50	62,5%	75,0%	12,5%
mean E		71,9%	86,7%	14,8%
mean C		72,7%	74,2%	1,6%

for group [$F(1,30)$ = 23.77, $p < .05$], vowel [$F(1,30)$ = 18.9, $p < .05$], and time [$F(1,30)$ = 18.2, $p < .05$], there was no significant group x vowel interaction [$F(1,30)$ = 2.26, $p < .11$], nor a vowel x time interaction [$F(1,30)$ = 1.6, $p < .19$]. This resulted in the fact that group, vowel, and time did not interact [$F(1,30)$ = 2.23, $p < .11$]. Although the effects are evident enough to be detected in isolation for vowel, time and group, they seem to be not relevant enough in the case of vowels as to be able to interact with time and group.

A test of simple main effects revealed no significant differences between the groups at Time 1. However, the test showed that there was a significant difference in group from Time 1 to Time 2. Moreover, a test of simple main effects showed no significant difference between the Times (before and after) for the control group, but a significant difference between the times for the experimental group.

The statistical results reveal that in fact the experimental group improved significantly more than the control group, and the lack of either vowel x time or group x vowel interactions shows that improvement on all vowel contrasts was about the same. A plausible explanation here could be given to training and not just to the fact that the students could have become familiar with the tests at Time 1 and Time 2.

3.5 INTERPRETATION OF THE RESULTS

The contrasting pair of vowels /i/-/ɪ/ revealed results which could possibly be attributed to the following factors. First, the students' awareness of the way the two sounds are produced may have increased. This was achieved through the explanation at the beginning of the unit, comparing the /i/-/ɪ/ sounds to situations where similar allophones are used in Spanish. The students themselves reported (during the training sessions) that the introductory part of each unit had helped them a lot. During this

section, specifically for the sound /ɪ/, the students were asked to remember the Spanish word *sí,* as pronounced when they were listening to someone telling them a story, and they did not believe what the person was saying. In Spanish this dubious *sí,* is the closest sound we have to the sound /ɪ/ in English.

Second, because these sounds were treated as "similar" based on the SLM, we assumed that it was going to be difficult for the students to perceive the "slight" differences between the sounds /i/-/ɪ/. So, careful attention was paid to establishing specific differences between them. Each sound was introduced on a separate day, and special emphasis was placed in the perception and production of the sound /ɪ/ as this short vowel does not have a counterpart in Spanish.

Third, at the end of Unit I (Appendix F), the students in the experimental group were exposed to many instances where these sounds contrast in the English language. As we said in Chapter II, this pair of contrasting vowels has a high functional load (95%). After introducing Lesson 3, the students became aware of the differences between 79 minimal pairs contrasting the /i/-/ɪ/ vowel sounds, and the exercises in this lesson provided them with opportunities to perceive these differences.

Factors which may have favored the improvement of the contrasting pair of vowels /u/-/ʊ/ may be similar to the ones pointed out for the /i/-/ɪ/ sounds. The explanation of how the sounds differ increased the students awareness in the perception of both /u/ and /ʊ/. The students expressed their preference for the introductory part of each lesson where articulatory differences between the sounds were explained.

Just as with the /ɪ/ sound, special attention was given to phoneme /ʊ/ during training. This might suggest that whenever there is a contrasting pair of vowels, where length is one of the problems, the more practice is given to the short vowel, the better the results in perception are. This assumption may have important implications in second language teaching.

Another possible explanation could be that the /u/-/ʊ/ vowel contrast has a very low functional load. This means that the number of pairs of vowels posing the contrast is very little. At the same time this implies that there was constant repetition of this restricted group in the training sessions, providing more opportunities for the participants to listen to the selected group of minimal pairs.

Finally, the /ɑ/-/ʌ/ contrast specially caught our attention. As we said before, the two groups showed very high scores in the pretest. Both /ɑ/ and /ʌ/ are "new" (following Flege's SLM way of labeling the sounds which L2 learners perceive as not having a direct correspondence to sounds in the L1), because from a perceptual point of view they do not have a direct analog in Spanish. The students had to create a new category for both sounds and this might explain why they had such good scores. The initial results in the two groups proved that the students seemed to have created this category with a certain degree of success. We could also observe that there was a noticeable improvement of the perception of the vowel sounds /ɑ/-/ʌ/ in the experimental group at posttest which might be attributable to training, and that the control group improved slightly as well.

CHAPTER IV

RESULTS AND ANALYSIS ON PRODUCTION

4.1 INTRODUCTION

This chapter includes the results of the study related to production. It first lays out the results on the category goodness test and includes a detailed description of all the procedures carried out to analyze the data with a certain degree of reliability. Then, the chapter presents and discusses the results related to the comprehensibility test. At the end, it statistically analyzes the results of the category goodness test and the comprehensibility test.

4.2 RESULTS

4.2.1 Results for the category goodness test

4.2.1.1 Results on a 9-point scale

In Chapter II we noted that we had selected 12 words for the category goodness test. As a reminder, the words are:

/ɑ/-/ʌ/	/ʊ/-/u/	/ɪ/-/i/
caught-cut	pull-pool	Tim-seems
boss-bus	took-too	mix-weeks

The final individual words used for the category goodness test were 816 speech samples. A part of this sample was elicited from the participants (32 participants x 2

times x 12 words = 768) and the other part was elicited from the native speakers (4 NSs x 12 words = 48).

The rating procedure took place in the phonetics lab in the Department of Linguistics at Simon Fraser University. The raters were three female NSs who were recruited from this department. They had grown up in Canada west of Quebec. The three of them were graduate students and reported normal hearing. Two of these raters had experience teaching ESL, one did not.

On a 9-point scale, the raters had to rate how well the word was pronounced, 1 being very poor and 9 being very good. The mean rating scores for all the words representing the pairs of vowels /ɑ/-/ʌ/, /ʊ/-/u/ and /ɪ/-/i/ are grouped in Table 4.1.

Table 4.1. Three-rater category goodness mean scores on a 9-point scale for all the words representing the contrasting pairs of vowels.

Stimuli	Control			Experimental		
	Pretest	Posttest	Difference	Pretest	Posttest	Difference
caught	2,6	3,0	0,4	4,0	4,7	0,7
cut	3,6	3,8	0,2	4,1	4,8	0,7
boss	4,7	5,2	0,5	5,2	4,9	-0,3
bus	6,3	6,2	-0,1	6,2	6,5	0,3
pool	3,9	4,1	0,2	3,5	3,5	0,0
pull	5,1	4,1	-1,0	5,0	4,3	-0,7
too	6,1	6,8	0,7	5,5	5,9	0,4
took	5,4	4,8	-0,6	4,1	5,0	0,9
weeks	6,2	6,5	0,3	6,7	7,1	0,4
mix	4,9	6,1	1,2	4,6	4,6	0,0
seems	7,1	6,1	-1,0	6,5	6,6	0,1
Tim	3,9	3,9	0,0	4,0	3,8	-0,2
Total			0,8			2,3

Overall there was a slight improvement between Time 1 and Time 2 in both groups. The difference between Time 1 and Time 2 in the experimental group was greater (2,3) than in the control group (0,8).

However, taking into consideration that the data collected was the result of judgements made by three different raters on the stimuli presented to them, we had to consider how reliable the data was. In order to do this, we needed to estimate the inter-rater reliability of the scores. Inter-rater reliability means the extent to which the three different raters agreed on the data collected. The degree of reliability is usually expressed as the percentage of the total number of ratings that are the same for all raters (Selinger and Shohamy, 1990).

A simple comparison of the ratings revealed little agreement among the three raters in the study. After computing the results of the word *caught*, for example, we noticed that there were only a few instances where there was total agreement among the raters (Table 4-2).

We noticed that when the three raters heard the word *caught* they totally agreed only in two occasions (13%) for the control group, and not even once for the experimental group (0%). The three raters gave the highest score (9) to two of the four native speakers (50%). But in our analysis, we did not want to include items on which there was not good agreement.

To increase the number of "items used" for the analysis, we grouped the ratings into three sets of scores: low, medium, and high. On the 9-point scale, low scores included the ratings closer to a very poor pronunciation (1, 2, and 3); medium scores indicated an acceptable pronunciation (4, 5, and 6); and high scores included the ratings

Table 4.2 Three-rater mean scores on a 9-point scale for the word *caught* at pretest.

Group	Code	Rater 1	Rater 2	Rater 3	mean	Agreement
C1	6	1	4	6	3,7	0
C2	4	5	2	2	3,0	0
C3	7	1	1	1	1,0	1
C4	9	2	4	6	4,0	0
C5	10	3	2	5	3,3	0
C6	11	1	3	7	3,7	0
C7	12	1	1	1	1,0	1
C8	19	6	2	3	3,7	0
C9	23	1	3	1	1,7	0
C10	22	1	6	3	3,3	0
C11	30	1	2	1	1,3	0
C12	41	1	2	1	1,3	0
C13	42	8	1	1	3,3	0
C14	44	1	3	1	1,7	0
C15	48	1	2	1	1,3	0
C16	50	3	4	7	4,7	0
E1	2	2	2	9	4,3	0
E2	15	8	4	1	4,3	0
E3	17	1	3	7	3,7	0
E4	18	4	4	2	3,3	0
E5	14	9	7	5	7,0	0
E6	20	2	1	5	2,7	0
E7	24	5	3	6	4,7	0
E8	25	1	2	1	1,3	0
E9	26	9	5	3	5,7	0
E10	33	1	4	2	2,3	0
E11	36	3	3	1	2,3	0
E12	37	8	6	1	5,0	0
E13	39	7	4	1	4,0	0
E14	40	6	1	5	4,0	0
E15	45	8	8	7	7,7	0
E16	46	1	3	2	2,0	0
Mean C		2,3	2,6	2,9	2,6	2
Mean E		4,7	3,8	3,6	4,0	0
					33%	13%
					44%	0%
native	F1	9	9	9	9,0	1
native	F2	4	8	9	7,0	0
native	M1	9	9	9	9,0	1
native	M2	9	7	9	8,3	0
					Total NS	2
						50%

58

related to a very good pronunciation (7, 8, and 9). These items would indicate at least a certain degree of agreement among the raters.

We decided to see if the regrouping would yield more number of items to be analyzed with a certain degree of reliability. Table 4-3 shows that, with this new way of interpreting the results, the percentages representing the number of items on which 'agreement' occurred increased from 13% to 44% for the control group, and from 0% to 25% in the experimental group. All of the items (100%) in the NSs' group could be analyzed too.

A more careful look at the word *caught* indicated that by removing the cases where there was more than a two-point spread, we could be able to analyze the data. But we noticed that if we did that, we would be left out with a very small population and with no statistical equivalence among the groups. The population would decrease from 32 subjects to 11—7 students in the control group and only 4 in the experimental group.

Still in search for a solution to increase the number of items used for the word *caught*, we noticed that there was more agreement between Raters 1 and 2 than with Rater 3 —who did not have experience teaching ESL. Just for a test, we excluded Rater 3 from the study, and there was an evident increase of items to be used (Table 4-4). We had nine cases in both groups representing 56% in each group.

We computed the results for the word *caught* at posttest, and the words *cut, boss, bus, pool, pull, too, took, weeks, mix, seems,* and *Tim* at pretest and posttest. We examined the results and they revealed that the number of items we could use was not enough for our analysis (Table 4-5). We decided keep Rater 3, and analyze the possible causes of this lack of agreement taking into consideration the regrouped data as shown in

Table 4.3 Three-rater regrouped mean scores on a 9-point scale for the word *caught* at pretest.

Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	1	4	6	3,7	0
C2	4	5	2	2	3,0	0
C3	7	1	1	1	1,0	1
C4	9	2	4	6	4,0	0
C5	10	3	2	5	3,3	0
C6	11	1	3	7	3,7	0
C7	12	1	1	1	1,0	1
C8	19	6	2	3	3,7	0
C9	23	1	3	1	1,7	1
C10	22	1	6	3	3,3	0
C11	30	1	2	1	1,3	1
C12	41	1	2	1	1,3	1
C13	42	8	1	1	3,3	0
C14	44	1	3	1	1,7	1
C15	48	1	2	1	1,3	1
C16	50	3	4	7	4,7	0
E1	2	2	2	9	4,3	0
E2	15	8	4	1	4,3	0
E3	17	1	3	7	3,7	0
E4	18	4	4	2	3,3	0
E5	14	9	7	5	7,0	0
E6	20	2	1	5	2,7	0
E7	24	5	3	6	4,7	0
E8	25	1	2	1	1,3	1
E9	26	9	5	3	5,7	0
E10	33	1	4	2	2,3	0
E11	36	3	3	1	2,3	1
E12	37	8	6	1	5,0	0
E13	39	7	4	1	4,0	0
E14	40	6	1	5	4,0	0
E15	45	8	8	7	7,7	1
E16	46	1	3	2	2,0	1
Mean C		2,3	2,6	2,9	2,6	7
Mean E		4,7	3,8	3,6	4,0	4
				Total C	33%	44%
				Total E	44%	25%
native	F1	9	9	9	9,0	1
native	F2	4	8	9	7,0	1
native	M1	9	9	9	9,0	1
native	M2	9	7	9	8,3	1
				Mean	8,3	4
					Total	100%

60

Table 4.4 Two-rater mean scores on a 9-point scale for the word *caught* at pretest.

Group	Code	Rater 1	Rater 2	Mean	Items used
C1	6	1	4	2,5	0
C2	4	5	2	3,5	0
C3	7	1	1	1,0	1
C4	9	2	4	3,0	0
C5	10	3	2	2,5	1
C6	11	1	3	2,0	1
C7	12	1	1	1,0	1
C8	19	6	2	4,0	0
C9	23	1	3	2,0	1
C10	22	1	6	3,5	0
C11	30	1	2	1,5	1
C12	41	1	2	1,5	1
C13	42	8	1	4,5	0
C14	44	1	3	2,0	1
C15	48	1	2	1,5	1
C16	50	3	4	3,5	0
E1	2	2	2	2,0	1
E2	15	8	4	6,0	0
E3	17	1	3	2,0	1
E4	18	4	4	4,0	1
E5	14	9	7	8,0	1
E6	20	2	1	1,5	1
E7	24	5	3	4,0	0
E8	25	1	2	1,5	1
E9	26	9	5	7,0	0
E10	33	1	4	2,5	0
E11	36	3	3	3,0	1
E12	37	8	6	7,0	0
E13	39	7	4	5,5	0
E14	40	6	1	3,5	0
E15	45	8	8	8,0	1
E16	46	1	3	2,0	1
Mean C		2,3	2,6	2,5	9
Mean E		4,7	3,8	4,3	9
				Total C	56%
				Total E	56%
native	F1	9	9	9,0	1
native	F2	4	8	6,0	1
native	M1	9	9	9,0	1
native	M2	9	7	8,0	1
			Mean	8,0	4
				Total	100%

Table 4-5. Rating scores on NS production on a 9-point scale for the category goodness test.

Stimuli	mean score	items used	% of items used
caught	8,3	3	75%
cut	5,0	0	0%
boss	7,2	1	25%
bus	7,8	2	50%
pool	7,2	1	25%
pull	5,3	0	0%
too	7,5	2	50%
took	5,8	0	0%
weeks	7,8	3	75%
mix	8,5	3	75%
seems	7,1	1	25%
Tim	8,2	3	75%
Total	**7,1**	**19**	**40%**

Table 4-3[8]. The results on the total of items used for the words *caught, cut, boss, bus, pool, pull, too, took, weeks, mix, seems* and *Tim*, are presented in Table 4-6. The highest percentage achieved was for the word *cut* at posttest (56%). But still the number of items we could use representing a certain degree of reliability was very low.

We analyzed what the causes of this low inter-rater agreement could have been, and we arrived at the conclusion that the task the raters had been asked to solve was a

---

[8] Tables 4-2 and 4-3 give an example of the order in which the data was regrouped in order to increase the reliability of the scores. We just included the data related to the word *caught* in the main body of our thesis. However, Appendix G includes the detailed data related to the category goodness test on a 9-point scale for all the words.

difficult one. We noticed that the raters had agreed either when the pronunciation was very poor (ratings 1 or 2) or very good (ratings 8 or 9). This wide range made us think that if the point scale were dropped to 5, the difficulty of the task would decrease, and the raters would have a narrower scope to rate the samples.

We redesigned the listening task using the same program we had used before (Audio Experiment Generator 4.0), but this time we just used half of the sample re-rated in order to see if the agreement among the raters increased when the scale was dropped.

4.2.1.2 Results on a 5-point scale

The words *caught, boss, pull, took, Tim*, and *mix* were selected for the re-rating procedure. The criterion for selecting this sample was simple. If the raters gave good rating scores for the words representing the sounds which Spanish speakers do not have in their L1, we could predict that the production of equal sounds was going to be good.

The re-rating procedure took place two weeks after the initial rating. Because the period of time between the initial and the second rating tasks was so short, we decided to recruit three new raters. This recruitment was necessary because first, the stimuli were the same, and second, because the rating task was similar, and there was a possibility that the first raters would become used to the pronunciation of the Spanish speakers, thus giving them better ratings.

The new raters were three female NSs who were enrolled as undergraduate students in the Department of Linguistics at Simon Fraser University. These raters had grown up in Canada west of Quebec. The three of them reported normal hearing and two of them had experience teaching ESL, but one did not.

Because it was crucial for the study to have inter-rater agreement, we computed the results of only one word first, to establish a comparison between the first and the

Table 4-6. Total of items used on a 9-point scale for the category goodness test

Stimuli	Pretest			
	Control	Experimental	Total	%
caught	7	4	11	34%
cut	9	5	14	44%
boss	1	0	1	3%
bus	5	7	12	38%
pool	1	3	4	13%
pull	3	4	7	22%
too	9	5	14	44%
took	3	1	4	13%
weeks	6	3	9	28%
mix	2	4	6	19%
seems	6	5	11	34%
Tim	2	5	7	22%
Total	54	46	100	26%

Stimuli	Posttest			
	Control	Experimental	Total	%
caught	4	3	7	22%
cut	11	7	18	56%
boss	1	1	2	6%
bus	7	3	10	31%
pool	2	4	6	19%
pull	6	2	8	25%
too	8	7	15	47%
took	4	4	8	25%
weeks	6	7	13	41%
mix	4	4	8	25%
seems	4	6	10	31%
Tim	6	6	12	38%
Total	63	54	117	30%

second ratings. As an example of what we did, this time we included the results for the word *mix*. We followed the same procedure we carried out for the word *caught*. The initial results revealed that there was 0% of total agreement among the raters in both the experimental and the control groups. The raters totally agreed in one instance with the native speakers. Table 4-7 summarizes these results.

The next step included regrouping the scores in search for more number of items which could be used in the analysis of the data. This time we considered excluding an item if there was more than a two-point spread among the raters. For example, lets suppose that the stimulus is the word *mix* elicited by Student Code C-50. At pretest, Raters A, B, and C gave the scores 5, 1, and 4 respectively. According to our criteria, we could keep the scores given by Raters A and C, but not the one given by Rater A. The difference between the scores of Raters A and B (5 − 1 = 4), equals four, and that between Raters C and B (4 − 1 = 3) equals three. So we concluded that there was the item had to be left out.

Based on the previous idea, we regrouped the data for the word *mix* again, and the number of items used increased from 0% in both groups to 50% in the control group and to 56% in the experimental group. In the NSs' group it increased from 25% to 75% (Table 4-8).

Looking at these results, we realized that there were two raters who agreed a lot among themselves (Raters A and C), and not that much with the other rater (Rater B). When we removed rater B from the data, the number of items we could used increased from 50% in the control group to 100%, and from 56% in the experimental group to 100%.

Table 4.7 Three-rater mean scores on a 5-point scale for the word *mix* at pretest.

Group	Code	Rater A	Rater B	Rater C	mean	agreement
C1	10	4	2	3	3,0	0
C2	11	5	2	3	3,3	0
C3	12	5	4	5	4,7	0
C4	19	4	1	3	2,7	0
C5	22	4	1	3	2,7	0
C6	23	5	2	4	3,7	0
C7	30	5	4	4	4,3	0
C8	4	5	1	4	3,3	0
C9	41	4	3	3	3,3	0
C10	42	5	3	4	4,0	0
C11	44	5	1	3	3,0	0
C12	48	4	1	3	2,7	0
C13	50	5	1	4	3,3	0
C14	6	5	3	3	3,7	0
C15	7	4	2	2	2,7	0
C16	9	3	1	3	2,3	0
E1	14	5	1	3	3,0	0
E2	15	5	3	3	3,7	0
E3	17	5	1	4	3,3	0
E4	18	5	2	3	3,3	0
E5	2	3	1	2	2,0	0
E6	20	4	2	3	3,0	0
E7	24	5	2	2	3,0	0
E8	25	3	1	3	2,3	0
E9	26	4	1	3	2,7	0
E10	33	4	2	3	3,0	0
E11	36	3	2	3	2,7	0
E12	37	3	2	2	2,3	0
E13	39	5	3	4	4,0	0
E14	40	4	1	4	3,0	0
E15	45	5	2	3	3,3	0
E16	46	5	4	3	4,0	0
Mean C		4,5	2	3,4	3,3	0
Mean E		4,3	1,9	3,0	3,0	0
				Total C	66%	0%
				Total E	60%	0%
Native	F2	5	4	4	4,3	0
Native	F1	5	5	5	5,0	1
Native	M2	5	2	5	4,0	0
Native	M1	5	4	4	4,3	0
				Mean	4,4	1
				Total		25%

66

Table 4.8 Three-rater regrouped mean scores on a 5-point scale for the word *mix* at pretest

Group	Code	Rater A	Rater B	Rater C	mean	items used
C1	10	4	2	3	3,0	1
C2	11	5	2	3	3,3	0
C3	12	5	4	5	4,7	1
C4	19	4	1	3	2,7	0
C5	22	4	1	3	2,7	0
C6	23	5	2	4	3,7	0
C7	30	5	4	4	4,3	1
C8	4	5	1	4	3,3	0
C9	41	4	3	3	3,3	1
C10	42	5	3	4	4,0	1
C11	44	5	1	3	3,0	0
C12	48	4	1	3	2,7	0
C13	50	5	1	4	3,3	0
C14	6	5	3	3	3,7	1
C15	7	4	2	2	2,7	1
C16	9	3	1	3	2,3	1
E1	14	5	1	3	3,0	0
E2	15	5	3	3	3,7	1
E3	17	5	1	4	3,3	0
E4	18	5	2	3	3,3	0
E5	2	3	1	2	2,0	1
E6	20	4	2	3	3,0	1
E7	24	5	2	2	3,0	0
E8	25	3	1	3	2,3	1
E9	26	4	1	3	2,7	0
E10	33	4	2	3	3,0	1
E11	36	3	2	3	2,7	1
E12	37	3	2	2	2,3	1
E13	39	5	3	4	4,0	1
E14	40	4	1	4	3,0	0
E15	45	5	2	3	3,3	0
E16	46	5	4	3	4,0	1
Mean C		4,5	2	3,4	3,3	8
Mean E		4,3	1,9	3,0	3,0	9
				Total C	66%	50%
				Total E	60%	56%
Native	F2	5	4	4	4,3	1
Native	F1	5	5	5	5,0	1
Native	M2	5	2	5	4,0	0
Native	M1	5	4	4	4,3	1
				Mean	4,4	3
				Total		75%

67

Table 9-10 shows that now there was 100% of items we could use for the NSs group as well . These results were very encouraging, so we regrouped the data for the rest of the words and excluded Rater B[9] from the study. We computed these results which appear in Tables 4-10, 4-11, and 4-12.

Table 4-10 shows that the percentage of items used was more than 93 % and Table 4-11 reveals an increase rate of agreement on NS production (from 75% to 83.3%). To re-test the reliability of these scores, we recruited a phonetically trained subject and asked him to rate the words *caught, pull,* and *mix.* When we computed these results, we noticed that agreement among Raters A and C and the phonetically trained subject was kept making the data reliable and worthy of being analyzed.

The results in Table 4-12 indicate very little improvement for the control group (0,3) and regression in the experimental group (–0,5). This was the reality in most of the cases except for the word *caught.* The word *caught* shows the highest improvement in both groups (0,5 in the control group and 0,7 in the experimental group). These results cannot be attributed to training. To begin with, the mean scores in both groups increased. Second, the English vowel sound /ɑ/ is represented by the vowels *au* in writing, and in general, the spelling of the word is somewhat "strange" for Spanish speakers. The Spanish speakers seemed to give special attention to the grapheme-phoneme correspondence of this word, thus improving the production of the vowel /ɑ/. Accordingly, we predict that this knowledge could be transferred to the production of the words *taught, bought, ought,* etc.

When we looked at how individual students had been rated in the different words,

---

[9] Rater B did not have experience teaching ESL.

Table 4.9 Two-rater mean scores on a 5-point scale for the word *mix* at pretest

Group	Code	Rater A	Rater C	Mean	items used
C1	10	4	3	3,5	1
C2	11	5	3	4,0	1
C3	12	5	5	5,0	1
C4	19	4	3	3,5	1
C5	22	4	3	3,5	1
C6	23	5	4	4,5	1
C7	30	5	4	4,5	1
C8	4	5	4	4,5	1
C9	41	4	3	3,5	1
C10	42	5	4	4,5	1
C11	44	5	3	4,0	1
C12	48	4	3	3,5	1
C13	50	5	4	4,5	1
C14	6	5	3	4,0	1
C15	7	4	2	3,0	1
C16	9	3	3	3,0	1
E1	14	5	3	4,0	1
E2	15	5	3	4,0	1
E3	17	5	4	4,5	1
E4	18	5	3	4,0	1
E5	2	3	2	2,5	1
E6	20	4	3	3,5	1
E7	24	5	2	3,5	1
E8	25	3	3	3,0	1
E9	26	4	3	3,5	1
E10	33	4	3	3,5	1
E11	36	3	3	3,0	1
E12	37	3	2	2,5	1
E13	39	5	4	4,5	1
E14	40	4	4	4,0	1
E15	45	5	3	4,0	1
E16	46	5	3	4,0	1
Mean C		4,5	3,4	3,9	16
Mean E		4,3	3,0	3,6	16
			Total C	78%	100%
			Total E	72%	100%
Native	F2	5	4	4,5	1
Native	F1	5	5	5,0	1
Native	M2	5	5	5,0	1
Native	M1	5	4	4,5	1
			Mean	4,8	4
			Total		100%

Table 4-10. Total of items used on a 5-point scale (category goodness test)

Stimuli	Pretest			Posttest		
	Control	Experimental	Total	Control	Experimental	Total
caught	12	16	28	14	15	29
boss	13	15	28	16	16	32
pull	16	16	32	16	16	32
took	16	15	31	15	16	31
mix	16	16	32	16	16	32
Tim	16	16	32	15	15	30
Total	89	94	183	92	94	186
% items	93%	98%	95%	96%	98%	97%

Table 4-11. Rate of agreement on NS production on a 5-point scale:
Category goodness test

Stimuli	mean score	agreement	% of agreement
caught	5,0	4	100%
boss	4,7	3	75%
pull	4,9	3	75%
took	5,0	4	100%
mix	4,8	2	50%
Tim	5,0	4	100%
Total	4.9	20	83.3%

Table 4-12. Two-rater mean-rating scores on a 5-point scale for the words *caught, boss, pull, took, mix,* and *Tim.*

Stimuli	Control			Experimental		
	Pretest	Posttest	Difference	Pretest	Posttest	Difference
caught	2,3	2,8	0,5	2,6	3,3	0,7
boss	3,4	3,5	0,1	3,4	3,3	-0,1
pull	3,2	3,2	0,0	3,4	3,2	-0,2
took	3,5	3,2	-0,3	3,4	3,2	-0,2
mix	3,9	4,0	0,1	3,6	3,2	-0,4
Tim	3,4	3,3	-0,1	3,0	2,7	-0,3
Total			0,3			-0,5

we found that the mean rating scores had in fact increased from Time 1 to Time 2. Table 4-13 shows that five students in the control group increased their mean scores from Time 1 to Time 2 for the vowel /ɑ/ represented by the words *caught* and *boss* (Codes C6, C12, C19, C44, and C48). The results revealed that 69% of the students (11 out of 16) improved their pronunciation from Time 1 to Time 2 for the word *caught*, and 48 % (7 out of 16) for the word *boss* (detailed results for all the words are given in Appendix H).

In the experimental group, four students improved their pronunciation for the vowel sound /ɑ/ (Codes E15, E17, E26, and E40). The scores increased for both words *caught* and *boss*. The word caught showed a 75% improvement (12 students) and six students (38%) in the experimental group improved their pronunciation for the word *boss*.

The results for the English vowel sound /ʊ/ revealed that two students in the control group had improved their pronunciation (Codes C6, and C10), and two students in the experimental group achieved better scores at posttest (Codes E14, and E46). In all, seven students in the control group increased their scores for the word *pull* (44%), and only three for the word *took* (19%). This low result for the word *took* may be attributed to spelling. When Spanish speakers see a word written with a double *o* (oo) they tend to pronounce the English vowel /u/ as there are many words with the sound /u/ written with double *o* too, and as the students were reading, they might have been confused. The results of the experimental group indicated that eight students had improved their pronunciation of the words *pull* and *took* respectively. Four students pronounced the word *pull* with a better pronunciation (25%) and the other four students were given a higher rating for the word *took* (25%).

Table 4-13. Mean rating scores in the category goodness test for the words *caught, pull, took, mix,* and *Tim* on a 5-point scale.

Group	Code	CAUGHT 1		BOSS 2		PULL 3		TOOK 4		MIX 5		TIM 6	
		Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
C1	10	3,0	2,5	4,0	3,0	2,5	3,5	3,0	3,5	3,5	3,0	4,0	1,5
C2	11	2,0	3,0	4,0	3,0	3,0	4,0	4,5	3,5	4,0	4,5	2,5	3,0
C3	12	2,0	2,5	3,0	5,0	3,0	4,0	3,5	3,5	5,0	5,0	4,0	3,5
C4	19	2,5	3,5	2,5	3,0	3,0	3,0	3,5	3,5	3,5	4,5	4,0	4,0
C5	22	3,0	3,0	4,5	3,5	4,0	2,0	3,0	3,0	3,5	4,0	4,0	4,0
C6	23	2,0	3,5	4,5	3,0	4,5	2,0	3,5	2,0	4,5	4,5	2,5	2,0
C7	30	2,5	2,0	2,5	4,0	4,0	4,5	4,0	2,5	4,5	5,0	2,5	4,0
C8	4	2,0	2,5	3,0	3,0	1,5	1,5	4,5	4,5	4,5	4,5	4,0	3,0
C9	41	2,5	3,0	2,5	5,0	4,0	4,0	2,0	1,0	3,5	3,5	4,0	4,0
C10	42	1,0	1,0	3,5	3,0	3,5	2,5	3,5	4,5	4,5	3,5	3,0	2,5
C11	44	1,5	2,0	2,5	3,5	3,0	2,0	4,0	3,5	4,0	4,0	4,0	4,5
C12	48	2,5	3,5	3,0	5,0	3,0	4,0	3,5	3,5	3,5	4,0	3,0	4,0
C13	50	3,5	3,0	4,0	1,5	3,5	4,0	3,5	3,5	4,5	4,0	3,0	3,5
C14	6	3,0	3,5	3,5	4,0	3,0	4,5	2,5	3,0	4,0	3,0	4,5	3,5
C15	7	1,5	2,0	3,0	3,0	3,0	3,0	4,0	2,5	3,0	4,5	3,0	2,5
C16	9	3,0	3,5	4,0	4,0	3,0	3,0	4,0	3,0	3,0	3,0	3,0	3,5
E1	14	4,0	1,0	4,0	3,5	2,5	3,0	4,0	4,5	4,0	4,5	2,0	3,5
E2	15	1,5	3,0	3,0	5,0	3,0	5,0	4,0	4,0	4,0	3,5	5,0	2,0
E3	17	3,5	4,0	3,5	4,0	3,0	2,5	3,5	3,0	4,5	3,5	2,5	2,5
E4	18	1,5	2,5	3,0	2,5	2,5	2,5	2,0	2,5	4,0	4,0	3,5	4,5
E5	2	3,0	3,5	2,0	2,0	2,5	2,0	4,0	3,0	2,5	1,5	2,5	2,0
E6	20	2,5	4,0	3,5	3,0	3,0	3,0	3,5	3,0	3,5	4,5	3,0	4,0
E7	24	4,0	3,5	3,5	4,5	4,0	4,5	4,0	4,0	3,5	2,0	3,0	3,0
E8	25	2,5	2,5	4,0	4,5	5,0	3,0	2,0	1,0	3,0	3,5	4,0	2,5
E9	26	3,0	4,0	3,5	4,5	3,5	2,5	3,5	2,5	3,5	2,5	3,0	2,5
E10	33	2,5	3,5	3,5	3,0	4,5	3,5	3,0	3,5	3,5	3,5	2,5	2,5
E11	36	1,0	3,5	3,5	2,5	4,0	2,5	4,5	3,5	3,0	3,0	2,0	2,5
E12	37	2,0	4,0	2,0	2,0	2,0	1,5	3,0	3,0	2,5	2,0	2,5	2,0
E13	39	1,0	3,0	4,0	3,5	4,5	2,5	4,0	3,0	4,5	2,5	3,0	2,0
E14	40	3,5	4,0	3,0	3,5	4,5	4,0	4,0	3,0	4,0	4,5	3,5	2,0
E15	45	3,5	2,5	4,5	2,5	3,5	3,5	3,0	3,5	4,0	3,5	3,0	2,5
E16	46	2,0	3,5	4,0	3,0	2,5	5,0	2,5	4,0	4,0	2,5	3,5	3,0
Cont.		2,3	2,8	3,4	3,5	3,2	3,2	3,5	3,2	3,9	4,0	3,4	3,3
Exp.		2,6	3,3	3,4	3,3	3,4	3,2	3,4	3,2	3,6	3,2	3,0	2,7

73

The last two words (*mix and Tim*) representing the English vowel sound /ɪ/ revealed that 3 students in the control group had improved their pronunciation for both words (Codes C11, C30, and C48). Only two students in the experimental group increased their means scores from Time 1 to Time 2 (Codes E14 and E20). In all 12 students in the control group had better scores for the words *mix* and *Tim* (six each which equals 38%), while eight students showed improvement on the pronunciation of these words (four each, 25%).

It is interesting to note that the results for Student Code C6 show that this student improved his pronunciation for the vowel sounds /ɑ/ and /ɪ/, while Student Code E14 improved her pronunciation for the sounds /ʊ/ and /ɪ/. These last results are more consistent than those presented by the student in the experimental group. Our interpretation of these results is that the vowel sounds /ʊ/ and /ɪ/ present the same type of difficulty for the Spanish speaker. The duration differences, and the fact that both are produced with the muscle more relaxed makes us think that in fact, this specific student perceived these differences and transferred them to the production of these sounds. There are other factors that can be attributed to the previous results. These factors will be explained in Chapter V.

4.2.2 Results for the comprehensibility test

4.2.2.1 Results on a 9-point scale

As we said in Chapter II, three sentences were selected from the paragraph for the comprehensibility test. The sentences were:

She is always sick.

she:　　　　　representing the English vowel sound /i/

always: representing the English vowel sound /ɑ/

sick: representing the English vowel sound /ɪ/

I think he is a fool.

think: representing the English vowel sound /ɪ/

fool: representing the English vowel sound /u/

Luke is strong.

Luke: representing the English vowel sound /u/

strong: representing the English vowel sound /ɑ/

The final sentence stimuli used for the comprehensibility test consisted of 204 speech samples. A group of this sample was elicited from the participants (32 participants x 2 times x 3 sentences = 192) and the other group was elicited from the native speakers (4 NSs x 3 words = 12).

The rating procedure took place under the same conditions and with the same raters used for the category goodness. On a 9-point scale, the raters had to rate how comprehensible the sentences were: 1 being poor pronunciation and 9 being native like. The mean rating scores for all the sentences are grouped in Table 4-14.

These initial results revealed that overall there had been a slight improvement in the global pronunciation of the vowels included in the study. However, the experience from the analysis of the results of the category goodness test indicated that the reliability of these scores had to be tested.

We followed the same procedures as for the category goodness test (total agreement, regrouping, native speaker agreement). The results revealed that, after

regrouping the scores[10] the items in which the raters had agreed were very few, and the percentages indicating agreement among the raters was really low. Table 4-15 summarizes these results.

Table 4-14. Comprehensibility test mean-rating scores on a 9-point scale for all the sentences.

	Control			Experimental		
Stimuli	Pretest	Posttest	Difference	Pretest	Posttest	Difference
Sentence 1	5,3	5,0	-0,3	4,8	4,7	-0,1
Sentence 2	4,9	5,2	0,3	4,6	4,8	0,2
Sentence 3	4,0	4,9	0,9	3,3	4,4	1,0
Total			0,9			1,1

Sentence 1: She is always sick.
Sentence 2: I think he is a fool.
Sentence 3: Luke is strong.

Table 4-15. Total of items used on a 9-point scale (comprehensibility test)

	Pretest			Posttest		
Stimuli	Control	Experimental	% items used	Control	Experimental	% items used
Sentence 1	5	6	34%	6	4	31%
Sentence 2	4	5	28%	4	3	22%
Sentence 3	3	2	16%	4	5	28%
Total	12	13	26%	14	12	27%

Although the raters had not agreed among themselves when rating the stimuli elicited by the students, they had better agreement when they rated the native speakers

---

[10] The scores were regrouped following the same procedure as the category goodness test (high-medium-low). The detailed results are presented in Appendix I.

(Table 4-16). Based on the previous results, we knew that we could not analyze the data because it was not reliable. So we dropped the scale from a 9 to a 5-point.

Table 4-16. Rate of agreement on NS production on a 9-point scale: Comprehensibility test

Stimuli	mean score	agreement	% of agreement
Sentence 1	8,9	2	50%
Sentence 2	8,8	3	75%
Sentence 3	8,9	2	50%
Total mean	8,9	7	58.3%

Sentence 1: She is always sick.
Sentence 2: I think he is a fool.
Sentence 3: Luke is strong.

4.2.2.2 Results on a 5-point scale

We used the same three sentences presented in the 9-point scale analysis and we followed a similar procedure as the one explained before except that this time, we asked raters A, B, and C to report information related to the quality of the recordings.

The reason that we did this was because some of the raters had indicated that the quality of the recordings was not that good, and that sometimes it was difficult for them to give an appropriate rating.

We met with Raters A, B, and C and had a listening session where the raters were asked to rate the stimuli. They were provided with a booklet where they had to write their code, and mark their rating scores (Appendix J). The person conducting the experiment

played each stimulus once and, when the raters indicated that they had completed the task, the next stimulus was played.

There was one stimulus from the control group (C-41 at pretest) which did not play at all because of a computer error, neither did C-4 at posttest. So we removed these items from the analysis of the sentences.

When we analyzed these results, we noticed that only one rater had reported problems with the quality of the recordings. Then, we did not compute these results. After entering the data in the computer, we noticed an increase in the number of items showing better agreement among the raters (Tables 4-17 and 4-18).

Table 4-17. Total of items used on a 5-point scale (comprehensibility test).

Stimuli	Pretest			Posttest		
	Control	Experimental	% items used	Control	Experimental	% items used
Sentence 1	15	16	100%	16	16	100%
Sentence 2	15	16	100%	15	16	100%
Sentence 3	15	16	100%	16	16	100%
Total	45	48	100%	47	48	100%
% items used	100%	100%	100%	100%	100%	100%

Table 4-18. Rate of agreement on NS production on a 5-point scale: Comprehensibility test

Stimuli	mean score	agreement	% of agreement
Sentence 1	5,0	4	100%
Sentence 2	5,0	4	100%
Sentence 3	5,0	4	100%
Total mean	5,0	4	100%

Sentence 1: She is always sick.
Sentence 2: I think he is a fool.
Sentence 3: Luke is strong.

We computed the mean scores for the three sentences (Appendix K) and there was no noticeable overall improvement between the experiment and the control groups at Time 1 and Time 2. A difference could be noticed when we analyzed individual sentences (Table 4-19). As we can see, the sentence *Luke is strong*, was the one with the higher improvement in both groups. This sentence has words with two of the sounds included in our study: /u/ represented by the word *Luke* and the sound /ɑ/ represented by the word *strong*. We can attribute this improvement to a few assumptions. The study revealed an improvement for the word *caught*, and previously we had noted that we could not attribute this improvement to the effects of training. However, one assumption could be that if the pronunciation of the word *caught* improved, the students at least could have an unconscious idea of what changes take place within their articulatory apparatus when producing the vowel sound /ɑ/, thus pronouncing the word *strong* better. Another assumption could be the environment where the vowel sound /ɑ/ is found. The fact that the vowel sound /ɑ/ is preceded by the consonant vowel /r/ and the nasal sound /n/ could have been the cause of a better pronunciation of the sound /ɑ/. In either case, we cannot explain the causes of the improvement of pronunciation of this sentence.

The overall results for the three sentences can be seen in Table 4-20. Keeping with the results of the category goodness test, the comprehensibility test also shows that individual students improved in the global production of vowels. For example, the sentence with the highest improvement in both groups was *Luke is strong*. Nine students in the control group improved (60%) and eight students in the experimental group (50%) were rated with a better score. In the control group, three students improved their pronunciation for the sentence *She is always sick*, and four in the experimental group.

79

The results for the sentence *I think he is a fool* reveal that seven students in the control group increased their scores, while five students in the experimental group were rated with a higher score.

Two students in the control group showed improvement at Time 2 in the global production of the vowels in all three sentences (Codes C-48 and C6). It is interesting that student C6 showed an improvement within the one-month period in which the pretest and posttest were administered. Again, these individual differences may be attributed not to training but to other factors which will be discussed in the following chapter.

Table 4-19. Comprehensibility test mean-rating scores on a 5-point scale for all the sentences.

	Control			Experimental		
Stimuli	Pretest	Posttest	Difference	Pretest	Posttest	Difference
Sentence 1	3,0	2,9	-0,1	2,6	2,8	0,2
Sentence 2	2,7	2,90	0,2	2,5	2,5	0,0
Sentence 3	2,7	3,2	0,5	2,6	3,0	0,4
Total			0,6			0,6

4.3 STATISTICAL ANALYSIS

A direct comparison between groups reveals no improvement between Time 1 and Time 2 as an effect of training. However, we carried out a statistical analysis to evaluate the previous statement.

Table 4-20. Mean rating scores in the comprehensibility test for the three sentences on a 5-point scale

Group	Code	SENTENCE 1		SENTENCE 2		SENTENCE 3	
		Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
C1	6	1,5	3,0	2,0	3,0	2,5	3,0
C2	4	4,0	3,5	2,0	1,5	3,0	
C3	7	3,0	2,5	3,5	4,5	2,5	3,0
C4	9	3,5	3,0	2,0	3,0	3,0	3,0
C5	10	3,0	3,0	2,5	3,5	3,5	3,5
C6	11	3,0	3,0	3,0	2,5	2,0	3,0
C7	12	2,5	3,0	2,5	3,0	2,5	2,5
C8	19	3,5	3,0	2,0	4,0	2,5	3,0
C9	23	3,0	3,0	1,5	3,0	2,5	3,0
C10	22	3,5	3,0	3,5	3,0	2,5	2,5
C11	30	3,0	2,0	4,0	4,0	4,0	3,0
C12	41		3,0		4,0		3,0
C13	42	2,5	2,5	2,5	3,5	2,5	3,0
C14	44	2,0	2,0	3,0	3,0	2,0	2,0
C15	48	3,0	3,5	2,5	3,0	2,5	3,0
C16	50	4,0	3,0	4,0	3,0	3,0	3,0
E1	2	2,0	2,0	2,5	2,0	2,0	2,0
E2	15	2,5	2,5	3,5	3,5	3,0	2,0
E3	17	3,0	3,0	2,0	4,0	2,0	3,0
E4	18	3,0	2,5	3,0	3,0	1,5	2,5
E5	14	2,0	3,5	3,5	3,5	3,0	3,0
E6	20	2,5	2,5	3,0	3,0	2,5	2,0
E7	24	2,5	3,0	3,5	3,5	2,5	2,5
E8	25	3,0	3,5	2,5	3,0	3,5	2,5
E9	26	3,0	2,5	1,0	3,5	3,0	2,0
E10	33	3,0	3,0	3,0	3,0	2,0	3,0
E11	36	3,0	3,0	1,0	2,0	2,5	2,5
E12	37	2,5	2,0	2,5	3,0	2,5	2,0
E13	39	3,5	3,0	2,5	3,0	2,5	3,0
E14	40	2,0	3,0	2,5	2,0	2,5	2,5
E15	45	2,5	2,5	3,0	3,5	3,0	3,5
E16	46	2,0	3,0	2,5	2,5	2,0	2,5
Cont.		3,0	2,9	2,7	3,2	2,7	2,9
Exp.		2,6	2,8	2,6	3,0	2,5	2,5

On the production tests, we did 6 ANOVAS for the analysis of the words and 3 ANOVAS for the analysis of the sentences. We created two families of tests using a Bonferroni adjustment to determine the appropriate *p* values.

For the words we divided 0,05 (which is the *p* value generally thought as significant) between the number of words (6 words) and this resulted in a *p* value of 0,008. For the sentences we divided 0,05 between the number of sentences (3 sentences) and the result was a *p* value of 0.017.

The ANOVA designs had one between factor (group: experimental and control) and one within factor (time: before and after). Each ANOVA gives three different *F*-ratios. In this analysis one *F*-ratio indicates the effect of group, another one indicates the effect of time and the last one indicates the interaction of group and time. Tables 4-21 and 4-22 summarize the results for the words *caught, boss, pull, took, mix*, and *Tim* and for the three sentences included in the comprehensibility test.

Table 4-21. Results on the 6 ANOVAS carried out for the words

Words	Group		Time		Interaction	
	F-ratio	*p*-value	*F*-ratio	*p*-value	*F*-ratio	*p*-value
caught	2.635	.1150	8.821	.0058	.583	.4510
boss	.148	.7032	.049	.8268	.265	.6103
pull	.064	.8025	.357	.5546	.357	.5546
took	.038	.8461	4.640	.0394	.321	.5750
mix	7.103	.0123	1.616	.2134	3.860	.0588
Tim	6.809	.0140	1.668	.2063	.363	.5512

For the word *caught*, the results showed no significant main effect of group [F(1, 30) = 2.635, p >.05]. However, there was a significant main effect of time [F(1,30) = 8.821, p < .008]. Moreover, there was not a significant main effect when group and time

interacted [$F(1,30)$ = .583, p > .05]. The rest of the words did not show any significant main effect.

Table 4-22. Results on the 3 ANOVAS carried out for the sentences

Sentences	Group		Time		Interaction	
	F-ratio	*p*-value	*F*-ratio	*p*-value	*F*-ratio	*p*-value
1	2.399	.1323	.011	.9186	1.697	.2029
2	.436	.5142	8.410	.0070	.040	.8423
3	4.023	.0546	1.249	.2732	.694	.4118

Again, when we compared the results of the experimental group with those of the control group in the comprehensibility test, we immediately noticed no improvement between Time 1 and Time 2. The ANOVA designs had one between factor (group: experimental and control) and one within factor (time: before and after).

For the sentence *Luke is strong*, the results showed no significant main effect of group [$F(1, 30)$ = .436, p >.05]. However, there was a significant main effect of time [$F(1,30)$ = 8.410, p < .017]. There was not a significant main effect when group and time interacted [$F(1,30)$ = .040, p > .05]. The rest of the sentences did not show any significant main effect.

The statistical analysis might suggest that there was an improvement in the global production of vowels /u/ and /ɑ/. The vowel /u/ should not have been a problem for the Spanish speakers, but the vowel /ɑ/ represented by the word *strong*, may have been produced with better pronunciation. The implications of the results of the study will be given in Chapter V.

CHAPTER V

CONCLUSIONS

5.1 INTRODUCTION

This chapter presents the conclusions of the thesis. The overview of the results of the study includes the answers to the research questions proposed at the beginning of the project and a comparison between perception and production in the light of the results of the present research. Then, it discusses some of the factors that may have contributed to these results. The chapter concludes laying out the strengths and weaknesses of this specific study and includes future directions in the teaching of pronunciation.

5.2 AN OVERVIEW OF THE PRESENT STUDY

One of the main objectives of the study was to see if there was significant improvement as a result of training Spanish speakers in the perception and production of English vowel sounds /i/-/ɪ/, /u/-/ʊ/, and /ɑ/-/ʌ/ in a regular ESL classroom setting. First, this research was carried out taking into consideration the findings of other studies, and, at the same time, providing more reality to what ESL teachers would face in their actual teaching situations.

The results presented in Chapter III suggest that training Spanish speakers in the perception of the contrasting pairs of vowels /i/-/ɪ/, /u/-/ʊ/, and /ɑ/-/ʌ/ in a regular ESL classroom setting was effective. The experimental group revealed a significant

improvement in the percentage of correct responses between the pretest and posttest (from 60.1% at pretest to 83.3% at posttest).

The experimental group showed a noticeable improvement in the perception of the specific pair of vowel /ɑ/-/ʌ/. The control group also showed an improved performance in the perception of this specific pair. According to the SLM, these two sounds are "new" for Spanish speakers. The high scores achieved by both groups in the pretest suggest that there were students who had already formed a new category for these sounds. These students could improve their perceptual categories with training and those who had not created the new category might have been trained to do so. The percentages of correct responses noticeably increased at posttest for the experimental group, which suggest the effectiveness of training.

The previous findings have a practical implication. Although it is true that recovery of sensitivity is very difficult in adult SL learners, second language teachers have very economical ways to try to remedy this situation .

Chapter IV described the steps carried out to investigate if training had had an effect on production. Two tests were used to assess the students' performance before and after training: a category goodness test and a comprehensibility test. The category goodness test yielded no significant results for the experimental group in relation to the control group. So no correlation could be established between the perceptual learning achieved by the students in a three-week training period and production.

The vowel sound /ɑ/ showed an improved performance in the production of the word *caught*. This improvement is not attributed to training as both groups showed better performances from Time 1 to Time 2. We might assume that the correspondence between spelling and sound could have played an important part in the results of the pronunciation

for this specific word. Another plausible explanation might be that the pronunciation could have been modified through experience. However, there is no evidence to back up the previous assumptions.

The results of the comprehensibility test for the sentences *She is always sick, Luke is strong,* and *I think he is a fool* revealed no evidence that training had influenced the students' global production of the contrasting pairs of vowels /i/-/ɪ/, /u/-/ʊ/, and /ɑ/-/ʌ/. However, there was a significant improvement in the production of the sentence *Luke is strong* in both groups. We might speculate that the sound /u/ represented in this sentence by the word *Luke*, is very similar to the Spanish /u/, so the students would not have a problem pronouncing this word. As for the sound /ɑ/ represented in this sentence by the word *strong*, we have two assumptions: First, the environment where the vowel sound /ɑ/ is, could have led to an improved pronunciation of the word *strong*. The sound is preceded by the consonant sound /r/ and followed by the nasal sound /n/. Second, the vowel sound /ɑ/ achieved the highest percentage scores in the pretest for both the perception and production parts. As the SLM suggests, this is a "new" sound and it is easier for the students to create a new category. Again, these suppositions lack empirical evidence to support a possible relationship between the high scores in the perception of the vowel sound /ɑ/ and the high scores in the production of the word *strong*.

The relationship between perception and production is still very controversial. Our study provides ample evidence that speech perception is modifiable through training. However, it fails to prove its relationship with production. Our assumption is that speech perception might be the starting link in the chain of processes involved in second language acquisition, and that production is based on this initiating link. The need to

address the issue of whether there are perceptual bases for production difficulties in L2 students still exists.

Many researchers have related the causes of improvement in pronunciation not necessarily to perceptual learning but to other individual factors which go beyond the efforts of ESL instructors in the classrooms. We will discuss some of the factors which may have contributed to the results of our research.

5.3 FACTORS WITHIN LEARNERS THAT AFFECT PRONUNCIATION

In Chapter I we mentioned that transfers from L1 to L2 can be a factor which influences pronunciation learning, but these are just some of the factors that affect pronunciation. There are other aspects over which we have little or no control, that can clearly influence L2 learning, and more specifically the learning of pronunciation.

The first factor is age. There is a commonly held belief that there is a strong relationship between L2 pronunciation ability and age and there is very little controversy on whether there is an age-related limit on the mastery of pronunciation. Some researchers support the idea that children under the age of puberty have an excellent chance of having a native like accent if they have continued exposure to a native context and that beyond the age of puberty, SL adult learners will surely keep the foreign accent. That is, a seventy-year old SL learner will be just as successful as a twenty-year old SL learner provided all other learning factors are equal for both (Scovel, 1969; Krashen, 1973). But as we know, puberty can occur as early as 9 years and as late as 16 years. In 1996, Munro, Flege, and Mckay concluded that it is not possible to make a meaningful statement about the relevance of puberty to accent.

We have said that there are researchers who claim that the critical period is not a factor to be taken into consideration in SL pronunciation acquisition. This has been substantiated by the fact that there are accounts about adult learners who have studied a L2 after puberty and have achieved a native like pronunciation. For example, vowel production accuracy may vary among individuals who began their L2 learning in adulthood. In our study, we noticed that three students in the experimental group –Codes 15, 24, and 33 had started studying English at the ages of eight, six, and nine respectively. Their performance production tests at Time 1 and Time 2 can be compared to students who started studying the language later on in life. Student Code No. C-6, for example, started studying English at age of 14. This student showed a very good performance in both the results on the perception part of study and the production part of the study. Even the student who started studying English at the age of 29 (Code E-2) had very good scores in the perception and production of most of the vowels. So, these differences raise the question of whether improvement in vowel production by adults might be caused by the age a person starts learning a second language or by another factor.

When we analyze exposure to a second language, we see that sometimes people live in a foreign country but do not take advantage of exposure to the foreign language. In our study, the students estimated that during 62% of their time they spoke English to a native speaker. This was good, and we know that the school has an "English Only" policy which the students had to follow. However, the average time the majority of the students were exposed to an English speaking environment was three months, and studies have shown that length of residence does make a difference in the accuracy of the production of foreign sounds (Bohn & Flege, 1992). If according to research most of the

improvement in L2 vowel production takes place within a 2-year period (Flege, Bohn, and Jang, 1997), a correlation between length of residence and vowel production accuracy could not be established in our study. The mean length of residence of the subjects was only three months. But there are still other factors to consider.

Individual analysis of students suggests that exposure to the language may in fact influence accuracy in the production of English vowels. The student who improved the most in the perception and production of the vowels included in the study was student Code No. 6. This student started studying English in Colombia at the age of 14 and, prior to coming to Canada, he had been enrolled in an intensive English program in his home country. In the extemporaneous speech sample we collected, he expressed that he was going to speak a lot of English when he went back to his country because his job (an international business company) would demand this from him.

Phonetic ability is the capacity some people have to discriminate foreign sounds; that is; *to have an ear for a foreign language* (Brown, 2001; Nunan, 2000). Our teaching experience tells us that good discriminators are able to benefit from pronunciation drills, while poor discriminators are not. So this is an ability which learners bring to the learning situation and there is little control over it.

The ability to adopt and develop a good pronunciation in a foreign language is related with the extent to which the learners want to identify themselves with the foreign culture. This is known as acculturation. Most of the subjects who participated in the study showed no interest in adapting to this new culture, because the majority were returning to their native countries in six months. They were international students who had come to Canada either to learn English or to improve their English skills.

Generally speaking, the subjects wanted to have a better pronunciation in English and that is why they participated in the study. When the subjects were randomly assigned to the groups, the students in the control group expressed their willingness to be trained after the study was concluded. Once the students did the posttest, we talked to the instructor who had conducted the lessons in the experimental group and he agreed to train the control group. Training was given using the same materials as those used in the experimental group.

This tells us that if motivation is high, the effort to achieve a better pronunciation will be considerably high. Most of the students who had better scores in the production test, reported having specific interest in improving their pronunciation. This was so because, once they went back to their native countries, they were going to be promoted depending on how good their English skills were.

All of the above factors suggest that learners who really want to learn to pronounce a L2 clearly and intelligibly can do it. Teachers can assist them with specific instruction.

As language instructors, we have had the opportunity to teach a second language to students who are real beginners, and have enjoyed watching the progressive improvement of these students' performance. We have noticed that changes do not take place overnight. Most of us cannot believe how well some of these students do at the end of the school year, and some of us look for the most appropriate solutions to the problems we encounter on a daily basis in our ESL classes.

5. 4. STRENGTHS, WEAKNESSES, AND FUTURE DIRECTIONS

Many of the studies on perceptual training of novel sounds have investigated non-native consonant contrasts. Not very many studies have investigated the impact of

perceptual training of non-native vowel contrasts in adults. This thesis provides enough evidence that L2 adults can be trained to perceive L2 vowel contrasts.

All of the studies, where perceptual training is present, have been carried out under laboratory conditions. The empirical data provided in this study indicates that L2 adults can be trained to perceive novel sounds in a regular ESL classroom setting.

The study was very successful in the perceptual training of the pair of vowels /i/-/ɪ/ and /u/-/ʊ/. Because these pairs of vowel sounds were treated as "similar", emphasis was made on the sound not having an exact counterpart for the Spanish speaker. More emphasis was placed on the differences in the position of the tongue and the degree of tenseness than the difference in length. The significant results might suggest that this procedure could be useful when teaching the pronunciation of sounds with similar characteristics.

One limitation of this study was the use of a 9-point scale for the initial rating procedure. This scale proved to be not very effective for the rating task. As we saw, raters tended to agree either when the pronunciation was really good or when it was really poor. The 5-point scale yielded better results in our study.

Previous studies using rating procedures have asked experienced ESL teachers to carry out the rating tasks. In our study, the data provided by the two raters who did not have experience teaching English had to be excluded in the analysis of the results, because their judgments showed inconsistencies in agreement with the other two experienced raters. The findings strongly suggest the use of experienced ESL teachers for similar rating tasks.

When I wrote this thesis, the two most important parts of the teaching learning process never left my mind: the teachers and the students. I think this thesis can help ESL

teachers understand why their students make pronunciation mistakes, and look for the conditions that facilitate the development of their students' ability to differentiate and produce new phonetic categories.

The study proves that there are students out there who do want to improve their pronunciation. Now we know there are practical ways to meet their demands. This thesis is just a wake up call. Now we, ESL teachers have to get to work.

REFERENCES

Asher, J. J. (1977). *Learning Another Language Through Actions: The Complete Teacher's Guidebook.* Los Gatos, CA: Sky Oaks Production.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman and H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224) Cambridge, MA: MIT Press.

Bloomfield, L. (1933). *Language.* New York: Henry Holdt and Company.

Bohn, O. S. & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition, 14,* (pp. 131-158).

Bowen, J. D., Madsen, H., and Hilferty, A. (1985). *TESOL:Techniques and Procedures.* Cambridge MA: Newbury House.

Bradlow, A. R., and Pisoni, D. B. (1997). Training Japanese listeners to identify English /r/ and /l/: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101,* (pp. 2299-2310).

Brown, A. (1988). Functional load and the teaching of pronunciation. In A. Brown (Ed.), *Teaching English Pronunciation* (pp. 211-224).

Brown, H. D. (2001). *Teaching by Principles: An Interactive Approach to Language Pedagogy.* (Second edition). White Plains, NY: Addison Wesley Longman.

Catford, J. C. (1987). Phonetics and the Teaching of Pronunciation: A Systemic Description of English Phonology. In J. Morley (Ed.), *Current Perspectives on Pronunciation: Practices Anchore in Theory.* (pp. 87-100). TESOL.

Cattegno, C. (1972) *Teaching Foreign Languages in Schools: The Silent Way,* (2nd edn.) New York: Educational Solutions.

Chomsky, N. (1957). *Syntactic Structures.* The Hague: Mouton.

Celce-Murcia, M., Brinton, D. M., and Goodwin, J. M. (1996). *Teaching Pronunciation: A reference for Teachers of English to Speakers of Other Languages.* Cambridge: Cambridge University Press.

Corder, S. P. (1974). Idiosyncratic dialects and error analysis. In Schumann & N. Stenson (Eds.) *New frontiers in second language learning.* Rowley, MA: Newbury House.

Curran, C. A. (1972). *Counseling-Learning: A Whole-Person Model for Education.* New York: Grune and Stratton.

Derwing, T., Munro, M. and Wiebe, G. (1998). Pronunciation instruction for "fossilized" learners, can it help?. *Applied Language Learning, Vol. 8, No. 2,* (pp. 217-235).

Derwing, T., Munro, M. and Wiebe, G. (1998). Evidence in Favor of a Broad Framework for Pronunciation Instruction. *Language Learning, 48,* (pp. 393-410)

Derwing, T., and Munro, M. (1997). Accent, intelligibility, and comprehensibility: evidence from four L1s. *Studies in Second Language Acquisition. No. 19,* (pp. 1-16).

Eimas, P. D. (1975). Developmental studies in speech perception. In L.B. Cohen and P. Salapatek, (Eds.) *Infant Perception: From Sensation to Cognition, Vol. 2,* (pp. 193-231). Academic Press.

Flege, J. E., Mackay, I. R.,and Meador, D. (1999). Native Italian speaker's perception and production of English vowels. *Journal of the Acoustical Society of America, 106,* (pp. 2973-2987).

Flege, J. E., Bohn, 0.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25,* (pp. 437-470).

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America, 93,* (pp. 1589-1608).

Flege, J. E. (1991a) Perception and production: The relevance of phonetic input to L2 phonological learning, In T. Heubner, & C. Ferguson (Eds.) *Crosscurrents in second language acquisition and linguistic theory.* (pp. 249-284). Philadelphia: John Benjamins.

Flege, J. E. (1991b). The interlingual identification of Spanish and English vowels: Orthographic evidence. *Quarterly Journal of Experimental Psychology, 43,* (pp. 701-731).

Flege, J. E., and Wang, C. (1989). Native-Language phonotactic constraints affect how well Chinese subjects perceive the word final English /t/-/d/ contrast. *Journal of Phonetics 17,* (pp. 299-315).

Flege, J. E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders: A review* (pp. 224-401), Norwood, NJ: Ablex Publishing.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics, 15,* (pp. 47-65).

Flege, J. E. (1981). The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly, 15,* (pp. 443-455).

Gall, J. P., Gall, M. D., Borg W. (1999). *Applying Educational Research: A practical guide* (4ᵗʰ Edition). New York: NY, Addison and Wesley Longman.

Garcia-Perez, G. M. (1999) *Training Spanish Speakers in the Perception and Production of Vowel Sounds* (unpublished booklet).

Garcia-Perez, G. M. (1997). Training Korean speakers in the perception of /r/and /l/. Unpublished paper.

Goodwin, J. (2001). Teaching Pronunciation. In Marianne Celce-Murcia (ed.) *Teaching English as a Second or Foreign Language*. (Third edition, pp. 117-138). Boston, MA: Heinle & Heinle.

Goto, H. (1971). Auditory perception by normal Japanese adults of the "L' and "R." *Neuropsychologia, 9*, (pp. 317-323).

Ioup, G. (1995). Evaluating the need for input enhancement in post-critical period language acquisition. In D. Singleton and Z. Lengyel (Eds.), *The Age Factor in Second Language Acquisition* (pp. 95-123). Clevedon: Multilingual Matters.

Ioup, G. & Weinberger, S. H. (1987). *Interlanguage phonology: The acquisition of a second language sound system*. New York: Newbury House.

Kelly, L. G. (1969). *25 Centuries of Language Teaching*. Rowley, MA: Newbury House.

Krashen, S. D. (1973). Lateralization, language learning, and the critical period: Some new evidence. *Language Learning, 23*, (pp. 63-74).

Liberman, A. L., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, (pp. 431-461).

Lightbown, P. M., and Spada, N. (1999). *How Languages are Learned*. Oxford: Oxford University Press.

Logan, J. S., Lively, S. E., and Pisoni, D.B. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *Journal of the Acoustical Society of America, 89*, (pp. 874-886).

Lozanov, G. (1978). *Suggestology and Outlines of Suggestopedy*. New York: Gordon and Breach.

Munro, M. (1999). *Audio Stimulus Playback 4*: (computer software program). Department of Linguistics, Simon Fraser University.

Munro, M. (1999). *Audio Experiment Generator 4*: (computer software program). Department of Linguistics, Simon Fraser University.

Munro, M., Flege, J. E., and Mackay, I. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics 17*, (pp. 313-334).

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: acoustic measurements and accentedness ratings. *Language and Speech, 36*, (pp. 39-66).

Nunan, D. (2000). Language Teaching Methodology: A textbook for language teachers. (Second edition). London: Longman.

Odisho, E. Y. (1990). Teaching English Vowels to Adult Hispanics. Paper presented at the 17th Annual Illinois Convention of TESOL/Bilingual Education. Chicago. Illinois.

O'Grady, W. and Archibald, J. (2000) Contemporary Linguistic Analysis. An Introduction. Fourth Edition. Pearson Education Canada Inc. Toronto

Perlmutter, M. (1989). Intelligibility rating of L2 speech pre- and postintervention. *Perceptual and Motor Skills, 68*, (pp. 515-521).

Pisoni, D. B. (1978). Speech Perception. In W. K. Estes (Eds.) *Handbook of learning and cognitive processes* (pp. 167-233). Hillsdale, NJ. Erlbaum.

Pisoni, D. B., Lively, S. E., and Logan, J. S. (1994). Perceptual Learning of Nonnative Speech Contrasts: Implications for Theories of Speech Perception. In J. Goodman and H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 121-166). Cambridge, MA: MIT Press.

Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance. No. 8*, (pp. 297-314).

Richards, J. C. (1971). A non contrastive approach to error analysis. *English Language Teaching Journal, 25*, (pp. 204-219).

Richards, J. C., and Rodgers, T. S. (1986). Approaches and Methods in Language Teaching. Cambridge: Cambridge University Press.

Rochet, B. L (1995). Perception and production of L2 speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross language research* (pp. 379-410). Timonium, MD: York Press.

Rogers, H. (1991). *Theoretical and Practical Phonetics*. Coop Clark Pitman Ltd. A Longman Company, Toronto.

Schmidt, A. M. (1996). Cross language identification of consonants. Part I. Korean perception of English. *Journal of the Acoustical Society of America, 99*, (pp. 3201-3211).

Scovel, T. (1969). Foreign Accent. Language acquisition and cerebral dominance. *Language Learning, 19*, (pp. 245-254).

Selinger, H. W. & Shohamy, E. (1990). Second Language Research Methods. Oxford: Oxford University Press.

Selinker, L. (1972). Interlanguage. *IRAL 10* (pp. 209-231).

Sheldon, A. and Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3,* (pp. 243-261).

Singleton, D. (1995). Introduction: A critical look at the Critical Period Hypothesis in second language acquisition research. In D. Singleton and Z. Lengyel (Eds.), *The Age Factor in Second Language Acquisition* (pp. 1-29). Clevedon: Multilingual Matters.

Stockwell, R. P. & Bowen, J. (1965). The sounds of English and Spanish. Chicago: The University of Chicago Press.

Strange, W. and Jenkins, J. (1978). Role of linguistic experience in the perception of speech. In R. D. Walk and H. L. Pick (eds.) *Perception and experience*. New York: Plenum Press.

Streeter, L. A. (1976). Language perception of two-month old infants shows effects of both innate mechanisms and experience. *Nature, 259*, (pp. 39-41).

Wang, X. (2002). *Training Mandarin and Cantonese speakers to identify English vowel contrasts: Long Term retention and effects on production.* Unpublished Ph. D. thesis, Simon Fraser University, Burnaby, British Columbia.

Wang, X. (1997). *The acquisition of English Vowels by Mandarin ESL learners: A study of production and perception.* Unpublished master's thesis, Simon Fraser University, Burnaby, British Columbia.

Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly, 4* (2), (pp. 123-130).

Werker, J. F. (1989). Becoming a Native Listener. *American Scientist, 77*, (pp. 54-59).

Widdowson, H. (1978). Notional-Functional Syllabuses:1978 Part Four. In C.H. Blatchford and J. Schachter, eds. On *TESOL '78: EFL*

Policies, Programs, Practices. Washington, D.C.: TESOL, (pp. 33-35).

Williams, L. (1979). The modification of speech perception and production in second language learning. *Perception & Psycholinguistics 26*, (pp. 95-104).

SIMON FRASER UNIVERSITY

OFFICE OF RESEARCH ETHICS

BURNABY, BRITISH COLUMBIA
CANADA V5A 1S6
Telephone: 604-291-3447
FAX: 604-268-6785

June 17, 2003

Ms. Grisel Maria Garcia-Perez
Graduate Student
Department of Linguistics
Simon Fraser University

Dear Ms Garcia-Perez:

Re: Training Spanish Speakers in the Perception
and Production of English Vowels

The above-titled ethics application has been granted approval by the
Simon Fraser Research Ethics Board, in accordance with Policy R 20.01,
"Ethics Review of Research Involving Human Subjects".

Sincerely,

Dr. Hal Weinberg, Director
Office of Research Ethics

---

* For inclusion in thesis/dissertation/extended essays/research project report, as submitted to the university library in
fulfillment of final requirements for graduation. Note: correct page number required.

Are you **18 or older**?

Is **Spanish** your first language?

Do you want to **improve your pronunciation** in English?

Would you like to participate in a study of pronunciation learning?
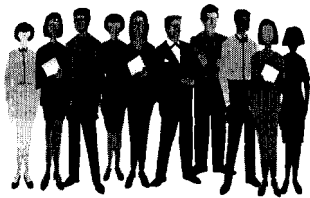
The Spanish speakers Club is waiting for you!!!

For more information, contact Dale.

Appendix C. Language background questionnaire given to all participants at the beginning of the study.

Language Background Questionnaire – Spanish Speakers

Date:_____   Participant Code:_____

1. Is your hearing normal?   Yes____ No____

2. What is your first language?   _____

3. Where were you born? (city, country)   _____,_____

4. What is your date of birth?   ____/____/____
                                                        DD   MM   YY

5. When did you come to Canada?   ____/____/____
                                                        DD   MM   YY

6. Did you study English before coming to Canada?   _____
   If so, for how long?

7. At what age did you start studying English?   _____

8. How long have you been taking ESL courses   _____
in Canada?

9. Have you taken any courses on English pronunciation? If so, give details.

_____

10. How important is it for you to improve your pronunciation?

Not very   1   2   3   4   5   6   7 ....**Very**
important   **important**

11. How often does your pronunciation cause difficulty when you are talking to others?

never   1   2   3   4   5   6   7   **very often**

12. How often do you speak English...

never	1	2	3	4	5	6	7 **very often**

a) at home?   1   2   3   4   5   6   7

b) with your   1   2   3   4   5   6   7
friends?

c) with a Native   1   2   3   4   5   6   7
speaker of
English?

d) with a non-   1   2   3   4   5   6   7
native speaker
of English?

13. List all the places where you have lived for more than 6 months

Place   How long

_____   _____

_____   _____

_____   _____

Appendix D. Consent form given to the subjects before the experiment started.

SIMON FRASER UNIVERSITY

<u>INFORMED CONSENT BY SUBJECTS TO PARTICIPATE
IN A RESEARCH PROJECT OR EXPERIMENT</u>

The University and those conducting this project subscribe to the ethical conduct of research and to the protection at all times of the interests, comfort, and safety of subjects. This form and the information it contains are given to you for your own protection and full understanding of the procedures. Your signature on this form will signify that you have received a document which describes the procedures, possible risks, and benefits of this research project, that you have received an adequate opportunity to consider the information in the document, and that you voluntarily agree to participate in the project.

Although your name will be recorded, it is not required for the analysis of any data collected in the study. It will be kept separate from the data collected. Materials will be held in a secure location, and will be destroyed after the completion of the study.

Having being asked by Grisel María García-Pérez of the Linguistics Department of Simon Fraser University to participate in a research project experiment, I have read the procedures specified in the information for participants sheet.

I understand the procedures to be used in this experiment.

I understand that I may withdraw my participation in this experiment at any time.

I also understand that I may register any complaint I might have about the experiment with the researcher named above or with Dr. Paul McFetridge, Chair of the Linguistics Department of Simon Fraser University (Phone: 291-4798).

I may obtain copies of the results of this study, upon its completion, by contacting Grisel María García-Pérez, Linguistics Department, Simon Fraser University, Burnaby, BC, V5A-1S6.

I have been informed that the research material will be held confidential by the Principal Investigator.

Page 2.

Please read the "Information Sheet for Participants" before completing the section below.

I agree to participate by

a) completing a language background questionnaire
b) participating in a 12 hour training period, if applicable
c) being tested before and after the training period on the perception and production of English vowels

as described in the document referred to above, during the time period February, 1999 to January, 2000 at Van West College, 1215 West Broadway, Vancouver.

NAME (please type or print legibly):_____

ADDRESS:_____

_____

SIGNATURE: _____**WITNESS:**_____

DATE: _____

ONCE SIGNED, A COPY OF THIS CONSENT FORM AND A SUBJECT FEEDBACK FORM SHOULD BE PROVIDED TO THE SUBJECT.

Appendix E-1: Pre-posttest used for the perception part of the study

Pre-test (Van West College)
(Training Spanish speakers in the perception and production of English vowels)
Student code:_____

d) The following pairs of words are identical except for the vowel sound. Circle the word that you hear.

a) cheap	chip	m) cuff	cough
b) feet	fit	n) cut	caught
c) he's	his	o) bus	boss
d) leave	live	p) lung	long
e) lung	long	q) fool	full
f) cut	caught	r) look	Luke
g) bus	boss	s) kook	cook
h) cuff	cough	t) who'd	hood
i) fool	full	u) cheap	chip
j) who'd	hood	v) leave	live
k) look	Luke	w) he's	his
l) kook	cook	x) feet	fit

Appendix E-2. Pre-posttest used for the production part of the study

e) Read the following sentences: (orally)

1. Jean was released three weeks ago.
2. She seems to be extremely pleased.
3. Mix it with a stick.
4. Jim caught a big fish.
5. It is easy for Tim to miss the meeting.
6. He cut her cuffs on the bus.
7. It was fun to fill the cup on the truck.
8. That dog has a long collar.
9. I wonder why my socks are lost.
10. Sometimes my boss drives me nuts.
11. He looks like a good cook.
12. Who took my book?
13. The pool is too cool.
14. I knew this juice was going to turn blue.
15. Don't pull out my tooth!

f) Read the following paragraph: (orally)

Paul, Luke, Jean and Tim live together. They are best buddies. Paul says he is rich. He likes good food and talks a lot. Luke is strong. He has blue eyes and red cheeks. Jean is not fun at all. She is always sick. Tim is the one I know the least. I think he is a fool. I could tell you more about them all, but I wonder if I should.

g) Answer the following questions: (orally)

When you go back to your country, will you be speaking English …

a) a lot? (Explain)

b) just a little? (Explain)

UNIT I: The vowel sounds /i/ and /ɪ/

Lesson # 1: The vowel sound /i/

1. **Articulation.** How is the sound /i/ produced?

The vowel sound /i/ is produced by placing the tongue higher and more forward than for the other vowels. The tongue is tense and the lips are more spread and less open than for any of the other vowel sounds. (Remember the Spanish word *¡mira!)*. The vowel sound /i/ is relatively long.

2. **Pronunciation practice**. Listen and repeat. Pay attention to the way the English vowel sound /i/ is produced.

beach	greet	people
bean	heal	queen
beat	heat	reach
breeches	jeep	real
cheap	kneel	receive
cheat	lead	seal
cheek	lean	seed
deal	least	seek
dean	leave	she
deep	machine	taxi
each	mean	teal
ease	meat	team
eat	meet	these
feel	neat	weak
feet	peach	week
freeze	peak	
green	peel	

3. **Listening practice**. Listen to the following sentences. Circle the words which are pronounced with the vowel sound /i/. Practice reading the sentences out loud. Exaggerate the /i/ sound just a little bit.

a) Rita really needed to eat meat.
b) These books are easy.
c) Even Jean's teacher was speechless.
d) He is leaving next week.

e) She was really sleeping when I started reading.

4. **Dictation**. Write the following sentences. Circle the words which are pronounced with the vowel sound /i/. Practice reading the sentences out loud. Exaggerate the /i/ sound just a little bit.

a) Steve believes I like to eat cheese.
b) Peter is pleased to speak to the teacher.
c) Her team played three weeks ago.
d) He teaches reading every evening.
e) It is extremely easy to meet you at the beach.

5. **Spelling awareness**.
 Note that the sound /i/ is usually written with the letters *e*, as in b*e*; *ee*, as in b*ee*n; *ea*, as in b*ea*t; or *ie*, as in p*ie*ce. Other spellings are *y*, as in alread*y*; *ey*, as in vall*ey*; *ei*, as in n*ei*ther; *i*, as in tax*i*; *i...e*, as in mach*i*n*e*; and *eo*, as in p*eo*ple. Look at the list of words in exercise two, and find examples for each of the spellings listed above.

6. **Conversation Practice**. Find a partner. Ask your partner:

a) the time he/she is leaving today
b) if he/she likes to eat meat, cheese, beans...
c) if he/she likes to drink coffee or tea
d) if he/she is going to watch a movie tonight
e) the time he/she is going to sleep

7. **More Practice.** Reread the list of words in Exercise 2. Look up the meaning of the words you do not know. Read these words out loud. When you read them, try to look at yourself in a mirror so that you see the position of your lips.

Lesson # 2:   The vowel sound /ɪ/

1. Articulation. How is the vowel sound /ɪ/ produced?

The vowel sound /ɪ/ is produced with the front part of the tongue. The tongue is more relaxed and the sound is shorter than vowel sound /i/. The lips are less widely spread and more open than for the /i/ sound. (Remember the Spanish word *sí*, as pronounced when you are listening to someone telling you a story, and do not believe what this person is saying).

2. Pronunciation practice. Listen and repeat. Pay attention to the way the English vowel sound /ɪ/ is produced.

bitch	gyp	mill	Sick
bin	his	Min	sin
bit	hill	mitt	sip
britches	hip	knit	shill
chip	hit	pitch	shin
chit	it	pick	skit
chick	is	pill	spill
dill	ill	Quinn	till
din	kill	rich	Tim
dim	kin	rill	tin
dip	kid	rim	will
itch	lick	rip	wick
fill	Lynn	rid	gin
fit	lip	Rick	
frizz	list	skim	
grin	live	sill	
grit	Liz	Sid	

109

3. Listening practice. Listen to the following sentences. Circle the words that are pronounced with the vowel sound /ɪ/. Practice reading the sentences out loud. Exaggerate the /ɪ/ sound just a little bit.

 a) Six editions have been printed
 b) Miss Sylvia Miller lives in Michigan.
 c) Mix it with a stick.
 d) I will be here till mid-winter.
 e) Give him a little milk instead.

4. Dictation. Write the following sentences. Circle the words that are pronounced with the vowel sound /ɪ/. Practice reading the sentences out loud. Exaggerate the /ɪ/ sound just a little bit.

 a) He is very busy.
 b) It is simple to print the original script.
 c) This is a different fish.
 d) The winning ticket is inside this building.
 e) These are simple reading skills.

5. Spelling Awareness. Note that the sound /ɪ/ is usually written with the letter *i*. Other spellings are *y* , as in s*y*stem ; *u*, as in b*u*siness; *o*, as in w*o*men; *e*, as in *E*nglish and pr*e*tty; and *ui*, as in b*ui*lding. Look at the list of words in Exercise 2, and find examples for each of the spellings listed above.

6. Conversation Practice. Find a partner. Ask your partner:

 a) What he/she thinks about children
 b) When he/she will finish studying English
 c) If he/she has considered visiting England in the near future
 d) If he/she is rich
 e) What he/she does when he/she is sick

7. More Practice. Reread the list of words in Exercise 2. Look up the meaning of the words you do not know. Read these words out loud. When you read them, try to look at yourself in a mirror so that you see the position of your lips.

Lesson # 3: The vowel sounds /i/ and /ɪ/

1. **Articulation.** How are the vowel sounds /i/ and /ɪ/ produced?

Vowel /i/ is relatively long and tense. The lips are spread and less open than for any of the vowel sounds. (Remember the Spanish word ¡mira!). Vowel /ɪ/ is shorter and more relaxed than the vowel /i/. The lips are less widely spread and more open than for the /i/ sound. (Remember the Spanish word sí, as pronounced when you are listening to someone telling you a story, and you do not believe what this person is saying)

2. **Pronunciation practice.** Listen and repeat. Pay attention to the way the English vowel sounds /iː/ and /ɪ/ are produced.

/i/	/ɪ/	/i/	/ɪ/
1. beach	bitch	32. keen	kin
2. bean	bin	33. keyed	kid
3. beat	bit	34. kneel	nil
4. breeches	britches	35. lead	lid
5. cheap	chip	36. leak	lick
6. cheat	chit	37. lean	Lynn
7. cheek	chick	38. leap	lip
8. deal	dill	39. least	list
9. dean	din	40. leave	live
10. deem	dim	41. lees	Liz
11. deep	dip	42. meal	mill
12. each	itch	43. mean	Min
13. ease	is	44. meat/meet	mitt
14. eat	it	45. neat	knit
15. eel	ill	46. peach	pitch
16. feast	fist	47. peak	pick
17. feel	fill	48. peal/peel	pill
18. fees	fizz	49. peat/Pete	pit
19. feet	fit	50. queen	Quinn
20. fleet	flit	51. reach	rich
21. freeze	frizz	52. real	rill
22. greased	grist	53. ream	rim
23. green	grin	54. reap	rip
24. greet	grit	55. reed	rid
25. he's	his	56. reek	Rick
26. heal	hill	57. scheme	skim
27. heap	hip	58. seal	sill
28. heat	hit	59. seed	Sid
29. Jean/gene	gin	60. seek	sick
30. jeep	gyp	61. seen	sin
31. keel	kill	62. seep	sip

111

63. she'll	shill	72. team	Tim
64. sheen	shin	73. teen	tin
65. skeet	skit	74. tweet	twit
66. sleet	slit	75. we'll	will
67. sneaker	snicker	76. weak	wick
68. spiel	spill	77. wean	win
69. steeple	stipple	78. week	wick
70. teak	tick	79. wheeze	whiz
71. teal	till		

3. **Reading Practice**. Read the following sentences out loud. Exaggerate the /i/ and /ɪ/ sounds just a little bit.

   a) Tim thinks that she is not sick.
   b) I will visit Jean next week.
   c) If he gives me a kiss, I will feel extremely pleased.
   d) Rick likes to pick on me.
   e) We decided to leave because we had a meeting.
   f) Leave him alone! He is very busy.

4. **Conversation Practice**. Find someone who...

   a) speaks English on a regular basis in his/her native country
   b) never speaks English in his/her native country
   c) will be visiting England in the near future
   d) has his own business back home
   e) is always very busy
   f) is never sick
   g) feels extremely pleased with this course

5. **Quiz**. Listen to the following sentences. Circle the word that you hear.

   a) Spell the word (seek- **sick**)
   b) I think this is the other (sheep- **ship**)
   c) She was (sleeping- **slipping**) when I found her.
   d) I want to (**feel**- fill) that bag.
   e) They lost the (**lead**- lid)

Appendix G. Three-rater regrouped mean scores on a 9-point scale for the words *cut, boss, bus, pool, pull, too, took, weeks, mix, seems*, and *Tim* at pre-and posttest, and for the word *caught* at posttest.

Stimulus: *caught*						Time: posttest
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	3	6	3	4,0	0
C2	4	1	6	3	3,3	0
C3	7	3	4	3	3,3	0
C4	9	1	5	4	3,3	0
C5	10	2	2	3	2,3	1
C6	11	1	1	5	2,3	0
C7	12	1	1	2	1,3	1
C8	19	2	1	1	1,3	1
C9	23	8	4	6	6,0	0
C10	22	1	1	4	2,0	0
C11	30	2	2	4	2,7	0
C12	41	1	3	1	1,7	1
C13	42	7	6	1	4,7	0
C14	44	4	7	1	4,0	0
C15	48	3	6	7	5,3	0
C16	50	7	6	4	5,7	0
E1	2	9	8	7	8,0	1
E2	15	1	2	7	3,3	0
E3	17	1	8	8	5,7	0
E4	18	6	3	2	3,7	0
E5	14	3	4	1	2,7	0
E6	20	8	4	4	5,3	0
E7	24	7	2	2	3,7	0
E8	25	2	5	3	3,3	0
E9	26	9	5	6	6,7	0
E10	33	8	2	4	4,7	0
E11	36	4	8	5	5,7	0
E12	37	1	6	8	5,0	0
E13	39	1	2	1	1,3	1
E14	40	7	8	7	7,3	1
E15	45	2	3	8	4,3	0
E16	46	1	7	6	4,7	0
Mean C		2,9	3,8	3,3	3,3	4
Mean E		4,4	4,8	4,9	4,7	3
				Total C	33%	25%
				Total E	56%	19%

113

Stimulus: *cut*				Time: pretest	
Group	**Code**	**Rater 1**	**Rater 2**	**Rater 3**	**Mean**	**Items used**
C1	6	3	3	8	4,7	0
C2	4	2	1	1	1,3	1
C3	7	6	2	5	4,3	0
C4	9	7	5	6	6,0	0
C5	10	9	8	8	8,3	1
C6	11	2	1	1	1,3	1
C7	12	1	2	3	2,0	1
C8	19	2	2	1	1,7	1
C9	23	7	7	9	7,7	1
C10	22	7	6	4	5,7	0
C11	30	1	3	2	2,0	1
C12	41	1	2	4	2,3	0
C13	42	1	3	4	2,7	0
C14	44	1	1	3	1,7	1
C15	48	1	3	1	1,7	1
C16	50	3	5	3	3,7	0
E1	2	7	1	9	5,7	0
E2	15	1	2	3	2,0	1
E3	17	2	2	2	2,0	1
E4	18	8	3	5	5,3	0
E5	14	4	6	8	6,0	0
E6	20	1	1	4	2,0	0
E7	24	1	3	1	1,7	1
E8	25	2	3	4	3,0	0
E9	26	9	3	9	7,0	0
E10	33	9	6	7	7,3	0
E11	36	5	2	3	3,3	0
E12	37	1	1	1	1,0	1
E13	39	2	8	8	6,0	0
E14	40	6	7	7	6,7	0
E15	45	8	6	4	6,0	0
E16	46	1	2	1	1,3	1
		3,4	3,4	3,9	3,6	9
		4,2	3,5	4,8	4,1	5
				Total C	40%	56%
				Total E	46%	31%
native	F1	1	1	9	3,7	0
native	F2	1	1	9	3,7	0
native	M1	6	2	9	5,7	0
native	M2	9	3	9	7,0	0
				Mean	5,0	0
				Total		0%

Stimulus: *cut*				Time: posttest	
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	7	4	2	4,3	0
C2	4	3	2	1	2,0	1
C3	7	7	4	2	4,3	0
C4	9	8	7	8	7,7	1
C5	10	3	5	9	5,7	0
C6	11	1	1	2	1,3	1
C7	12	7	9	9	8,3	1
C8	19	1	1	1	1,0	1
C9	23	2	2	4	2,7	0
C10	22	7	5	8	6,7	0
C11	30	1	2	1	1,3	1
C12	41	2	2	2	2,0	1
C13	42	1	2	1	1,3	1
C14	44	8	8	9	8,3	1
C15	48	1	3	2	2,0	1
C16	50	1	2	2	1,7	1
E1	2	9	8	9	8,7	1
E2	15	1	2	1	1,3	1
E3	17	1	4	1	2,0	0
E4	18	9	6	5	6,7	0
E5	14	6	5	7	6,0	0
E6	20	9	9	7	8,3	1
E7	24	7	3	7	5,7	0
E8	25	1	1	2	1,3	1
E9	26	6	6	8	6,7	0
E10	33	2	3	3	2,7	1
E11	36	6	3	1	3,3	0
E12	37	1	5	3	3,0	0
E13	39	7	8	8	7,7	1
E14	40	8	3	8	6,3	0
E15	45	7	4	5	5,3	0
E16	46	1	1	3	1,7	1
Mean C		3,8	3,7	3,9	3,8	11
Mean E		5,1	4,4	4,9	4,8	7
				Total C	42%	69%
				Total E	53%	44%

Stimulus: *boss*				Time: pretest		
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	7	3	2	4,0	0
C2	4	7	6	2	5,0	0
C3	7	6	4	3	4,3	0
C4	9	6	3	2	3,7	0
C5	10	2	6	6	4,7	0
C6	11	7	5	4	5,3	0
C7	12	6	8	9	7,7	0
C8	19	1	4	4	3,0	0
C9	23	7	3	3	4,3	0
C10	22	2	4	8	4,7	0
C11	30	1	3	2	2,0	1
C12	41	3	5	9	5,7	0
C13	42	1	1	5	2,3	0
C14	44	3	6	3	4,0	0
C15	48	9	8	6	7,7	0
C16	50	6	6	7	6,3	0
E1	2	6	7	5	6,0	0
E2	15	4	7	9	6,7	0
E3	17	4	3	4	3,7	0
E4	18	2	4	3	3,0	0
E5	14	7	4	7	6,0	0
E6	20	2	5	6	4,3	0
E7	24	6	7	7	6,7	0
E8	25	4	7	8	6,3	0
E9	26	6	5	8	6,3	0
E10	33	4	1	5	3,3	0
E11	36	6	6	8	6,7	0
E12	37	2	1	4	2,3	0
E13	39	5	7	7	6,3	0
E14	40	3	4	7	4,7	0
E15	45	5	9	6	6,7	0
E16	46	3	3	8	4,7	0
Mean C		4,6	4,7	4,7	4,7	1
Mean E		4,3	5,0	6,4	5,2	0
				Total C	52%	6%
				Total E	58%	0%
native	F1	3	5	6	4,7	0
native	F2	9	6	9	8,0	0
native	M1	8	6	9	7,7	0
native	M2	9	7	9	8,3	1
				Mean	7,2	1
				Total		25%

Stimulus: *boss*				Time: posttest		
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	3	7	7	5,7	0
C2	4	7	2	3	4,0	0
C3	7	8	6	7	7,0	0
C4	9	2	2	4	2,7	0
C5	10	4	6	3	4,3	0
C6	11	9	2	5	5,3	0
C7	12	8	6	9	7,7	0
C8	19	2	3	6	3,7	0
C9	23	4	7	8	6,3	0
C10	22	2	4	7	4,3	0
C11	30	4	4	6	4,7	1
C12	41	8	6	9	7,7	0
C13	42	3	8	4	5,0	0
C14	44	2	4	6	4,0	0
C15	48	5	8	7	6,7	0
C16	50	4	6	3	4,3	0
E1	2	2	9	4	5,0	0
E2	15	6	4	9	6,3	0
E3	17	4	5	8	5,7	0
E4	18	3	7	2	4,0	0
E5	14	8	6	6	6,7	0
E6	20	2	3	4	3,0	0
E7	24	6	7	2	5,0	0
E8	25	7	3	9	6,3	0
E9	26	6	2	5	4,3	0
E10	33	6	3	3	4,0	0
E11	36	3	8	8	6,3	0
E12	37	1	1	1	1,0	1
E13	39	5	7	9	7,0	0
E14	40	5	2	7	4,7	0
E15	45	5	7	3	5,0	0
E16	46	3	3	7	4,3	0
Mean C		4,7	5,1	5,9	5,2	1
Mean E		4,5	4,8	5,4	4,9	1
				Total C	58%	6%
				Total E	54%	6%

117

Stimulus: *bus*				Time: pretest		
Group	**Code**	**Rater 1**	**Rater 2**	**Rater 3**	**Mean**	**Items used**
C1	6	6	4	5	5,0	1
C2	4	9	5	9	7,7	0
C3	7	4	1	1	2,0	0
C4	9	8	8	9	8,3	1
C5	10	4	2	9	5,0	0
C6	11	9	8	9	8,7	1
C7	12	6	4	9	6,3	0
C8	19	7	4	8	6,3	0
C9	23	7	7	9	7,7	1
C10	22	7	4	6	5,7	0
C11	30	8	5	8	7,0	0
C12	41	9	5	6	6,7	0
C13	42	2	1	8	3,7	0
C14	44	6	7	9	7,3	0
C15	48	5	3	6	4,7	0
C16	50	9	8	9	8,7	1
E1	2	7	2	9	6,0	0
E2	15	9	7	9	8,3	1
E3	17	5	3	7	5,0	0
E4	18	6	8	6	6,7	0
E5	14	9	8	9	8,7	1
E6	20	4	4	4	4,0	1
E7	24	1	1	8	3,3	0
E8	25	6	8	9	7,7	0
E9	26	2	3	9	4,7	0
E10	33	9	9	9	9,0	1
E11	36	8	6	8	7,3	0
E12	37	1	3	6	3,3	0
E13	39	8	7	9	8,0	1
E14	40	4	6	9	6,3	0
E15	45	9	8	9	8,7	1
E16	46	3	3	2	2,7	1
Mean C		6,6	4,8	7,5	6,3	5
Mean E		5,7	5,4	7,6	6,2	7
				Total C	70%	31%
				Total E	69%	44%
native	F1	9	8	9	8,7	1
native	F2	7	5	9	7,0	0
native	M1	7	6	9	7,3	0
native	M2	8	8	8	8,0	1
				Mean	7,8	2
				Total		50%

Stimulus: *bus*				Time: posttest		
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	8	8	9	8,3	1
C2	4	6	1	9	5,3	0
C3	7	7	6	6	6,3	0
C4	9	9	5	7	7,0	0
C5	10	7	8	9	8,0	1
C6	11	9	7	9	8,3	1
C7	12	7	8	7	7,3	1
C8	19	2	3	8	4,3	0
C9	23	5	9	7	7,0	0
C10	22	4	4	4	4,0	1
C11	30	5	2	3	3,3	0
C12	41	8	6	7	7,0	0
C13	42	2	4	8	4,7	0
C14	44	8	7	9	8,0	1
C15	48	3	3	3	3,0	1
C16	50	5	7	9	7,0	0
E1	2	8	5	9	7,3	0
E2	15	6	6	8	6,7	0
E3	17	4	5	6	5,0	1
E4	18	5	5	7	5,7	0
E5	14	8	8	9	8,3	1
E6	20	8	6	9	7,7	0
E7	24	7	7	9	7,7	1
E8	25	3	1	6	3,3	0
E9	26	5	7	9	7,0	0
E10	33	7	3	9	6,3	0
E11	36	6	5	8	6,3	0
E12	37	6	2	3	3,7	0
E13	39	6	4	9	6,3	0
E14	40	6	8	9	7,7	0
E15	45	8	6	8	7,3	0
E16	46	9	6	9	8,0	0
Mean C		5,9	5,5	7,1	6,2	7
Mean E		6,4	5,3	7,9	6,5	3
				Total C	69%	44%
				Total E	72%	19%

119

Stimulus: *pool*				Time: pretest		
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	7	5	8	6,7	0
C2	4	2	5	6	4,3	0
C3	7	1	4	1	2,0	0
C4	9	2	3	6	3,7	0
C5	10	1	4	8	4,3	0
C6	11	2	2	5	3,0	0
C7	12	2	2	8	4,0	0
C8	19	1	5	6	4,0	0
C9	23	3	8	7	6,0	0
C10	22	2	6	6	4,7	0
C11	30	1	4	2	2,3	0
C12	41	3	5	6	4,7	0
C13	42	1	2	4	2,3	0
C14	44	2	3	6	3,7	0
C15	48	1	6	4	3,7	0
C16	50	2	3	3	2,7	1
E1	2	1	5	1	2,3	0
E2	15	1	1	1	1,0	1
E3	17	2	2	6	3,3	0
E4	18	2	4	7	4,3	0
E5	14	1	3	7	3,7	0
E6	20	1	5	1	2,3	0
E7	24	2	3	4	3,0	0
E8	25	1	2	2	1,7	1
E9	26	1	5	3	3,0	0
E10	33	2	6	3	3,7	0
E11	36	2	4	7	4,3	0
E12	37	2	9	2	4,3	0
E13	39	2	8	9	6,3	0
E14	40	1	1	2	1,3	1
E15	45	1	6	7	4,7	0
E16	46	6	4	9	6,3	0
Mean C		2,1	4,2	5,4	3,9	1
Mean E		1,8	4,3	4,4	3,5	3
				Total C	43%	6%
				Total E	39%	19%
native	F1	1	9	9	6,3	0
native	F2	8	2	9	6,3	0
native	M1	6	9	7	7,3	0
native	M2	8	9	9	8,7	1
				Mean	7,2	1
				Total		25%

Stimulus: *pool*				Time: posttest		
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	6	4	7	5,7	0
C2	4	1	3	8	4,0	0
C3	7	2	8	7	5,7	0
C4	9	1	3	7	3,7	0
C5	10	1	7	8	5,3	0
C6	11	2	5	8	5,0	0
C7	12	3	7	7	5,7	0
C8	19	1	1	6	2,7	0
C9	23	1	1	1	1,0	1
C10	22	6	2	8	5,3	0
C11	30	1	3	7	3,7	0
C12	41	2	5	6	4,3	0
C13	42	1	2	8	3,7	0
C14	44	2	3	9	4,7	0
C15	48	2	2	1	1,7	1
C16	50	1	5	6	4,0	0
E1	2	6	9	1	5,3	0
E2	15	2	1	7	3,3	0
E3	17	2	2	8	4,0	0
E4	18	1	3	1	1,7	1
E5	14	2	2	8	4,0	0
E6	20	2	4	1	2,3	0
E7	24	1	2	3	2,0	1
E8	25	2	1	2	1,7	1
E9	26	6	5	2	4,3	0
E10	33	1	4	1	2,0	0
E11	36	7	4	8	6,3	0
E12	37	1	5	9	5,0	0
E13	39	1	2	8	3,7	0
E14	40	1	4	9	4,7	0
E15	45	3	3	7	4,3	0
E16	46	1	1	3	1,7	1
Mean C		2,1	3,8	6,5	4,1	2
Mean E		2,4	3,3	4,9	3,5	4
				Total C	46%	13%
				Total E	39%	25%

121

Stimulus: *pull*				Time: pretest	
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	1	2	6	3,0	0
C2	4	2	5	6	4,3	0
C3	7	3	7	6	5,3	0
C4	9	6	6	3	5,0	0
C5	10	1	1	4	2,0	0
C6	11	5	6	3	4,7	0
C7	12	2	2	8	4,0	0
C8	19	2	4	5	3,7	0
C9	23	7	8	2	5,7	0
C10	22	8	8	9	8,3	1
C11	30	2	6	5	4,3	0
C12	41	9	7	7	7,7	1
C13	42	6	4	7	5,7	0
C14	44	7	5	4	5,3	0
C15	48	6	4	6	5,3	1
C16	50	8	8	5	7,0	0
E1	2	2	7	5	4,7	0
E2	15	1	1	5	2,3	0
E3	17	8	3	4	5,0	0
E4	18	9	8	3	6,7	0
E5	14	1	2	8	3,7	0
E6	20	1	6	1	2,7	0
E7	24	1	3	9	4,3	0
E8	25	7	8	8	7,7	1
E9	26	5	6	2	4,3	0
E10	33	7	9	7	7,7	1
E11	36	8	8	8	8,0	1
E12	37	5	5	3	4,3	0
E13	39	1	5	7	4,3	0
E14	40	6	5	7	6,0	0
E15	45	5	6	4	5,0	1
E16	46	1	3	6	3,3	0
Mean C		4,7	5,2	5,4	5,1	3
Mean E		4,3	5,3	5,4	5,0	4
				Total C	57%	19%
				Total E	56%	25%
native	F1	2	1	9	4,0	0
native	F2	1	1	9	3,7	0
native	M1	8	2	9	6,3	0
native	M2	9	3	9	7,0	0
				Mean	5,3	0
				Total		0%

Stimulus: *pull*				Time: posttest	
Group	**Code**	**Rater 1**	**Rater 2**	**Rater 3**	**Mean**	**Items used**
C1	6	2	4	7	4,3	0
C2	4	1	2	1	1,3	1
C3	7	9	7	7	7,7	1
C4	9	2	3	4	3,0	0
C5	10	7	4	7	6,0	0
C6	11	8	6	6	6,7	0
C7	12	3	2	7	4,0	0
C8	19	1	2	9	4,0	0
C9	23	1	1	1	1,0	1
C10	22	8	6	1	5,0	0
C11	30	7	4	9	6,7	0
C12	41	4	6	4	4,7	1
C13	42	1	3	5	3,0	0
C14	44	1	2	1	1,3	1
C15	48	2	2	4	2,7	1
C16	50	2	4	9	5,0	0
E1	2	9	7	4	6,7	0
E2	15	8	4	7	6,3	0
E3	17	1	6	6	4,3	0
E4	18	5	5	2	4,0	0
E5	14	2	3	6	3,7	0
E6	20	6	2	2	3,3	0
E7	24	1	1	8	3,3	0
E8	25	7	3	8	6,0	0
E9	26	6	6	3	5,0	0
E10	33	1	3	3	2,3	1
E11	36	1	3	6	3,3	0
E12	37	1	3	1	1,7	1
E13	39	1	1	7	3,0	0
E14	40	9	2	7	6,0	0
E15	45	1	1	6	2,7	0
E16	46	9	5	9	7,7	0
Mean C		3,7	3,6	5,1	4,1	6
Mean E		4,3	3,4	5,3	4,3	2
				Total C	46%	38%
				Total E	48%	13%

Stimulus: *took*				Time: pretest	
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	6	7	4	5,7	0
C2	4	8	6	8	7,3	0
C3	7	6	6	7	6,3	0
C4	9	2	7	4	4,3	0
C5	10	1	2	9	4,0	0
C6	11	8	7	8	7,7	1
C7	12	5	5	4	4,7	1
C8	19	2	7	8	5,7	0
C9	23	8	9	3	6,7	0
C10	22	8	8	5	7,0	0
C11	30	9	9	7	8,3	1
C12	41	6	1	5	4,0	0
C13	42	6	1	6	4,3	0
C14	44	2	4	4	3,3	0
C15	48	1	3	6	3,3	0
C16	50	2	5	3	3,3	0
E1	2	7	8	4	6,3	0
E2	15	1	3	7	3,7	0
E3	17	7	6	8	7,0	0
E4	18	2	8	5	5,0	0
E5	14	3	3	3	3,0	1
E6	20	3	1	9	4,3	0
E7	24	1	3	6	3,3	0
E8	25	4	1	7	4,0	0
E9	26	1	4	4	3,0	0
E10	33	2	4	3	3,0	0
E11	36	2	4	5	3,7	0
E12	37	1	4	4	3,0	0
E13	39	7	6	5	6,0	0
E14	40	2	5	6	4,3	0
E15	45	1	4	2	2,3	0
E16	46	5	3	3	3,7	0
Mean C		5,0	5,4	5,7	5,4	3
Mean E		3,1	4,2	5,1	4,1	1
				Total C	60%	19%
				Total E	46%	6%
native	F1	5	5	9	6,3	0
native	F2	3	5	9	5,7	0
native	M1	6	6	9	7,0	0
native	M2	3	5	5	4,3	0
				Mean	5,8	0
				Total		0%

124

Stimulus: *took*				Time: posttest	
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	2	5	3	3,3	0
C2	4	7	6	6	6,3	0
C3	7	8	8	3	6,3	0
C4	9	8	8	7	7,7	1
C5	10	7	8	4	6,3	0
C6	11	1	1	3	1,7	1
C7	12	8	4	2	4,7	0
C8	19	2	6	6	4,7	0
C9	23	7	7	2	5,3	0
C10	22	7	7	3	5,7	0
C11	30	8	9	6	7,7	0
C12	41	1	1	1	1,0	1
C13	42	2	6	9	5,7	0
C14	44	7	5	4	5,3	0
C15	48	2	3	7	4,0	0
C16	50	1	2	2	1,7	1
E1	2	9	9	4	7,3	0
E2	15	3	4	6	4,3	0
E3	17	6	7	6	6,3	0
E4	18	5	7	3	5,0	0
E5	14	4	4	5	4,3	1
E6	20	6	3	7	5,3	0
E7	24	6	4	5	5,0	1
E8	25	2	2	5	3,0	0
E9	26	8	5	2	5,0	0
E10	33	5	7	5	5,7	0
E11	36	8	9	4	7,0	0
E12	37	5	2	2	3,0	0
E13	39	8	6	3	5,7	0
E14	40	9	8	8	8,3	1
E15	45	2	2	3	2,3	1
E16	46	2	4	2	2,7	0
Mean C		4,9	5,4	4,3	4,8	4
Mean E		5,5	5,2	4,4	5,0	4
				Total C	53%	25%
				Total E	56%	25%

Stimulus: *too*					Time: pretest	
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	4	3	5	4,0	0
C2	4	9	8	9	8,7	1
C3	7	4	4	5	4,3	1
C4	9	8	5	6	6,3	0
C5	10	6	3	7	5,3	0
C6	11	7	8	9	8,0	1
C7	12	8	7	9	8,0	1
C8	19	1	1	7	3,0	0
C9	23	6	6	7	6,3	0
C10	22	8	7	8	7,7	1
C11	30	3	2	7	4,0	0
C12	41	8	8	7	7,7	1
C13	42	3	3	3	3,0	1
C14	44	9	8	9	8,7	1
C15	48	8	8	7	7,7	1
C16	50	3	4	6	4,3	0
E1	2	6	6	9	7,0	0
E2	15	4	5	7	5,3	0
E3	17	9	4	7	6,7	0
E4	18	1	2	1	1,3	1
E5	14	8	6	9	7,7	0
E6	20	2	3	9	4,7	0
E7	24	1	1	5	2,3	0
E8	25	1	2	4	2,3	0
E9	26	5	5	6	5,3	1
E10	33	8	6	7	7,0	0
E11	36	8	5	4	5,7	0
E12	37	8	7	9	8,0	1
E13	39	6	6	7	6,3	0
E14	40	6	4	8	6,0	0
E15	45	8	7	8	7,7	1
E16	46	6	4	6	5,3	1
Mean C		5,9	5,3	6,9	6,1	9
Mean E		5,4	4,6	6,6	5,5	5
				Total C	68%	56%
				Total E	61%	31%
native	F1	6	5	8	6,3	0
native	F2	7	9	9	8,3	1
native	M1	6	7	8	7,0	0
native	M2	8	8	9	8,3	1
				Mean	7,5	2
				Total		50%

Stimulus: *too*				Time: posttest	
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used

Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	6	6	7	6,3	0
C2	4	8	8	5	7,0	0
C3	7	8	6	7	7,0	0
C4	9	9	6	9	8,0	0
C5	10	7	7	8	7,3	1
C6	11	6	1	7	4,7	0
C7	12	8	9	7	8,0	1
C8	19	4	3	7	4,7	0
C9	23	7	9	8	8,0	1
C10	22	7	9	9	8,3	1
C11	30	4	2	2	2,7	0
C12	41	9	8	9	8,7	1
C13	42	3	8	8	6,3	0
C14	44	7	8	9	8,0	1
C15	48	8	7	9	8,0	1
C16	50	6	4	6	5,3	1
E1	2	9	9	7	8,3	1
E2	15	8	6	8	7,3	0
E3	17	6	4	4	4,7	1
E4	18	4	2	7	4,3	0
E5	14	7	7	8	7,3	1
E6	20	5	3	9	5,7	0
E7	24	8	4	9	7,0	0
E8	25	1	3	1	1,7	1
E9	26	7	5	8	6,7	0
E10	33	5	6	6	5,7	1
E11	36	9	9	7	8,3	1
E12	37	2	1	6	3,0	0
E13	39	8	6	8	7,3	0
E14	40	8	7	9	8,0	1
E15	45	8	5	6	6,3	0
E16	46	4	1	5	3,3	0
Mean C		6,7	6,3	7,3	6,8	8
Mean E		6,2	4,9	6,8	5,9	7
				Total C	76%	50%
				Total E	66%	44%

127

Stimulus: *mix*				Time: pretest		
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	8	7	5	6,7	0
C2	4	6	8	3	5,7	0
C3	7	1	2	5	2,7	0
C4	9	8	6	1	5,0	0
C5	10	1	3	3	2,3	1
C6	11	8	5	1	4,7	0
C7	12	6	4	9	6,3	0
C8	19	1	1	3	1,7	1
C9	23	8	5	6	6,3	0
C10	22	9	4	4	5,7	0
C11	30	2	7	6	5,0	0
C12	41	1	8	8	5,7	0
C13	42	7	2	5	4,7	0
C14	44	1	4	1	2,0	0
C15	48	8	6	7	7,0	0
C16	50	8	6	8	7,3	0
E1	2	4	1	1	2,0	0
E2	15	8	6	4	6,0	0
E3	17	8	8	8	8,0	1
E4	18	7	4	4	5,0	0
E5	14	7	7	6	6,7	0
E6	20	1	3	3	2,3	1
E7	24	1	7	4	4,0	0
E8	25	5	7	3	5,0	0
E9	26	7	5	4	5,3	0
E10	33	8	5	5	6,0	0
E11	36	3	5	5	4,3	0
E12	37	1	1	1	1,0	1
E13	39	7	4	4	5,0	0
E14	40	1	2	1	1,3	1
E15	45	9	7	4	6,7	0
E16	46	2	6	5	4,3	0
Mean C		5,2	4,9	4,7	4,9	2
Mean E		4,9	4,9	3,9	4,6	4
				Total C	54%	13%
				Total E	51%	25%
native	F1	7	8	9	8,0	1
native	F2	9	6	9	8,0	0
native	M1	9	9	9	9,0	1
native	M2	9	9	9	9,0	1
				Mean	8,5	3
				Total		75%

128

Stimulus: *mix*					Time: posttest	
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	7	7	4	6,0	0
C2	4	9	8	5	7,3	0
C3	7	8	5	6	6,3	0
C4	9	6	5	3	4,7	0
C5	10	3	4	2	3,0	0
C6	11	9	6	8	7,7	0
C7	12	9	8	7	8,0	1
C8	19	1	1	3	1,7	1
C9	23	8	7	7	7,3	1
C10	22	8	8	6	7,3	0
C11	30	5	7	7	6,3	0
C12	41	8	6	6	6,7	0
C13	42	7	7	4	6,0	0
C14	44	8	8	7	7,7	1
C15	48	8	4	4	5,3	0
C16	50	6	7	8	7,0	0
E1	2	1	7	1	3,0	0
E2	15	7	6	3	5,3	0
E3	17	7	6	8	7,0	0
E4	18	7	6	9	7,3	0
E5	14	8	7	9	8,0	1
E6	20	3	4	8	5,0	0
E7	24	1	3	1	1,7	1
E8	25	3	3	4	3,3	0
E9	26	7	3	4	4,7	0
E10	33	7	4	5	5,3	0
E11	36	8	4	4	5,3	0
E12	37	1	2	1	1,3	1
E13	39	7	3	4	4,7	0
E14	40	2	5	5	4,0	0
E15	45	7	7	6	6,7	0
E16	46	1	3	1	1,7	1
Mean C		6,9	6,1	5,4	6,1	4
Mean E		4,8	4,6	4,6	4,6	4
				Total C	68%	13%
				Total E	51%	13%

Stimulus: *weeks*				Time: pretest		
Group	**Code**	**Rater 1**	**Rater 2**	**Rater 3**	**Mean**	**Items used**
C1	6	7	8	9	8,0	1
C2	4	8	7	9	8,0	1
C3	7	2	2	6	3,3	0
C4	9	3	4	9	5,3	0
C5	10	3	1	9	4,3	0
C6	11	8	8	9	8,3	1
C7	12	5	6	7	6,0	0
C8	19	2	3	4	3,0	0
C9	23	8	9	9	8,7	1
C10	22	3	8	9	6,7	0
C11	30	3	6	6	5,0	0
C12	41	9	9	9	9,0	1
C13	42	2	1	7	3,3	0
C14	44	6	7	7	6,7	0
C15	48	8	9	7	8,0	1
C16	50	4	5	7	5,3	0
E1	2	7	7	9	7,7	1
E2	15	3	7	9	6,3	0
E3	17	3	5	9	5,7	0
E4	18	6	8	8	7,3	0
E5	14	4	4	9	5,7	0
E6	20	6	7	9	7,3	0
E7	24	7	6	9	7,3	0
E8	25	3	8	9	6,7	0
E9	26	8	8	9	8,3	1
E10	33	2	7	9	6,0	0
E11	36	3	8	9	6,7	0
E12	37	7	6	8	7,0	0
E13	39	7	9	9	8,3	1
E14	40	2	3	5	3,3	0
E15	45	4	7	9	6,7	0
E16	46	6	7	8	7,0	0
Mean C		5,1	5,8	7,7	6,2	6
Mean E		4,9	6,7	8,6	6,7	3
				Total C	69%	38%
				Total E	74%	19%
native	F1	7	9	9	8,3	1
native	F2	3	5	9	5,7	0
native	M1	8	8	9	8,3	1
native	M2	9	9	9	9,0	1
				Mean	7,8	3
				Total		75%

130

Stimulus: *weeks*				Time: posttest	
Group	**Code**	**Rater 1**	**Rater 2**	**Rater 3**	**Mean**	**Items used**
C1	6	9	8	9	8,7	1
C2	4	8	6	9	7,7	0
C3	7	6	3	9	6,0	0
C4	9	8	6	9	7,7	0
C5	10	8	9	9	8,7	1
C6	11	8	8	9	8,3	1
C7	12	8	8	9	8,3	1
C8	19	3	3	8	4,7	0
C9	23	4	6	9	6,3	0
C10	22	8	8	7	7,7	1
C11	30	2	1	4	2,3	0
C12	41	7	3	7	5,7	0
C13	42	6	8	9	7,7	0
C14	44	8	5	7	6,7	0
C15	48	6	4	9	6,3	0
C16	50	1	1	2	1,3	1
E1	2	7	8	6	7,0	0
E2	15	7	6	8	7,0	0
E3	17	4	8	9	7,0	0
E4	18	9	9	9	9,0	1
E5	14	2	6	8	5,3	0
E6	20	6	2	9	5,7	0
E7	24	9	8	9	8,7	1
E8	25	5	4	9	6,0	0
E9	26	7	8	9	8,0	1
E10	33	4	6	9	6,3	0
E11	36	8	8	9	8,3	1
E12	37	8	8	9	8,3	1
E13	39	7	8	9	8,0	1
E14	40	7	8	9	8,0	1
E15	45	6	8	9	7,7	0
E16	46	2	4	6	4,0	0
Mean C		6,3	5,4	7,8	6,5	6
Mean E		6,1	6,8	8,5	7,1	7
				Total C	87%	38%
				Total E	94%	44%

Stimulus: *seems*				Time: pretest	
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used

Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	8	5	9	7,3	0
C2	4	6	4	5	5,0	1
C3	7	6	3	6	5,0	0
C4	9	9	8	9	8,7	1
C5	10	2	4	9	5,0	0
C6	11	7	7	9	7,7	1
C7	12	9	8	9	8,7	1
C8	19	5	8	9	7,3	0
C9	23	9	5	9	7,7	0
C10	22	7	6	9	7,3	0
C11	30	9	8	8	8,3	1
C12	41	8	6	9	7,7	0
C13	42	2	1	9	4,0	0
C14	44	8	8	9	8,3	1
C15	48	9	6	9	8,0	0
C16	50	7	6	8	7,0	0
E1	2	9	6	9	8,0	0
E2	15	7	4	9	6,7	0
E3	17	7	5	6	6,0	0
E4	18	9	7	9	8,3	1
E5	14	9	6	9	8,0	0
E6	20	5	5	8	6,0	0
E7	24	8	6	9	7,7	0
E8	25	4	1	1	2,0	0
E9	26	8	7	9	8,0	1
E10	33	9	8	9	8,7	1
E11	36	7	3	3	4,3	0
E12	37	8	5	9	7,3	0
E13	39	8	7	9	8,0	1
E14	40	4	1	9	4,7	0
E15	45	8	2	8	6,0	0
E16	46	3	4	7	4,7	0
Mean C		6,9	5,8	8,4	7,1	6
Mean E		7,1	4,8	7,7	6,5	5
				Total C	79%	38%
				Total E	72%	31%
native	F1	9	7	9	8,3	1
native	F2	2	6	9	5,7	0
native	M1	9	6	9	8,0	0
native	M2	9	1	9	6,3	0
				Mean	7,1	1
				Total		25%

132

Stimulus: *seems*					Time: posttest	
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	8	6	9	7,7	0
C2	4	1	1	1	1,0	1
C3	7	1	1	1	1,0	1
C4	9	6	3	1	3,3	0
C5	10	7	7	9	7,7	1
C6	11	9	6	9	8,0	0
C7	12	9	6	9	8,0	0
C8	19	7	3	7	5,7	0
C9	23	3	5	7	5,0	0
C10	22	9	6	8	7,7	0
C11	30	8	6	8	7,3	0
C12	41	7	4	7	6,0	0
C13	42	9	6	9	8,0	0
C14	44	9	7	9	8,3	1
C15	48	8	2	5	5,0	0
C16	50	9	6	9	8,0	0
E1	2	9	8	7	8,0	1
E2	15	6	3	9	6,0	0
E3	17	7	7	8	7,3	1
E4	18	9	8	9	8,7	1
E5	14	8	5	9	7,3	0
E6	20	5	3	9	5,7	0
E7	24	8	3	9	6,7	0
E8	25	2	3	1	2,0	1
E9	26	8	7	9	8,0	1
E10	33	7	5	9	7,0	0
E11	36	5	3	4	4,0	0
E12	37	7	3	9	6,3	0
E13	39	8	9	9	8,7	1
E14	40	9	2	9	6,7	0
E15	45	9	4	9	7,3	0
E16	46	8	3	6	5,7	0
Mean C		6,9	4,7	6,8	6,1	4
Mean E		7,2	4,8	7,8	6,6	6
				Total C	68%	25%
				Total E	73%	38%

133

Stimulus: *Tim*				Time: pretest		
Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	5	6	2	4,3	0
C2	4	6	5	4	5,0	1
C3	7	6	2	1	3,0	0
C4	9	6	5	2	4,3	0
C5	10	1	1	5	2,3	0
C6	11	1	7	4	4,0	0
C7	12	6	3	4	4,3	0
C8	19	5	3	2	3,3	0
C9	23	4	4	3	3,7	0
C10	22	6	6	2	4,7	0
C11	30	5	5	2	4,0	0
C12	41	2	4	3	3,0	0
C13	42	1	1	1	1,0	1
C14	44	7	7	3	5,7	0
C15	48	6	5	8	6,3	0
C16	50	2	4	2	2,7	0
E1	2	1	1	2	1,3	1
E2	15	7	2	3	4,0	0
E3	17	5	7	4	5,3	0
E4	18	1	2	2	1,7	1
E5	14	9	8	2	6,3	0
E6	20	1	3	3	2,3	1
E7	24	7	5	3	5,0	0
E8	25	8	7	4	6,3	0
E9	26	6	5	3	4,7	0
E10	33	7	4	4	5,0	0
E11	36	6	6	2	4,7	0
E12	37	3	2	1	2,0	1
E13	39	3	4	2	3,0	0
E14	40	6	4	5	5,0	1
E15	45	4	4	1	3,0	0
E16	46	7	7	1	5,0	0
Mean C		4,3	4,3	3,0	3,9	2
Mean E		5,1	4,4	2,6	4,0	5
				Total C	43%	13%
				Total E	44%	31%
native	F1	8	9	9	8,7	1
native	F2	7	7	9	7,7	1
native	M1	9	9	9	9,0	1
native	M2	5	8	9	7,3	0
				Mean	8,2	3
				Total		75%

134

Stimulus: *Tim*				Time: posttest	
Group	**Code**	**Rater 1**	**Rater 2**	**Rater 3**	**Mean**	**Items used**
C1	6	5	6	3	4,7	0
C2	4	4	7	4	5,0	0
C3	7	6	6	1	4,3	0
C4	9	6	6	3	5,0	0
C5	10	5	4	2	3,7	0
C6	11	1	3	1	1,7	1
C7	12	2	1	3	2,0	1
C8	19	2	1	1	1,3	1
C9	23	2	2	3	2,3	1
C10	22	4	5	4	4,3	1
C11	30	2	5	4	3,7	0
C12	41	3	7	3	4,3	0
C13	42	6	4	1	3,7	0
C14	44	2	7	4	4,3	0
C15	48	7	8	9	8,0	1
C16	50	3	3	4	3,3	0
E1	2	8	6	1	5,0	0
E2	15	7	4	1	4,0	0
E3	17	2	2	2	2,0	1
E4	18	7	9	9	8,3	1
E5	14	1	6	1	2,7	0
E6	20	2	4	2	2,7	0
E7	24	4	3	2	3,0	0
E8	25	6	3	1	3,3	0
E9	26	7	4	5	5,3	0
E10	33	1	2	4	2,3	0
E11	36	7	4	2	4,3	0
E12	37	3	2	1	2,0	1
E13	39	6	6	4	5,3	1
E14	40	4	5	4	4,3	1
E15	45	2	3	3	2,7	1
E16	46	4	4	3	3,7	0
Mean C		3,8	4,7	3,1	3,9	6
Mean E		4,4	4,2	2,8	3,8	6
				Total C	43%	67%
				Total E	42%	67%

Appendix H. Two-rater regrouped mean scores on a 5-point scale for the words *caught, boss, pull, took,* and *Tim* at pre- and posttest, and for the word *mix* at posttest.

Group	Code	Rater A	Rater C	Mean	Items used
C1	10	4	2	3,0	1
C2	11	2	2	2,0	1
C3	12	2	2	2,0	1
C4	19	3	2	2,5	1
C5	22	4	2	3,0	1
C6	23	3	1	2,0	1
C7	30	3	2	2,5	1
C8	4	2	2	2,0	1
C9	41	4	1	2,5	0
C10	42	1	1	1,0	1
C11	44	2	1	1,5	1
C12	48	4	1	2,5	0
C13	50	4	3	3,5	1
C14	6	3	3	3,0	1
C15	7	2	1	1,5	1
C16	9	4	2	3,0	1
E1	14	4	4	4,0	1
E2	15	2	1	1,5	1
E3	17	3	4	3,5	1
E4	18	2	1	1,5	1
E5	2	3	3	3,0	1
E6	20	3	2	2,5	1
E7	24	4	4	4,0	1
E8	25	3	2	2,5	1
E9	26	3	3	3,0	1
E10	33	3	2	2,5	1
E11	36	1	1	1,0	1
E12	37	2	2	2,0	1
E13	39	1	1	1,0	1
E14	40	4	3	3,5	1
E15	45	4	3	3,5	1
E16	46	2	2	2,0	1
Mean C		2,9	1,8	2,3	12
Mean E		2,8	2,4	2,6	16
			Total C	46%	75%
			Total E	52%	100%
Native	F2	5	5	5,0	1
Native	F1	5	5	5,0	1
Native	M2	5	5	5,0	1
Native	M1	5	5	5,0	1
			Mean	5,0	4
			Total		100%

136

Results on a 5-point scale for the word *caught* at posttest

Group	Code	Rater A	Rater C	Mean	Items used
C1	10	3	2	2,5	1
C2	11	3	3	3,0	1
C3	12	3	2	2,5	1
C4	19	3	4	3,5	1
C5	22	3	3	3,0	1
C6	23	3	4	3,5	1
C7	30	3	1	2,0	1
C8	4	3	2	2,5	1
C9	41	5	1	3,0	0
C10	42	1	1	1,0	1
C11	44	2	2	2,0	1
C12	48	5	2	3,5	0
C13	50	4	2	3,0	1
C14	6	4	3	3,5	1
C15	7	2	2	2,0	1
C16	9	4	3	3,5	1
E1	14	1	1	1,0	1
E2	15	4	2	3,0	1
E3	17	4	4	4,0	1
E4	18	3	2	2,5	1
E5	2	3	4	3,5	1
E6	20	4	4	4,0	1
E7	24	4	3	3,5	1
E8	25	3	2	2,5	1
E9	26	5	3	4,0	1
E10	33	4	3	3,5	1
E11	36	5	2	3,5	0
E12	37	4	4	4,0	1
E13	39	3	3	3,0	1
E14	40	5	3	4,0	1
E15	45	2	3	2,5	1
E16	46	4	3	3,5	1
Mean C		3,2	2,3	2,8	14
Mean E		3,6	2,9	3,3	15
			Total C	56%	88%
			Total E	66%	94%

Results on a 5-point scale for the word *boss* at pretest.

Group	Code	Rater A	Rater C	Mean	Items used
C1	10	5	3	4,0	1
C2	11	4	4	4,0	1
C3	12	5	1	3,0	0
C4	19	2	3	2,5	1
C5	22	5	4	4,5	1
C6	23	5	4	4,5	1
C7	30	3	2	2,5	1
C8	4	4	2	3,0	1
C9	41	3	2	2,5	1
C10	42	5	2	3,5	0
C11	44	3	2	2,5	1
C12	48	3	3	3,0	1
C13	50	4	4	4,0	1
C14	6	5	2	3,5	0
C15	7	4	2	3,0	1
C16	9	4	4	4,0	1
E1	14	5	3	4,0	1
E2	15	4	2	3,0	1
E3	17	4	3	3,5	1
E4	18	3	3	3,0	1
E5	2	2	2	2,0	1
E6	20	4	3	3,5	1
E7	24	4	3	3,5	1
E8	25	4	4	4,0	1
E9	26	3	4	3,5	1
E10	33	4	3	3,5	1
E11	36	5	2	3,5	0
E12	37	3	1	2,0	1
E13	39	4	4	4,0	1
E14	40	3	3	3,0	1
E15	45	5	4	4,5	1
E16	46	5	3	4,0	1
Mean C		4,0	2,8	3,4	13
Mean E		3,9	2,9	3,4	15
			Total C	68%	81%
			Total E	68%	94%
Native	F2	5	5	5,0	1
Native	F1	5	3	4,0	1
Native	M2	5	5	5,0	1
Native	M1	5	5	5,0	1
			Mean	4,8	4
				Total	100%

Results on a 5-point scale for the word *boss* at posttest

Group	Code	Rater A	Rater C	Mean	Items used
C1	10	4	2	3,0	1
C2	11	3	3	3,0	1
C3	12	5	5	5,0	1
C4	19	4	2	3,0	1
C5	22	4	3	3,5	1
C6	23	4	2	3,0	1
C7	30	4	4	4,0	1
C8	4	3	3	3,0	1
C9	41	5	5	5,0	1
C10	42	3	3	3,0	1
C11	44	4	3	3,5	1
C12	48	5	5	5,0	1
C13	50	2	1	1,5	1
C14	6	4	4	4,0	1
C15	7	3	3	3,0	1
C16	9	5	3	4,0	1
E1	14	4	3	3,5	1
E2	15	5	5	5,0	1
E3	17	5	3	4,0	1
E4	18	3	2	2,5	1
E5	2	2	2	2,0	1
E6	20	3	3	3,0	1
E7	24	5	4	4,5	1
E8	25	5	4	4,5	1
E9	26	5	4	4,5	1
E10	33	3	3	3,0	1
E11	36	3	2	2,5	1
E12	37	3	1	2,0	1
E13	39	3	4	3,5	1
E14	40	4	3	3,5	1
E15	45	4	1	2,5	1
E16	46	3	3	3,0	1
Mean C		3,9	3,2	3,5	16
Mean E		3,8	2,9	3,3	16
			Total C	70%	100%
			Total E	66%	100%

139

Results on a 5-point scale for the word *pull* at pretest.

Group	Code	Rater A	Rater C	Mean	Items used
C1	10	3	2	2,5	1
C2	11	3	3	3,0	1
C3	12	3	3	3,0	1
C4	19	3	3	3,0	1
C5	22	4	4	4,0	1
C6	23	5	4	4,5	1
C7	30	5	3	4,0	1
C8	4	1	2	1,5	1
C9	41	4	4	4,0	1
C10	42	4	3	3,5	1
C11	44	4	2	3,0	1
C12	48	3	3	3,0	1
C13	50	3	4	3,5	1
C14	6	4	2	3,0	1
C15	7	3	3	3,0	1
C16	9	3	3	3,0	1
E1	14	3	2	2,5	1
E2	15	4	2	3,0	1
E3	17	3	3	3,0	1
E4	18	2	3	2,5	1
E5	2	3	2	2,5	1
E6	20	3	3	3,0	1
E7	24	4	4	4,0	1
E8	25	5	5	5,0	1
E9	26	4	3	3,5	1
E10	33	5	4	4,5	1
E11	36	4	4	4,0	1
E12	37	2	2	2,0	1
E13	39	5	4	4,5	1
E14	40	5	4	4,5	1
E15	45	4	3	3,5	1
E16	46	3	2	2,5	1
Mean C		3,4	3,0	3,2	16
Mean E		3,7	3,1	3,4	16
			Total C	64%	100%
			Total E	68%	100%
Native	F2	5	5	5,0	1
Native	F1	5	5	5,0	1
Native	M2	5	5	5,0	1
Native	M1	5	4	4,5	1
			Mean	4,9	4
				Total	100%

Results on a 5-point scale for the word *pull* at posttest

Group	Code	Rater A	Rater C	Mean	Items used
C1	10	4	3	3,5	1
C2	11	5	3	4,0	1
C3	12	4	4	4,0	1
C4	19	3	3	3,0	1
C5	22	2	2	2,0	1
C6	23	2	2	2,0	1
C7	30	5	4	4,5	1
C8	4	1	2	1,5	1
C9	41	4	4	4,0	1
C10	42	2	3	2,5	1
C11	44	3	1	2,0	1
C12	48	5	3	4,0	1
C13	50	4	4	4,0	1
C14	6	4	5	4,5	1
C15	7	4	2	3,0	1
C16	9	3	3	3,0	1
E1	14	3	3	3,0	1
E2	15	5	5	5,0	1
E3	17	3	2	2,5	1
E4	18	3	2	2,5	1
E5	2	2	2	2,0	1
E6	20	3	3	3,0	1
E7	24	5	4	4,5	1
E8	25	2	4	3,0	1
E9	26	2	3	2,5	1
E10	33	4	3	3,5	1
E11	36	3	2	2,5	1
E12	37	2	1	1,5	1
E13	39	2	3	2,5	1
E14	40	4	4	4,0	1
E15	45	4	3	3,5	1
E16	46	5	5	5,0	1
Mean C		3,4	3,0	3,2	16
Mean E		3,3	3,1	3,2	16
			Total C	64%	100%
			Total E	64%	100%

141

Results on a 5-point scale for the word *took* at pretest.

Group	Code	Rater A	Rater C	Mean	Items used
C1	10	3	3	3,0	1
C2	11	5	4	4,5	1
C3	12	4	3	3,5	1
C4	19	3	4	3,5	1
C5	22	4	2	3,0	1
C6	23	4	3	3,5	1
C7	30	5	3	4,0	1
C8	4	5	4	4,5	1
C9	41	3	1	2,0	1
C10	42	3	4	3,5	1
C11	44	5	3	4,0	1
C12	48	4	3	3,5	1
C13	50	3	4	3,5	1
C14	6	3	2	2,5	1
C15	7	4	4	4,0	1
C16	9	5	3	4,0	1
E1	14	4	4	4,0	1
E2	15	4	4	4,0	1
E3	17	4	3	3,5	1
E4	18	3	1	2,0	1
E5	2	5	3	4,0	1
E6	20	4	3	3,5	1
E7	24	5	3	4,0	1
E8	25	2	2	2,0	1
E9	26	5	2	3,5	0
E10	33	4	2	3,0	1
E11	36	5	4	4,5	1
E12	37	3	3	3,0	1
E13	39	4	4	4,0	1
E14	40	4	4	4,0	1
E15	45	3	3	3,0	1
E16	46	3	2	2,5	1
Mean C		3,9	3,1	3,5	16
Mean E		3,9	2,9	3,4	15
			Total C	70%	100%
			Total E	68%	94%
Native	F2	5	5	5,0	1
Native	F1	5	5	5,0	1
Native	M2	5	5	5,0	1
Native	M1	5	5	5,0	1
			Mean	5,0	4
			Total		100%

Results on a 5-point scale for the word *took* at posttest

Group	Code	Rater A	Rater C	Mean	Items used
C1	10	4	3	3,5	1
C2	11	4	3	3,5	1
C3	12	4	3	3,5	1
C4	19	4	3	3,5	1
C5	22	3	3	3,0	1
C6	23	2	2	2,0	1
C7	30	3	2	2,5	1
C8	4	5	4	4,5	1
C9	41	1	1	1,0	1
C10	42	5	4	4,5	1
C11	44	4	3	3,5	1
C12	48	5	2	3,5	0
C13	50	4	3	3,5	1
C14	6	3	3	3,0	1
C15	7	3	2	2,5	1
C16	9	4	2	3,0	1
E1	14	5	4	4,5	1
E2	15	5	3	4,0	1
E3	17	3	3	3,0	1
E4	18	3	2	2,5	1
E5	2	3	3	3,0	1
E6	20	3	3	3,0	1
E7	24	4	4	4,0	1
E8	25	1	1	1,0	1
E9	26	3	2	2,5	1
E10	33	4	3	3,5	1
E11	36	4	3	3,5	1
E12	37	3	3	3,0	1
E13	39	3	3	3,0	1
E14	40	3	3	3,0	1
E15	45	4	3	3,5	1
E16	46	5	3	4,0	1
Mean C		3,6	2,7	3,2	15
Mean E		3,5	2,9	3,2	16
			Total C	64%	94%
			Total E	64%	100%

143

Results on a 5-point scale for the word *Tim* at pretest.

Group	Code	Rater A	Rater C	Mean	Items used
C1	10	4	4	4,0	1
C2	11	3	2	2,5	1
C3	12	4	4	4,0	1
C4	19	5	3	4,0	1
C5	22	4	4	4,0	1
C6	23	3	2	2,5	1
C7	30	2	3	2,5	1
C8	4	4	4	4,0	1
C9	41	4	4	4,0	1
C10	42	3	3	3,0	1
C11	44	5	3	4,0	1
C12	48	4	2	3,0	1
C13	50	3	3	3,0	1
C14	6	4	5	4,5	1
C15	7	3	3	3,0	1
C16	9	4	2	3,0	1
E1	14	2	2	2,0	1
E2	15	5	5	5,0	1
E3	17	3	2	2,5	1
E4	18	4	3	3,5	1
E5	2	3	2	2,5	1
E6	20	3	3	3,0	1
E7	24	3	3	3,0	1
E8	25	5	3	4,0	1
E9	26	3	3	3,0	1
E10	33	3	2	2,5	1
E11	36	2	2	2,0	1
E12	37	3	2	2,5	1
E13	39	3	3	3,0	1
E14	40	4	3	3,5	1
E15	45	3	3	3,0	1
E16	46	4	3	3,5	1
Mean C		3,7	3,2	3,4	16
Mean E		3,3	2,8	3,0	16
			Total C	68%	100%
			Total E	60%	100%
Native	F2	5	5	5,0	1
Native	F1	5	5	5,0	1
Native	M2	5	5	5,0	1
Native	M1	5	5	5,0	1
			Mean	5,0	4
				Total	100%

Results on a 5-point scale for the word *Tim* at posttest

Group	Code	Rater A	Rater C	Mean	Items used
C1	10	2	1	1,5	1
C2	11	3	3	3,0	1
C3	12	5	2	3,5	0
C4	19	5	3	4,0	1
C5	22	5	3	4,0	1
C6	23	2	2	2,0	1
C7	30	4	4	4,0	1
C8	4	4	2	3,0	1
C9	41	4	4	4,0	1
C10	42	2	3	2,5	1
C11	44	5	4	4,5	1
C12	48	4	4	4,0	1
C13	50	4	3	3,5	1
C14	6	4	3	3,5	1
C15	7	3	2	2,5	1
C16	9	4	3	3,5	1
E1	14	5	2	3,5	0
E2	15	2	2	2,0	1
E3	17	2	3	2,5	1
E4	18	5	4	4,5	1
E5	2	2	2	2,0	1
E6	20	4	4	4,0	1
E7	24	3	3	3,0	1
E8	25	2	3	2,5	1
E9	26	2	3	2,5	1
E10	33	3	2	2,5	1
E11	36	3	2	2,5	1
E12	37	2	2	2,0	1
E13	39	2	2	2,0	1
E14	40	2	2	2,0	1
E15	45	2	3	2,5	1
E16	46	3	3	3,0	1
Mean C		3,8	2,9	3,3	15
Mean E		2,8	2,6	2,7	15
			Total C	66%	94%
			Total E	54%	94%

145

Results on a 5-point scale for the word *mix* at posttest

Group	Code	Rater A	Rater C	Mean	Items used
C1	10	3	3	3,0	1
C2	11	5	4	4,5	1
C3	12	5	5	5,0	1
C4	19	5	4	4,5	1
C5	22	5	3	4,0	1
C6	23	5	4	4,5	1
C7	30	5	5	5,0	1
C8	4	5	4	4,5	1
C9	41	4	3	3,5	1
C10	42	4	3	3,5	1
C11	44	5	3	4,0	1
C12	48	5	3	4,0	1
C13	50	5	3	4,0	1
C14	6	3	3	3,0	1
C15	7	5	4	4,5	1
C16	9	3	3	3,0	1
E1	14	5	4	4,5	1
E2	15	4	3	3,5	1
E3	17	4	3	3,5	1
E4	18	4	4	4,0	1
E5	2	2	1	1,5	1
E6	20	5	4	4,5	1
E7	24	2	2	2,0	1
E8	25	4	3	3,5	1
E9	26	3	2	2,5	1
E10	33	4	3	3,5	1
E11	36	3	3	3,0	1
E12	37	2	2	2,0	1
E13	39	2	3	2,5	1
E14	40	5	4	4,5	1
E15	45	4	3	3,5	1
E16	46	3	2	2,5	1
Mean C		4,5	3,6	4,0	16
Mean E		3,5	2,9	3,2	16
			Total C	80%	100%
			Total E	64%	100%

146

Appendix I. Three-rater regrouped mean scores on a 9-point scale for the sentences *She is always sick,* *Luke is strong, and I think he is a fool* **at pre- and posttests.**

Stimulus: *She is always sick*   Time: pretest

Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items Used
C1	6	4	3	4	3,7	0
C2	4	8	5	4	5,7	0
C3	7	5	8	3	5,3	0
C4	9	7	6	6	6,3	0
C5	10	8	6	5	6,3	0
C6	11	7	7	7	7,0	1
C7	12	4	6	4	4,7	1
C8	19	4	3	5	4,0	0
C9	23	6	6	7	6,3	0
C10	22	5	6	5	5,3	1
C11	30	5	5	6	5,3	1
C12	41	3	3	6	4,0	0
C13	42	3	3	4	3,3	0
C14	44	4	5	3	4,0	0
C15	48	7	7	7	7,0	1
C16	50	7	5	8	6,7	0
E1	2	5	4	3	4,0	0
E2	15	7	5	5	5,7	0
E3	17	4	4	5	4,3	1
E4	18	5	6	6	5,7	1
E5	14	3	4	2	3,0	0
E6	20	4	4	7	5,0	0
E7	24	6	5	5	5,3	1
E8	25	4	5	6	5,0	1
E9	26	4	4	5	4,3	1
E10	33	4	6	3	4,3	0
E11	36	4	4	3	3,7	0
E12	37	6	6	6	6,0	1
E13	39	5	7	4	5,3	0
E14	40	5	7	7	6,3	0
E15	45	4	3	6	4,3	0
E16	46	3	2	6	3,7	0
Mean C		5,4	5,3	5,3	5,3	5
Mean E		4,6	4,8	4,9	4,8	6
				Total C	59%	31%
				Total E	53%	38%
native	F1	9	9	9	9,0	1
native	F2	9	8	9	8,7	1
native	M1	9	8	9	8,7	1
native	M2	9	9	9	9,0	1
				Mean	8,9	4
					Total	100%

Stimulus: *She is always sick.*   Time: posttest

Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	4	4	6	4,7	1
C2	4	8	6	6	6,7	0
C3	7	6	3	5	4,7	0
C4	9	8	2	6	5,3	0
C5	10	5	5	3	4,3	0
C6	11	6	4	6	5,3	1
C7	12	6	6	6	6,0	1
C8	19	5	3	4	4,0	0
C9	23	5	4	6	5,0	1
C10	22	6	5	5	5,3	1
C11	30	4	3	4	3,7	0
C12	41	4	6	4	4,7	1
C13	42	3	4	2	3,0	0
C14	44	4	3	3	3,3	0
C15	48	8	7	6	7,0	0
C16	50	7	5	7	6,3	0
E1	2	4	1	6	3,7	0
E2	15	3	8	5	5,3	0
E3	17	5	6	6	5,7	1
E4	18	7	7	6	6,7	0
E5	14	7	4	4	5,0	0
E6	20	3	3	5	3,7	0
E7	24	4	3	3	3,3	0
E8	25	5	5	5	5,0	1
E9	26	4	4	2	3,3	0
E10	33	5	4	2	3,7	0
E11	36	5	4	5	4,7	1
E12	37	4	2	5	3,7	0
E13	39	4	4	3	3,7	0
E14	40	3	7	7	5,7	0
E15	45	4	7	6	5,7	0
E16	46	6	6	6	6,0	1
Mean C		5,6	4,4	4,9	5,0	6
Mean E		4,6	4,7	4,8	4,7	4
				Total C	56%	38%
				Total E	52%	25%

148

Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	5	2	3	3,3	0
C2	4	4	2	2	2,7	0
C3	7	6	5	2	4,3	0
C4	9	4	1	2	2,3	0
C5	10	5	4	3	4,0	0
C6	11	5	4	1	3,3	0
C7	12	5	3	3	3,7	0
C8	19	3	1	1	1,7	1
C9	23	5	3	3	3,7	0
C10	22	5	6	4	5,0	1
C11	30	8	5	7	6,7	0
C12	41	4	1	2	2,3	0
C13	42	7	6	6	6,3	0
C14	44	7	6	3	5,3	0
C15	48	6	3	1	3,3	0
C16	50	6	6	4	5,3	1
E1	2	6	1	1	2,7	0
E2	15	5	4	3	4,0	0
E3	17	4	1	1	2,0	0
E4	18	5	5	3	4,3	0
E5	14	5	6	3	4,7	0
E6	20	5	4	1	3,3	0
E7	24	7	2	5	4,7	0
E8	25	5	3	2	3,3	0
E9	26	6	2	1	3,0	0
E10	33	5	1	3	3,0	0
E11	36	3	1	1	1,7	1
E12	37	4	2	1	2,3	0
E13	39	5	3	2	3,3	0
E14	40	5	2	1	2,7	0
E15	45	6	4	4	4,7	1
E16	46	4	3	1	2,7	0
Mean C		5,3	3,6	2,9	4,0	3
Mean E		5,0	2,8	2,1	3,3	2
				Total C	44%	19%
				Total E	37%	13%
native	F1	9	9	9	9,0	1
native	F2	9	9	9	9,0	1
native	M1	9	9	9	9,0	1
native	M2	9	8	9	8,7	1
				Mean	8,9	4
				Total		100%

Stimulus: *Luke is strong.*　　　Time: posttest.

Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	6	5	3	4,7	0
C2	4	3	4	1	2,7	0
C3	7	9	7	8	8,0	1
C4	9	5	3	4	4,0	0
C5	10	5	3	3	3,7	0
C6	11	5	4	4	4,3	1
C7	12	6	5	2	4,3	0
C8	19	5	6	5	5,3	1
C9	23	4	4	1	3,0	0
C10	22	5	5	3	4,3	0
C11	30	8	3	6	5,7	0
C12	41	7	5	5	5,7	0
C13	42	5	5	4	4,7	1
C14	44	8	6	7	7,0	0
C15	48	7	4	4	5,0	0
C16	50	7	4	5	5,3	0
E1	2	3	5	2	3,3	0
E2	15	5	2	4	3,7	0
E3	17	7	5	7	6,3	0
E4	18	6	5	3	4,7	0
E5	14	5	6	6	5,7	1
E6	20	7	7	6	6,7	1
E7	24	5	2	1	2,7	0
E8	25	6	4	5	5,0	1
E9	26	5	4	3	4,0	0
E10	33	5	4	4	4,3	1
E11	36	6	6	4	5,3	1
E12	37	4	1	1	2,0	0
E13	39	5	4	3	4,0	0
E14	40	4	2	1	2,3	0
E15	45	4	7	9	6,7	0
E16	46	5	2	5	4,0	0
Mean C		5,9	4,6	4,1	4,9	4
Mean E		5,1	4,1	4,0	4,4	5
				Total C	54%	25%
				Total E	49%	31%

150

Stimulus: *I think he is a fool.*   Time: pretest.

Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	7	3	6	5,3	0
C2	4	7	6	4	5,7	0
C3	7	6	5	5	5,3	1
C4	9	4	7	7	6,0	0
C5	10	8	6	8	7,3	0
C6	11	4	1	4	3,0	0
C7	12	5	3	6	4,7	0
C8	19	5	2	4	3,7	0
C9	23	5	4	8	5,7	0
C10	22	4	6	7	5,7	0
C11	30	6	5	6	5,7	1
C12	41	5	4	4	4,3	1
C13	42	5	1	3	3,0	0
C14	44	4	4	3	3,7	0
C15	48	4	3	5	4,0	0
C16	50	5	4	5	4,7	1
E1	2	4	1	3	2,7	0
E2	15	6	6	6	6,0	1
E3	17	5	2	7	4,7	0
E4	18	3	1	2	2,0	1
E5	14	7	4	6	5,7	0
E6	20	7	2	2	3,7	0
E7	24	6	5	7	6,0	0
E8	25	5	7	6	6,0	0
E9	26	6	4	3	4,3	0
E10	33	4	5	5	4,7	1
E11	36	4	3	4	3,7	0
E12	37	7	3	4	4,7	0
E13	39	5	5	6	5,3	1
E14	40	5	4	7	5,3	0
E15	45	8	7	4	6,3	0
E16	46	3	2	3	2,7	1
Mean C		5,3	4,0	5,3	4,9	4
Mean E		5,3	3,8	4,7	4,6	5
				Total C	54%	25%
				Total E	51%	31%
native	F1	9	9	9	9,0	1
native	F2	9	9	9	9,0	1
native	M1	9	8	9	8,7	1
native	M2	9	8	9	8,7	1
				Mean	8,8	4
					Total	100%

Stimulus: *I think he is a fool.*   Time: posttest.

Group	Code	Rater 1	Rater 2	Rater 3	Mean	Items used
C1	6	5	2	4	3,7	0
C2	4	7	4	4	5,0	0
C3	7	6	3	7	5,3	0
C4	9	5	2	8	5,0	0
C5	10	6	6	7	6,3	0
C6	11	5	4	7	5,3	0
C7	12	4	6	6	5,3	1
C8	19	6	3	6	5,0	0
C9	23	7	2	5	4,7	0
C10	22	6	3	5	4,7	0
C11	30	5	4	4	4,3	1
C12	41	6	6	7	6,3	0
C13	42	7	7	9	7,7	1
C14	44	4	3	4	3,7	0
C15	48	7	5	7	6,3	0
C16	50	5	4	4	4,3	1
E1	2	5	2	2	3,0	0
E2	15	6	6	6	6,0	1
E3	17	5	7	8	6,7	0
E4	18	4	3	2	3,0	0
E5	14	6	7	7	6,7	1
E6	20	2	2	5	3,0	0
E7	24	4	5	4	4,3	1
E8	25	4	2	5	3,7	0
E9	26	5	3	4	4,0	0
E10	33	4	5	7	5,3	0
E11	36	4	3	6	4,3	0
E12	37	4	3	7	4,7	0
E13	39	5	3	7	5,0	0
E14	40	5	3	7	5,0	0
E15	45	7	6	6	6,3	0
E16	46	6	6	7	6,3	0
Mean C		5,7	4,0	5,9	5,2	4
Mean E		4,8	4,1	5,6	4,8	3
				Total C	58%	25%
				Total E	53%	19%

152

COMPREHENSIBILITY TEST

TRAINING SPANISH SPEAKERS IN THE PERCEPTION AND PRODUCTION OF ENGLISH VOWELS

Grisel María García Pérez

CODE_____

Listen carefully. You will hear ONE word. Each time you hear this word, your first task is to circle one number from 1 to 5, where 1 is very difficult to understand and 5 is very easy to understand. For example:

You hear word Mix and circle the following:

a)   **1**   2   3   4   5   ☐

b)   1   2   **3**   4   5   ☐

c)   1   2   3   4   **5**   ☐

You are telling me that:

. the first time you heard the word, it was very difficult to understand

. the second time, you understood what the speaker said

. the third time, it was very easy to understand what the speaker said

Your second task is to write an X on the square provided when your judgement is somehow affected by the quality of the recording, for example: too noisy

You will hear the sentence SHE IS ALWAYS SICK.

1.   1   2   3   4   5   ☐

2.   1   2   3   4   5   ☐

3.   1   2   3   4   5   ☐

4.   1   2   3   4   5   ☐

5.   1   2   3   4   5   ☐

6.   1   2   3   4   5   ☐

7.   1   2   3   4   5   ☐

8.   1   2   3   4   5   ☐

9.   1   2   3   4   5   ☐

10.   1   2   3   4   5   ☐

11.   1   2   3   4   5   ☐

12.   1   2   3   4   5   ☐

13.   1   2   3   4   5   ☐

14.   1   2   3   4   5   ☐

15.   1   2   3   4   5   ☐

16.   1   2   3   4   5   ☐

17.   1   2   3   4   5   ☐

18.   1   2   3   4   5   ☐

19.   1   2   3   4   5   ☐

20.   1   2   3   4   5   ☐

21.   1   2   3   4   5   ☐

22.   1   2   3   4   5   ☐

23.   1   2   3   4   5   ☐

24.   1   2   3   4   5   ☐

25.   1   2   3   4   5   ☐

26.   1   2   3   4   5   ☐

27.   1   2   3   4   5   ☐

28.   1   2   3   4   5   ☐

29.   1   2   3   4   5   ☐

30.   1   2   3   4   5   ☐

31.   1   2   3   4   5   ☐

156

32.	1	2	3	4	5
33.	1	2	3	4	5
34.	1	2	3	4	5
35.	1	2	3	4	5
36.	1	2	3	4	5
37.	1	2	3	4	5
38.	1	2	3	4	5
39.	1	2	3	4	5
40.	1	2	3	4	5
41.	1	2	3	4	5
42.	1	2	3	4	5
43.	1	2	3	4	5
44.	1	2	3	4	5
45.	1	2	3	4	5
46.	1	2	3	4	5

47.	1	2	3	4	5	☐
48.	1	2	3	4	5	☐
49.	1	2	3	4	5	☐
50.	1	2	3	4	5	☐
51.	1	2	3	4	5	☐
52.	1	2	3	4	5	☐
53.	1	2	3	4	5	☐
54.	1	2	3	4	5	☐
55.	1	2	3	4	5	☐
56.	1	2	3	4	5	☐
57.	1	2	3	4	5	☐
58.	1	2	3	4	5	☐
59.	1	2	3	4	5	☐
60.	1	2	3	4	5	☐
61.	1	2	3	4	5	☐

62.	1	2	3	4	5	☐
63.	1	2	3	4	5	☐
64.	1	2	3	4	5	☐
65.	1	2	3	4	5	☐
66.	1	2	3	4	5	☐
67.	1	2	3	4	5	☐
68.	1	2	3	4	5	☐

Appendix K. Two-rater regrouped mean scores on a 5-point scale for the sentences *Luke is strong* and *I think he is a fool* at pre- and posttest and for the sentence *She is always sick* at posttest.

Stimulus: *She is always sick*   **Time: posttest**

Group	Code	Rater A	Rater C	Mean	Items used
C1	6	3	3	3,0	1
C2	4	4	3	3,5	1
C3	7	3	2	2,5	1
C4	9	3	3	3,0	1
C5	10	3	3	3,0	1
C6	11	3	3	3,0	1
C7	12	3	3	3,0	1
C8	19	3	3	3,0	1
C9	23	3	3	3,0	1
C10	22	3	3	3,0	1
C11	30	2	2	2,0	1
C12	41	3	3	3,0	1
C13	42	3	2	2,5	1
C14	44	2	2	2,0	1
C15	48	3	4	3,5	1
C16	50	3	3	3,0	1
E1	2	2	2	2,0	1
E2	15	3	2	2,5	1
E3	17	3	3	3,0	1
E4	18	3	2	2,5	1
E5	14	3	4	3,5	1
E6	20	3	2	2,5	1
E7	24	3	3	3,0	1
E8	25	3	4	3,5	1
E9	26	2	3	2,5	1
E10	33	3	3	3,0	1
E11	36	2	4	3,0	1
E12	37	2	2	2,0	1
E13	39	3	3	3,0	1
E14	40	3	3	3,0	1
E15	45	2	3	2,5	1
E16	46	3	3	3,0	1
Mean C		2,9	2,8	2,9	16
Mean E		2,7	2,9	2,8	16
			Total C	58%	100%
			Total E	54%	100%

Stimulus: *Luke is strong*   **Time: pretest**

Group	Code	Rater A	Rater C	Mean	Items used
C1	6	2	2	2,0	1
C2	4	2	2	2,0	1
C3	7	4	3	3,5	1
C4	9	2	2	2,0	1
C5	10	2	3	2,5	1
C6	11	3	3	3,0	1
C7	12	2	3	2,5	1
C8	19	2	2	2,0	1
C9	23	1	2	1,5	1
C10	22	3	4	3,5	1
C11	30	4	4	4,0	1
C12	41				
C13	42	2	3	2,5	1
C14	44	3	3	3,0	1
C15	48	2	3	2,5	1
C16	50	4	4	4,0	1
E1	2	2	3	2,5	1
E2	15	4	3	3,5	1
E3	17	3	1	2,0	1
E4	18	3	3	3,0	1
E5	14	4	3	3,5	1
E6	20	3	3	3,0	1
E7	24	4	3	3,5	1
E8	25	3	2	2,5	1
E9	26	1	1	1,0	1
E10	33	3	3	3,0	1
E11	36	1	1	1,0	1
E12	37	2	3	2,5	1
E13	39	2	3	2,5	1
E14	40	3	2	2,5	1
E15	45	3	3	3,0	1
E16	46	3	2	2,5	1
Mean C		2,5	2,9	2,7	15
Mean E		2,8	2,4	2,6	16
			Total C	54%	100%
			Total E	52%	100%
Native	W1	5	5	5,0	1
Native	W2	5	5	5,0	1
Native	M1	5	5	5,0	1
Native	M2	5	5	5,0	1
			Mean	5,0	4
				Total	100%

Stimulus: *Luke is strong*   **Time: posttest**

Group	Code	Rater A	Rater C	Mean	Items used
C1	6	3	3	3,0	1
C2	4	2	1	1,5	1
C3	7	5	4	4,5	1
C4	9	3	3	3,0	1
C5	10	3	4	3,5	1
C6	11	2	3	2,5	1
C7	12	3	3	3,0	1
C8	19	4	4	4,0	1
C9	23	2	4	3,0	1
C10	22	3	3	3,0	1
C11	30	4	4	4,0	1
C12	41	4	4	4,0	1
C13	42	3	4	3,5	1
C14	44	3	3	3,0	1
C15	48	3	3	3,0	1
C16	50	3	3	3,0	1
E1	2	2	2	2,0	1
E2	15	3	4	3,5	1
E3	17	4	4	4,0	1
E4	18	3	3	3,0	1
E5	14	3	4	3,5	1
E6	20	3	3	3,0	1
E7	24	3	4	3,5	1
E8	25	3	3	3,0	1
E9	26	3	4	3,5	1
E10	33	3	3	3,0	1
E11	36	2	2	2,0	1
E12	37	2	4	3,0	1
E13	39	3	3	3,0	1
E14	40	2	2	2,0	1
E15	45	2	5	3,5	1
E16	46	2	3	2,5	1
Mean C		3,1	3,3	3,2	16
Mean E		2,7	3,3	3,0	16
			Total C	64%	100%
			Total E	60%	100%

Stimulus: *I think he is a fool.*   **Time: pretest**

Group	Code	Rater A	Rater C	Mean	Items used
C1	6	3	2	2,5	1
C2	4	2	4	3,0	1
C3	7	3	2	2,5	1
C4	9	3	3	3,0	1
C5	10	3	4	3,5	1
C6	11	2	2	2,0	1
C7	12	3	2	2,5	1
C8	19	3	2	2,5	1
C9	23	2	3	2,5	1
C10	22	3	2	2,5	1
C11	30	4	4	4,0	1
C12	41				
C13	42	2	3	2,5	1
C14	44	2	2	2,0	1
C15	48	2	3	2,5	1
C16	50	3	3	3,0	1
E1	2	2	2	2,0	1
E2	15	3	3	3,0	1
E3	17	2	2	2,0	1
E4	18	2	1	1,5	1
E5	14	4	2	3,0	1
E6	20	3	2	2,5	1
E7	24	3	2	2,5	1
E8	25	4	3	3,5	1
E9	26	3	3	3,0	1
E10	33	2	2	2,0	1
E11	36	2	3	2,5	1
E12	37	3	2	2,5	1
E13	39	2	3	2,5	1
E14	40	3	2	2,5	1
E15	45	3	3	3,0	1
E16	46	2	2	2,0	1
Mean C		2,7	2,7	2,7	15
Mean E		2,7	2,3	2,5	16
			Total C	54%	100%
			Total E	50%	100%
Native	W1	5	5	5,0	1
Native	W2	5	5	5,0	1
Native	M1	5	5	5,0	1
Native	M2	5	5	5,0	1
			Mean	5,0	4
			Total		100%

Stimulus: *I think he is a fool.*   **Time: posttest**

Group	Code	Rater A	Rater C	Mean	Items used
C1	6	3	3	3,0	1
C2	4				
C3	7	3	3	3,0	1
C4	9	3	3	3,0	1
C5	10	3	4	3,5	1
C6	11	3	3	3,0	1
C7	12	3	2	2,5	1
C8	19	3	3	3,0	1
C9	23	3	3	3,0	1
C10	22	3	2	2,5	1
C11	30	3	3	3,0	1
C12	41	3	3	3,0	1
C13	42	3	3	3,0	1
C14	44	2	2	2,0	1
C15	48	3	3	3,0	1
C16	50	3	3	3,0	1
E1	2	2	2	2,0	1
E2	15	3	1	2,0	1
E3	17	3	3	3,0	1
E4	18	3	2	2,5	1
E5	14	3	3	3,0	1
E6	20	2	2	2,0	1
E7	24	3	2	2,5	1
E8	25	3	2	2,5	1
E9	26	2	2	2,0	1
E10	33	3	3	3,0	1
E11	36	3	2	2,5	1
E12	37	2	2	2,0	1
E13	39	2	4	3,0	1
E14	40	2	3	2,5	1
E15	45	4	3	3,5	1
E16	46	2	3	2,5	1
Mean C		2,9	2,9	2,9	15
Mean E		2,6	2,4	2,5	16
			Total C	58%	94%
			Total E	50%	100%