A STUDY OF THE EFFECTS OF INDIVIDUAL DIFFERENCES IN WORKING MEMORY CAPACITY AND SYNCHRONOUS COMPUTER MEDIATED COMMUNICATION IN A SECOND LANGUAGE ON SECOND LANGUAGE ORAL PROFICIENCY DEVELOPMENT

By

JONATHAN SCOTT PAYNE

A thesis submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY
(Individual Interdisciplinary)

WASHINGTON STATE UNIVERSITY
The Graduate School

May 2000

To the Faculty of Washington State University:

The members of the Committee appointed to examine the thesis of JONATHAN SCOTT PAYNE find it satisfactory and recommend that it be accepted.

Chair

ACKNOWLEDGEMENT

I would like to express my sincere thanks to my advisor, Dr. Paul Whitney, and the other members of my committee for all of their guidance and support during this process. I appreciate the support given to me by the Department of Foreign Languages and Literatures during the data collection phase of my study. I would also like to thank the two excellent instructors of Spanish who participated in my research project, Juan Jimémez and Jennifer Wittenberg. Most of all, I would like to thank my wife, Claudia, for all of her patience and hard work in supporting me during my doctoral studies. Without her, none of this would have been possible.

A STUDY OF THE EFFECTS OF INDIVIDUAL DIFFERENCES IN WORKING MEMORY CAPACITY AND SYNCHRONOUS COMPUTER MEDIATED COMMUNICATION IN A SECOND LANGUAGE ON SECOND LANGUAGE ORAL PROFICIENCY DEVELOPMENT

Abstract

by Jonathan Scott Payne, Ph.D.
Washington State University
May 2000

Chair: Paul Whitney

This manuscript consists of two papers employing concepts from working memory and Levelt's (1989) model language production from cognitive psychology in an effort to better understand processes involved in second language production. The first paper outlines Levelt's model and explains how when expanded to incorporate concepts from working memory this model can effectively explain second language production processes. The second paper extends and tests these conclusions in an empirical context. Over a 15-week period, an experiment involving 58 participants tested the hypothesis that synchronous computer mediated communication in a second language can indirectly improve oral proficiency by developing the same cognitive mechanisms underlying spontaneous conversational speech in a second language. A second hypothesis tested in this study was that individual differences in working memory capacity can effectively predict the rate of L2 oral proficiency development for different types of learners in a chatroom setting. Findings confirmed the first hypothesis suggesting that synchronous on-line conferencing can indirectly improve speaking ability in a second language. Findings pertaining to the second hypothesis indicated that the constraints placed on learners of working memory capacity in oral production may be reduced in the chatroom setting.

TABLE OF CONTENTS

GENERAL INTRODUCTION

The manuscripts submitted in fulfillment of the dissertation requirement of the Graduate School represent an interdisciplinary effort at cross-fertilization between two disciplines: cognitive psychology and second language acquisition. Over the decades, cognitive psychology has accumulated a vast empirical literature and developed effective procedures and measurement techniques for studying linguistic behavior. The research agenda in second language acquisition, on the other hand, has been heavily influenced by Noam Chomsky's theory of linguistic knowledge and has sought to determine the source of second language knowledge, not language usage. Even on the applied side of second language acquisition research much of the focus has been on studying language input as an impetus for acquisition (Krashen, 1985). Until fairly recently (Swain, 1985), the role that output might play in second language acquisition has received less empirical attention.

The two papers in this manuscript employ Levelt's (1989) language production model augmented by concepts from working memory to explain the processes involved in second language production. In the first paper, Levelt's model is described and previous adaptations of this model to the second language context are discussed. It is argued that individual differences in working memory capacity can account for much of the linguistic behavior exhibited by lower and intermediate learners of another language. The second paper applies these two language processing models to explicate and make a priori predictions about the oral proficiency development of learners in a technology mediated second language learning environment.

Developing a better understanding of the cognitive mechanisms underlying second language performance can benefit researchers studying linguistic behavior in diverse settings and teachers seeking to develop effective instructional treatments that meet the needs of language learners. With instructional models that deviate from the conventional classroom setting, having a fundamental understanding of what types of interaction can engender acquisition appear to be even more pertinent. The goal of the this manuscript is to provide researchers and practitioners a new lens through which to view second language acquisition with the hope that this insight may benefit learners in their difficult task of learning another language.

Speaking Another Language: Combining Working Memory Theory and Levelt's Language Production Model to Explain L2 Speech

Learning to converse in another language is a complex endeavor. At lower and intermediate levels of proficiency, speaking a second language is very demanding, requiring each interlocutor to coordinate multiple language-related cognitive tasks: parsing incoming language streams, mapping this input onto existing syntactic and semantic structures in memory, deciding on an appropriate response, choosing the lexical items required to express that intent, selecting the correct syntactical elements needed to convey those communicative intentions, constructing the corresponding phonological plan for articulation, and then actually uttering this string of events in the form of speech (Levelt, 1989). This amalgam of processes must be orchestrated in concert, with very little margin for error. As if that were not enough, it is also necessary to keep track of the conversational context or what is referred to as the discourse record (Clark, 1996). This meta-level awareness is needed to ensure that utterances are appropriate to the current discussion (Grice, 1975), and to assist interlocutors in anticipating the direction the conversation is headed.

There is a body of literature addressing the cognitive processes of second language or bilingual speech production (see Dechert, Möhle, & Raupach, 1984; de Groot & Kroll, 1997, for examples). Unfortunately, few process models exist for describing L2 speech and guiding empirical inquiry. Employing process models has the distinct advantage of allowing researchers to make specific predictions about the performance of second language speakers under specific task requirements. In the L1 literature, Levelt's model of language production (1989; 1993; 1995) has received the most empirical attention and has been the basis of most models of L2 or bilingual language production processes (De Bot, 1992; De Bot & Schreuder, 1993; Poulisse & Bongaerts, 1994). The goal of this paper is to discuss Levelt's model of language production (1989, 1995) and show how when augmented by working memory theory, this process model may help guide performance-oriented second language acquisition research and explain some individual differences among L2 learners. Levelt's model, though considered a production model, is in many respects a conversation model. In fact, one iteration of the model has the comprehension side fully developed (Levelt, 1993). For the purposes of this paper, the focus will be on the original production model (Levelt, 1989).

Levelt's Production Model

The model described by Levelt (1989, see Figure 1) has four main components: the Conceptualizer, the Formulator, the Articulator, and the Speech Comprehension System (SCS). Utterances begin as a preverbal message or non-language specific thought in the Conceptualizer. The job of the Conceptualizer is to determine the semantic content of a spoken utterance and temporarily store that plan in working memory. The preverbal message generated by the Conceptualizer enters the Formulator where the conceptual plan is transformed into a linguistic plan. In the first stage of this process, the semantic units or lemmas that map onto the preverbal message are selected and encoded grammatically to form the syntactical surface structure of the utterance. The second task of the Formulator is to select the phonological representations or the lexemes for the selected lemmas in a process called phonological encoding. The resulting plan for articulation is stored in the Articulatory Buffer for the purpose of internal self-monitoring in the SCS. While the articulatory plan is being monitored internally, a phonological representation of that plan is stored in the buffer. If the articulatory plan that emerges from the Formulator after a round of monitoring matches the intended message, then the speech motor functions are engaged to produce the utterance as overt speech.

Modular and incremental processing are important constraints on the stages of Levelt's model. That is, once the Formulator has begun to access lemmas for the contents of the preverbal message, it is not possible to verify the semantic value of selected lemmas by checking back with the Conceptualizer. Nor is it possible for the Formulator to pass any information forward to the Articulator. Independent operation (modularity) within the stages of the model along with the sequential progression from stage to stage is what makes parallel processing possible. In other words, while one word is being uttered, the lemma and lexeme for another word is being selected, and in the Conceptualizer the speaker is still deciding what words will follow. In fact, as speakers we often begin expressing an idea before we have even determined how we are going to end it. This is what is meant by *incremental* in the model.

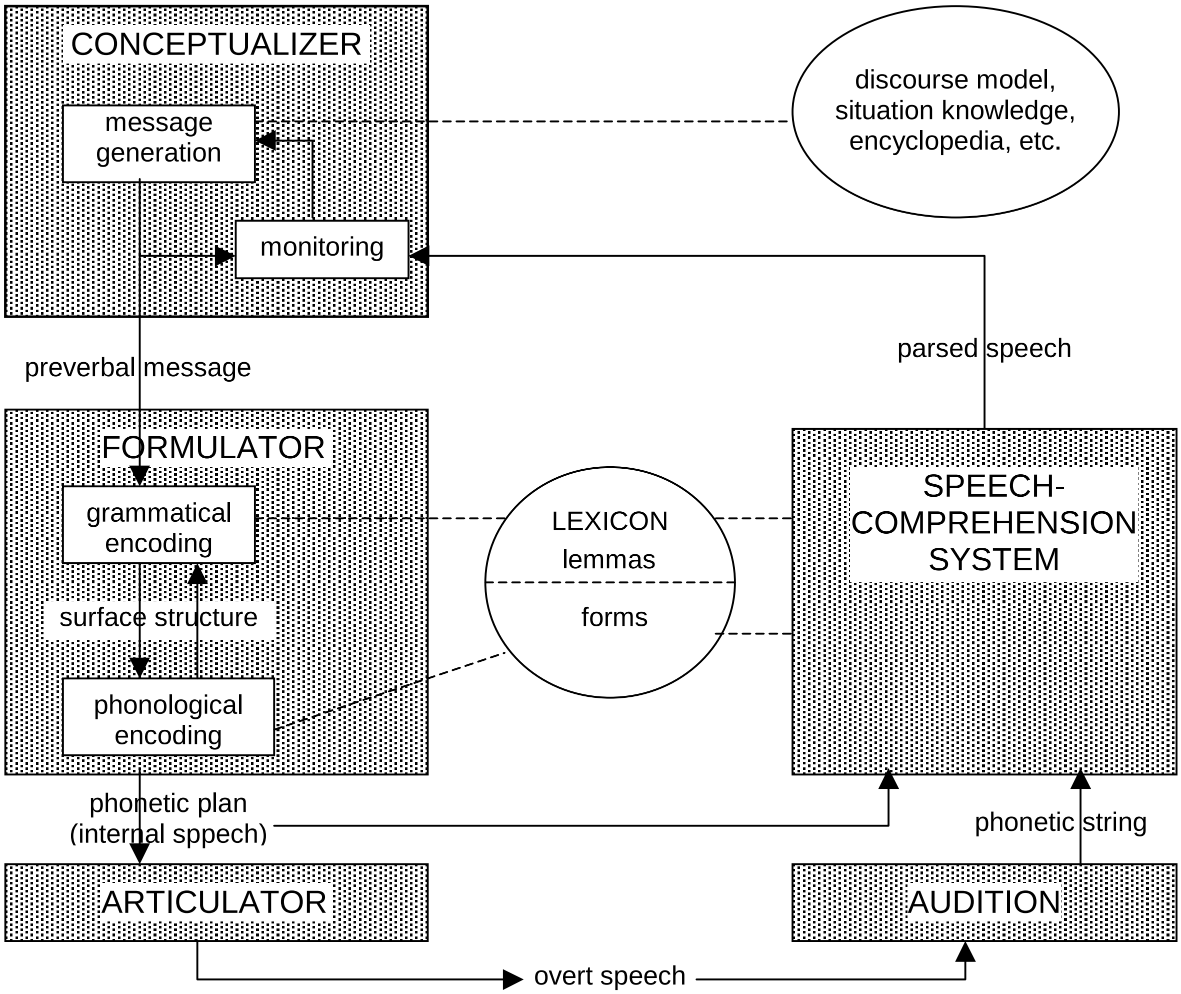

Figure 1: A blueprint for the speaker. Boxes represent processing components; circle and ellipse represent knowledge stores (Levelt, 1989, p. 9).

Levelt's Model and Bilingual Language Production

Adaptations of the model to illustrate bilingual language production processes have included augmenting the model with language-specific Formulators as a means of explaining fluent code-switching (de Bot, 1992). While the idea of speakers generating multiple parallel speech plans simultaneously appears reasonable, the proposal conflicts with Levelt's idea of modularity (Poulisse, 1997). De Bot claimed that the language of the utterance was selected in the Conceptualizer. However, such a proposal leaves fluent code-switching unexplained. If the preverbal message contains instructions specifying language for the Formulator, then it is not apparent how the speaker would be able to construct parallel speech plans. That is, not only would two Formulators be needed, but also two preverbal messages.

Later this proposal was abandoned in favor of an additional component called the Verbalizer located between the Conceptualizer and the Formulator (De Bot and Schreuder, 1993). The function of the Verbalizer is to chunk semantic data from the Conceptualizer. This addition was deemed necessary to account for a lack of one-to-one correspondence between words and semantic concepts. For instance, how should the Formulator know how to encode *Exit the library*, *Go out of the library*, or *Leave the library* from information in the preverbal message indicating that some individual has removed herself from the premises of the library?

A third adaptation of Levelt's ideas about language production (Poulisse and Bongaerts, 1994) didn't include de Bot and Schreuder's Verbalizer component, rather it borrowed the spreading activation approach employed by Green (1986) to explain how bilingual speakers can often so fluently switch back and forth between languages. Spreading activation as it relates to lexical access in speech purports that each lemma has language nodes. For example, a native English speaker learning German as a foreign language would have the lemmas *Tisch* and *table* associated with the concept of a piece of furniture with a flat surface parallel to the floor and having 3 or 4 legs. The table concept would have a node for German and a node for English connecting both lemmas to the same construct. With spreading activation theory, the node receiving the most activation is the one that is selected. If the L2 is being used, then the German node will receive more activation and *Tisch* will be selected and not *table*. However, since there is a link between the two lemmas, intentional or unintentional code-switching is possible.

Working Memory and Levelt's Model

Levelt's model and its adaptations have been useful for investigating L2 production processes, but neither the original model, nor bilingual versions can account for how individual differences in working memory may impact bilingual production. Levelt acknowledges the importance of the short-term storage of information in language production, but this aspect of the model remains undeveloped (1989, p.21). Working memory provides researchers with models and measurement techniques for determining an individual's capacity for temporarily maintaining verbal and visual-spatial information in memory, and for performing judgement or executive functions based changing conditions in one's immediate environment. Baddeley's (1986) notions of the phonological loop and the central executive will be combined and contrasted with Daneman and Carpenter's (1980) individual differences approach to modeling working memory.

The processes associated with lexical access (Formulator), articulation (Articulator), and parsing incoming speech (SCS) in L1 are thought to be largely automatic. Controlled processing is limited to the Conceptualizer where decisions about communicative intentions are made, comprehension of overt and internal speech occurs (Levelt, 1989, p. 21). Second language production on the other hand, presents a very different picture. Lexical access and articulation in a second language appear to be dominated more by controlled rather than automatic processing. If the assumption can be made that the bulk of language processing done by beginning to intermediate level speakers of a second language is controlled, then working memory theory should inform our understanding of how all four components of Levelt's model may perform under this different set of constraints.

Working Memory and the Conceptualizer

The Conceptualizer serves the purpose of planning what to say, deciding how to say it, keeping track of what has been said, and verifying that an utterance in progress fits with your communicative intentions and the conversation as a whole (Levelt, 1989). As previously stated, accomplishing these tasks successfully requires the focused attention of the speaker and implies an integral role for Working Memory.

Research findings on pausal phenomena may provide insight into the processes inherent in conceptualizing. Hesitations in the L1 can account for between 5% to 65% of speaking time depending on the speaker, and comprise an estimated 40% to 50% of speaking time during interviews and when people are asked to describe a simple scene (Goldman-Eisler, 1968). Hesitations may also account for a variety of linguistic-related behaviors. Deliberations in speech often occur before the first content word in a phrase or sentence and may reflect an individual's lexical access ability (Daneman & Green, 1986). Tentative behavior on the part L2 speakers could also reflect efforts to buy time for processing by slowing down the pace of discussion.

In a series of studies by the Kassel Group (Dechert, 1980, 1983; Dechert , Möhle, & Raupach, 1984; Dechert & Raupach, 1980a, 1980b, 1987) looking at pausal phenomena in L1 and L2 speech samples of German, French, and English, the level of processing was found to be the major difference between fluent and non-fluent L2 learners during periods of hesitation. For less fluent learners, attention seems to be focused on planning at the sentence level (micro-planning). Pauses in speech among fluent L2 speakers represent reflections on integration and the elaboration of main communicative goals with a series of sub-goals (macro-planning); much like the pausal behavior of native speakers (Schmidt, 1992). These findings suggest that the demands placed on working memory by less fluent L2 speakers may differ qualitatively and most likely quantitatively from more fluent L2 speakers. Less fluent speakers of a second language may expend a great deal of their attentional resources performing the various tasks related to conceptualizing.

For individuals with lower L2 proficiency, just comprehending input and keeping track of the discourse record constitutes a major challenge. Findings from Daneman and Green (1986) may shed some light on the role of Working Memory in comprehension and maintaining the discourse record. First, it should be noted that the individual differences approach to studying Working Memory employed by Daneman and colleagues is somewhat divergent from Baddeley's three-component conception. The individual differences approach views working memory storage capacity as a function of processing ability, with inefficient processors having less storage because they must devote more resources to processing functions. While working memory research testing Baddeley's model employs techniques for studying short-term memory performance for different input modalities and under various memory load conditions, the individual differences approach looks at working memory on a more global level by studying working memory function in complex cognitive tasks.

The instrument used in Daneman and Green's study was the reading span measure (Daneman and Carpenter, 1980). The reading span presents subjects with sets of sentences, ranging from 2 to 5 sentences, one at a time for a period of 7 seconds. Subjects are required to read aloud each sentence in a set remembering the last word of each sentence. Sentences are arranged on index cards and shown to the subject one at a time. As soon as a subject reads a sentence, the next sentence card is placed on top of the first and so on until the end of the set. After all sentences in a set have been viewed, subjects must recall the last word of each sentence.

The hypothesis tested by Daneman and Green (1986) was that an individual's ability to exploit contextual clues in text is in part due to that person's ability to efficiently coordinate the processing and storing functions of working memory. This hypothesized role of working memory was operationalized in this study by testing the prediction that working memory capacity would account for individual differences in the ability to learn vocabulary items from context. The use of context for learning may be the underlying mechanism accounting for high correlations among language-related measures of vocabulary knowledge, reading comprehension, and verbal intelligence (Sternberg & Powell, 1983).

To test the use of context to support vocabulary acquisition, subjects read a passage approximately 225 words in length and containing a word selected from dictionaries of rare and outlandish words (Byrne, 1974; Dickson, 1982). These words and their meanings (e.g. qualtagh - "the first person seen after leaving the house" or ghoom - "to hunt in the dark") were highly unusual making this task an excellent test of the ability to draw clues to a word's meaning from the surrounding context. The subjects were instructed to read each passage and to guess the meaning of the word based on the context. A pretest procedure was also done to ensure that none of the subjects knew any of the words used in the test. Each definition given by the subjects was scored on a 0 to 4 scale reflecting 1.) the preciseness of meaning and 2.) the degree of inference extending beyond a simple summarization of the contextual clues. The results from this experiment showed that reading span was strongly correlated with the ability to derive word meanings from context, $r(28) = .69$, $p < .01$.

Viewing these results from the perspective of Levelt's model, the task of learning new words from context would place a heavy burden on the Conceptualizer. As indicated by the strong correlation between reading span and the number of words that subjects could correctly define from the context, working memory appears to have a central role in conceptualizing. These results were generated from a first language task. In a second language having to use context for acquiring novel vocabulary and deciphering the meaning of textual and aural input is much more critical than in a L1 setting, especially for less proficient L2 learners. In second language instruction, being able to guess the meaning of words from context and to be able to piece together the gist of a text or conversation from words comprehended in isolation are two skills explicitly practiced.

Working Memory and the Formulator

As a starting point towards developing a better understanding L2 production processes, it may be useful to review the L1 literature related to aural production. In line with Bock's (1982) proposal that the development of spoken language ability reflects a progression from controlled to automatic processing of lexemes (phonological) in word production, Adams and Gathercole (1996) have suggested that verbal working memory may play a role in children learning to speak. Their reasoning is based on lexical access theories (Levelt, 1992; Dell & O'Seaghdha, 1992) describing a two step process in retrieving lexical items from memory. First, the lemma or the semantic unit is activated followed by the specifications of the associated phonological form or lexeme. If in the early stages of learning to speak, children process the phonological information for lemmas accessed in a controlled manner, then phonological working memory resources may be expended in order to temporarily maintain a phonological representation of the lemma to support articulation. Thus, the phonological loop may assist in lexeme acquisition in the same manner that it supports vocabulary learning. According to this proposal, once greater speaking fluency is obtained, the phonological loop would become redundant in lending further support to this process.

Several studies have explored the connection between phonological working memory and speech development of young children (Adams & Gathercole, 1995, 1996; Speidel, 1989, 1993). Adams and Gathercole (1996) found verbal working memory span as measured by nonword repetition to be an effective predictor of expressive language performance on the Bus Story (Renfrew, 1969). In this study, ninety-two children were given measures of phonological working memory (nonword repetition, memory span, and articulation rate), non-verbal cognitive ability, and language skills (vocabulary knowledge and expressive language as measured by the average length of the five longest utterances made by the children while re-telling the Bus Story). Employing hierarchical regression, nonword repetition was entered into the regression equation in the fourth step after age, non-verbal ability and the combined vocabulary scores. After the 27.1% of the variance accounted for by the first three steps, nonword repetition was the only factor able to account for a significant amount of additional variance (3.5%, $F(4,84) = 4.28$, $p < 0.05$ for Bus Story information scores and 3.5%, $F(4,84) = 4.17$, $p < 0.05$). The ability to reproduce novel words as measured by nonword repetition or accurately imitate sounds strings of adult speech is what Speidel (1989) proposes forms network connections involved in speech planning.

With first language production, the time requirements of searching for lemmas corresponding to the preverbal message is measured in milliseconds (Daneman & Green, 1986). Selecting the appropriate lemmas for an L2 preverbal message may require a much more extensive search. If access to the desired L2 lemmas is not automated, then resources must be expended to support the computations for translating the L1 equivalent. This fits with Kroll and Stewart's (1994) revised model of concept mediation and word association. According to this model, since L2 meaning is initially accessed via the L1, an unbalanced relationship between L1 and L2 at the lexical level occurs with links from L2 to L1 being relatively strong. The result is that L2 lemmas strongly activate their translated L1 equivalents (Kroll, Michael, &Sankaranarayanan, 1998). For beginning learners of a foreign language the connection between L1 and L2 is so strong that the L2 is almost exclusively accessed through word association or by translating L1 lexical entries. If this is the case, then it poses an interesting scenario for Levelt's model and working memory that differs from previous adaptations of Levelt's model (de Bot and Schreuder, 1993; Poulisse and Bongaerts, 1994). Could it be that beginners conceive of a communicative act in the conceptualizer, store the preverbal message for that intention in working memory, grammatically and phonologically encode the utterance in their L1, hold the plan for that utterance in the articulatory buffer, and then cycle back through the model translating the completed utterance from L1 to L2 prior to articulation? This would explain why beginning L2 learners are so encumbered in speaking, beyond their limited L2 mental lexicon. This proposal could also clarify why speakers of limited proficiency often blurt out their intended message in the L1 when unable to construct the appropriate utterance in the L2. This double-looping scenario would place immense pressure on the verbal and executive functions of working memory.

There is little data on working memory as it relates to L1 – L2 translation and the effects on second language vocabulary acquisition. One study by Nick Ellis and Alan Beaton (1993) investigated the effectiveness of keyword and repetition techniques for developing receptive and productive competence with German vocabulary items. The study employed three experimental groups: subjects utilizing a keyword method, subjects using simple rote repetition, and a third group where subjects used their "own strategy." This investigation sought to determine the effectiveness of keyword and repetition strategies in immediate and long-term recall of vocabulary items. Subjects were asked to translate foreign (German) words and native (English) words. For the keyword technique subjects were instructed "to IMAGINE a specific scene that links the sound of the English and German words together in some way." Prior to the experiment, subjects were asked to rate their own speaking, reading, writing, and listening ability for each language in which they showed proficiency other than English. A computer-based system calculated a score for NLANG - the total number of languages they knew, and FORLANG - the sum of their means scores over the four, language skill measures for each language in which they claimed proficiency other than their native language.

There are two aspects of these findings that are of direct interest. First, repetition proved to be superior to the keyword and own strategy conditions for translation tasks from English into German. The evidence indicates that production versus reception was a key element favoring repetition, as was the direction of translation. These findings neither support nor discount the notion of double-loop processing for L2 learners still using word association for gaining access to the second language. However, the connection between repetition and acquiring new vocabulary through L1 to L2 translation, implies a role for verbal working memory in L1 to L2 translation.

For more advanced L2 interlocutors who are able to access the appropriate lemmas without L1 mediation, it is still necessary to construct the syntax and phonological plan of an utterance. Again, if working memory resources are taken up by other processing tasks, then decrements may arise in these two areas. This may provide an explanation for a phenomenon that all communicative second language teachers observe with delight in conversation classes. When a discussion becomes stimulating for students and is no longer just an excuse for practicing speaking, the participants become focused on meaning. For those individuals wishing to have their opinions heard, lemma selection becomes a primary focus. If the level of discourse is pushing the boundaries a the student's proficiency level, then an imbalance in the distribution of working memory resources occurs between the planning and monitoring duties of the Conceptualizer and the lexical access function of the Formulator. Speakers begin to concentrate more on finding the words necessary to communicate their views. What teachers observe is that students shift their attention away from monitoring their own output and listening to what others have to say, with the result that attention to grammar and pronunciation nose-dives as students focus on getting their point across.

Working Memory and the Articulator

The articulator plays probably the least significant role in L1 production of the four components. The articulator's main function is to engage the musculature of the larynx and respiratory system to form the sounds specified by the phonetic plan in the articulatory buffer (Levelt, 1989, p.12). In the first language, this function occurs most of the time with little attention. In a second language containing novel sounds more attention must be given to producing overt speech. It is unclear as to the added burden that concentrating on pronunciation may place on working memory. It may imply that simply increasing the time required to produce overt speech by manipulating the vocal apparatus to produce newer sounds results in decay of the phonetic plan or at least requires more resources to maintain it.

Working Memory and the Speech Comprehension System

The job of the SCS is to parsing incoming speech, essentially performing the same functions of the Formulator, but in a decoding mode and in the reverse sequence (see Levelt, 1993). The first step of the SCS is to phonologically decode the acoustic signal. The second step is to search for the corresponding lemmas using the syntactical structure of the sentence and semantic context. It seems reasonable that phonological working memory or the phonological loop would play a significant supporting role in speech comprehension. Most of the first language research on phonological working memory in comprehension has focused on the need to temporarily store representations of language input for "off-line" processing during comprehension. The common ground among several theories of first language comprehension (Clark & Clark, 1977; Kintsch & Van Dijk, 1978; Howard & Butterworth, 1989) is that understanding clauses and sentences with simple syntax proceeds on-line. This means that long-term memory is accessed directly, thus by-passing working memory. However, as linguistic complexity increases, semantic and syntactic analysis must move off-line. During off-line processing, working memory temporarily stores a representation of the sentence for use in higher-level linguistic interpretation and for recognizing words later in the sentence (Gathercole & Baddeley, 1993). One view (Baddeley, Vallar & Wilson,1987) purports that verbal working memory acts as a "mnemonic window" for storing chunks during sentence processing. The limited capacity of the phonological loop becomes crucial when the chunk of information to be maintained exceeds phonological loop capacity, resulting in comprehension deficits. An alternative view sees verbal working memory as playing a role in storing the intermediate products of syntactic analysis (Caramazza & Berndt, 1985), whereas Martin (1987) has suggested that the phonological loop may function as a buffer for holding spoken language that was heard during periods of off-line processing for future analysis. These three views implicate phonological working memory at three different "passes" in processing sentence content. What is becoming the most widely held view among cognitive neuropsychologists is that phonological working memory is involved in sentence processing after syntactic analysis has occurred, but before the sentence has been fully interpreted (Gathercole & Baddeley, 1993). Based on their research with short-term memory patients, Waters, Caplan & Hildebrandt (1991) claim that the comprehension problems of people with impaired verbal working memory are due to a limited ability to compare the semantic interpretations derived from major lexical items (i.e. content words) within a sentence with the interpretations achieved from full syntactic analysis.

Applying these findings to Levelt’s model in a second language context may produce a different picture. For speakers of low to intermediate L2 proficiency levels it is likely that a large portion sentences heard are linguistically complex, making off-line semantic and syntactic analysis the norm. Of the three views presented above, all of them seem very plausible in the second language context. The “mnemonic window” view of Baddeley et al. (1987) makes sense in light of Kroll and Stewart’s (1994) revised hierarchical model as previously mentioned. Holding “chunks” of phonological code may be necessary to support comprehension, especially if the semantic value of lemmas is mediated via the L1 (concept mediation) and even more so if a translation is required (word association). Caramazza and Berndt’s (1985) proposal combined with Baddeley and colleagues could model the phonological loop as a store that initially holds strictly the sound structure of speech, but as parsing progresses the content of the phonological store becomes elaborated by syntactic and semantic attributes until moving on to the Conceptualizer. Of the three proposals, Martin's (1987) suggestion has immediate appeal for any lower-level foreign language learner. When attempting to comprehend speech in a foreign language, the listener must constantly be regulating how much time to spend deliberating on an incomprehensible word or phrase with the need to continue "capturing" the on-going conversation. Focusing too much on deciphering a word may result in no longer being able to follow the conversation, whereas not trying to comprehend novel words will limit understanding of what was said and stagnate Interlanguage development. Examining these three views of the role of phonological working memory in second language context would make a considerable contribution to the theory and practice of learning to understand another language.

Studies of L2 comprehension have chosen a different set of independent variables: speech rate as determined by pauses at constituent boundaries, compressed speech, and slowed speech (Griffiths 1990a, 1990b, 1991a, 1991b; Rader, 1990; Conrad, 1989; Long, 1985; Blau, 1990; Barshi & Healy, 1998), syntactic complexity (Blau, 1990), and the number of propositional units (Barshi & Healy, 1998). Findings from studies related to the rate of speech have been conflicting. Griffiths argued that below-average speech rates would facilitate instruction, but Rader's (1990) study found no significant differences in the comprehension of listening passages delivered at normal speed and at one of two slowed-down rates. Blau (1990) found that syntactic complexity did not affect comprehension, but pausing at constituent boundaries aided comprehension for lower and intermediate level listeners. Slowed speech rate was found to have no effect except with participants at very low proficiency levels.

In a study investigating the influence of propositional units on second language comprehension, Barshi & Healy (1998) tested the listening comprehension of 6 native English speakers and 12 nonnative speakers of English in three different experiments. Participants took a listening test that was based on the listening comprehension section of the TOEFL test. For this test they heard 10 sentences. For each sentence heard, participants selected the one sentence from four choices that was closest in meaning to the sentence heard. In each of the experiments, participants played the role of a pilot in a computerized test simulating the listening tasks of pilots when communicating with air-traffic controllers. The task was to click on the correct squares in the four, 4x4 grids visible on the computer screen based on auditory instructions delivered via the computer in different lengths and at different speech rates.

In all three experiments and for all three fluency levels of English, the number of propositions was determined to play the most decisive role in comprehension. When the number of propositions per message exceeded three, comprehension dropped significantly. Speech rate as measured by the length of pauses between words or the length of the words themselves, and the number of words in a proposition were not found to have any significant effect on comprehension, even among the low-fluency nonnative speakers of English. In the discussion, the authors noted that performance improved with practice and that the break-off point of 3 propositions parallels previous findings (Barshi, 1997a, 1997b) studying this task with native speakers. The fact that native speakers and low-fluency nonnative speakers could perform at the same level on a comprehension task is curious. How could nonnative speakers, even of low-intermediate proficiency levels, show the same comprehension patterns of native speakers? As mentioned by Barshi and Healy in their literature review, Conrad (1989) reported a comparison of L1 and L2 speakers listening to recordings of simple English sentences presented at varying rates of compression. The findings were that native speakers tended to focus more on key content words in the stimulus than nonnative speakers. This content word strategy may be what is responsible for equivalent results between L1 and L2 speakers the the Barshi and Healy study. If both groups are filtering input for key content words, then the nonnative speakers are effectively reducing the burden on working memory by engaging in much less syntactic processing and by reducing the overall amount of information that must be maintained.

The idea that how language input is processed can differ qualitatively based on proficiency level in a second language and between native speakers and nonnative speakers is emerging as a recurrent theme. Service (1987) proposes that native speakers make use their knowledge of language structure to make predictions about the direction the conversation is going. This amounts to a form of top-down hypothesis testing based on overlearned representations of the language. Since interpreting the speech signal comes at only a minor cost to working memory resources, listeners can focus most of their attentional resources on understanding the message. Service continues to say that this process is much different for the second language learner. Learners of an L2 at low and intermediate levels of language proficiency may employ more of a bottom-up analysis, starting their analysis with the phonological structure of the language stream (Service, 1987).

Differences between L1 and L2 speech on a micro-level are also pronounced. L2 speech tends to be more hesitant with longer and more frequent pauses, consist of shorter utterances, and contain many more slips of the tongue than L1 speech (Poulisse, 1997; Weise, 1984; Möhle, 1984; Lennon, 1990). As second language speakers become more fluent, speech rate and length of run increase, and the number of filled and unfilled pauses decrease (Lennon, 1990). The assumption is that fluency is a direct function of automatic language processing ability. Since controlled processing implicates working memory, limitations in working memory capacity should have an impact on L2 performance and consequently acquisition. Incidentally, many of the same dependent measures for this L2 speech research (e.g. articulation rate, pause length, length of run, slips of the tongue) have been the focus of working memory research in first language development. Most of these studies have investigated the role of phonological working memory capacity in the spoken language and vocabulary development of young children (Adams & Gathercole, 1995, 1996; Gathercole & Baddeley, 1989) including some studies of second language development in children (Speidel, 1989, 1993; Service, 1992; Service, & Kohonen, 1995). Findings from this line of research show that articulation rate among children between the ages of 4-7 is directly related to their phonological working memory capacity. In other words, the larger the capacity for temporarily storing and maintaining sound information in memory, the faster a child at an intermediate stage in language development is able to talk.

For adults speaking their first language, Daneman and Green (1986) developed a speaking span test that measures working memory capacity in production based on the maximum number of sentences a subject can generate from a list of unrelated words with the task of producing one sentence incorporating each word from a list of words. The second experiment of this study examined the hypothesis that "a speaker’s ability to produce an appropriate lexical item on-line is related to that speaker’s ability to coordinate the processing and storage functions of working memory (Daneman & Green, 1986)." Results from the speaking span test ($r(32) = .60$, $p < .01$) and a contextual vocabulary production task revealed that individuals with small speaking spans were less fluent at accessing context-appropriate lexical items and were slower in producing the words selected. Another interesting aspect of this study was the high correlations ($r(18) = -.71$, $p < .01$ for both) between in-context lexical codibility, context strength, and production fluency. This indicates that context increases the pool of candidates for selection thereby increasing fluency in production.

While Daneman and Green (1986) only depicted a relationship between speaking span and fluency at the word level, Daneman (1991) extended this line of inquiry to include production fluency and creativity with more complex usage, semantically and syntactically. The speaking span was also used in this study together with tasks eliciting speech generation, oral reading performance, and oral slips in the form of spoonerisms. The speaking span results were scored as either *strict* (the exact word must be contained in a grammatical sentence) or *lenient* (any version of the target word could be used as long as the sentence was grammatical). In the speech generation task, subjects were presented with a picture and asked to talk about the picture for 1 minute. Fluency was measured by the total number of words produced in the allotted time and a richness rating on a 1 (repetitious, semantically empty) to 5 (creative, semantically rich) scale. For the oral reading task subjects were asked to read a 320-word excerpt from *The Great Gatsby* by Scott Fitzgerald. Subjects were instructed to read the passage as fast as they could, though not so fast as to slur words and render the text unintelligible to the listener. From the tape-recorded protocols the number of errors for each category (repetition, false start, mispronunciation, addition, omission, and substitution) was tallied for each subject. Daneman's predictions that the strict speaking span score would be a better predictor of measures of fluency demanding accuracy and that the lenient speaking span score would predict more creative, open-ended language use were borne out in the results. Speaking span (lenient) showed a strong correlation with the number of words generated in the speech generation task and their richness in content $r(27) = .48$, $p < .01$ and $r(27) = .47$, $p < .01$, respectively. The speaking span (strict) demonstrated a strong relationship with the oral reading and oral slip tasks; both of which required accurate production.

These findings point to an interesting dissociation between speaking styles which in the SLA literature are labeled as "risk-takers" and "monitors." Risk-takers are those individuals who value participating in communicative exchange over always producing grammatically correct language. Monitors, on the other hand, focus much more attention on the grammaticality of their utterances and may forego the opportunity to speak if they are unsure about how to formulate their intentions in accordance with the target language grammar. As was found in Daneman (1991), those individuals who scored high on the lenient speaking span (risk-takers) demonstrated superior fluency and greater richness in their output, but at a cost to accuracy as determined by the oral reading and oral slip tasks. The "monitors," or those individuals scoring higher on the strict speaking span measure, were clearly less fluent than their counterparts, but performed better on the tasks where accuracy was important.

Implications

In light of the current paucity of process models for guiding applied research second language production, Levelt's model (1989) is promising. Augmenting Levelt's language production model with concepts from working memory can increase the variety of L2 phenomena that researchers can explain. Working memory research is currently enjoying the limelight in cognitive psychology. The constructs introduced in this paper are general in nature. The cutting-edge of L1 working memory research is the fractionization of working memory in an effort to develop a more detailed understanding of dissociations about separate working memory processes. It is apparent that working memory plays a central role in all forms of cognition, however the information processing demands placed on the second language learner suggest that bilingual processing may be an optimal test for current working memory models. This could account for the interest among cognitive psychologists in studying second language acquisition as evidenced by numerous recent publications (for a complete edition see Healy and Bourne, 1998).

For anyone involved in second language instruction, the question is how can this new knowledge inform practice? There are a couple of ways in which a better understanding of how working memory capacity affects language use may be helpful for teachers. First, being cognizant of burden placed on working memory by different types of activities can assist instructors in selecting and sequencing activities. This may help teachers understand why some activities may be too overwhelming for students at one level, but work well with another group of individuals at the same of a slightly different level.

Secondly, knowing more about working memory limitations can help us as foreign language professionals more effectively address the needs of individual learners. Activities and learning environments can be identified that dampen the impact of working memory limitations on learners and made integral components in the foreign language curriculum.

Bibliography

Adams, A.M. & Gathercole, S.E. 1995. Phonological working memory and speech production in preschool children. *Journal of Speech and Hearing Research* 38: 403-414.

Adams, A.M. & Gathercole, S.E. 1996. Phonological working memory and spoken language development in young children. *Quarterly Journal of Experimental Psychology* 49A(1): 216-233.

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A. D., Vallar, G., & Wilson, B. (1987). Sentence comprehension and phonological memory: Some neuropsychological evidence. In M. Coltheart (Eds.), Attention and performance, XII (pp. 509-529). Hove, UK: Lawrence Erlbaum Associates Ltd.

Baddeley, A.D., Thomson, N., & Buchanan, M. 1975. Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 9: 176-189.

Barshi, I. (1997a). Effect of message length in ATC communication: Linguistic and cognitive aspects. Paper presented at the Ninth International Symposium on Aviation Psychology, Columbus, OH.

Barshi, I. (1997b). Effects of linguistic properties and message length on misunderstandings in aviation communication. Unpublished doctoral dissertation, University of Colorado, Boulder.

Blau, E. K. (1990). The effect of syntax, speed, and pauses on listening comprehension. TESOL Quarterly, 24, 746-753.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, 89, 1-47.

Brooks, L.R. 1967. The suppression of visualisation by reading. *Quarterly Journal of Experimental Psychology* 19: 289-299.

Byrne, J. H. (1974). *Mrs. Byrne's dictionary of unusual, obscure, and preposterous words*. Secaucus, NJ: University Books.

Caramazza, A. & Berndt, R. S. (1985). A multicomponent view of agrammatic Broca's aphasia. In M.-L. Kean (Eds.), *Agrammatism*. New York: Academic Press.

Clark, H. H. (1996). *Using language*. New York : Cambridge University Press.

Clark, H. H., & Clark, E. (1977). *Psychology and language. An introduction to psycholinguistics*. New York: Harcourt Brace.

Conrad, L. R. (1989). The effects of time-compressed speech on native and EFL listening comprehension. Studies in Second Language Acquisition, 11, 1-16.

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior* , 19(4), 450-466.

Daneman, M. & Green, I. 1986. Individual differences in comprehending and producing words in context. *Journal of Memory and Language* 25: 1-18.

Daneman, M. 1991. Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research* 20 (6): 445-464.

De Bot, K. 1992. A bilingual production model: Levelt's 'Speaking' model adapted. *Applied Linguistics* 13 (1): 1-24.

De Bot, K. & Schreuder, R. 1993). Word production and the bilingual lexicon. In R. Schreuder & B. Weltens Eds.), *The bilingual lexicon* pp. 191-214). Amsterdam: John Benjamins.

Dechert, H. W. (1980). Pauses and intonation as indicators of verbal planning in second-language speech productions: Two examples from a case study. In H. W. Dechert & M. Raupach (Eds.), Temporal variables in speech (pp. 271-285). The Hague: Mouton.

Dechert, H. W. (1983). How a story is done in a second language. In C. Faerch & G. Kaspar (Eds.), *Strategies in interlanguage communication*. London: Longman.

Dechert, H. W., Möhle, D., & Raupach, M. (Eds.). (1984). *Second language productions*. Tübingen: Narr.

Dechert, H. W., Raupach, M. (Eds.). (1980a). Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler. The Hague: Mouton.

Dechert, H. W., & Raupach, M. (Eds.). (1980b). *Towards a cross-linguistic assessment of speech production*. Frankfurt a.M.: Lang.

Dechert, H. W., & Raupach, M. (Eds.). (1980). *Psycholinguistic models of production*. Norwood, NJ: Ablex.

De Groot, A.M.B. & Kroll, J.F. (1997). Tutorials in bilingualism: Psycholinguistic perspectives. Mahwah, NJ: Lawrence Erlbaum Associates.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. Cognition, 42, 287-314.

Dickson, P. (1982). *Words*. New York: Delacorte Press.

Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology*, 46A(3), 533-558.

Engle, R., Kane, M. & Tuholski, S. 1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake and P. Shah (Eds.), *Models of Working Memory: mechanisms of active maintenance and executive control* pp. 102-134). Cambridge: Cambridge University Press.

Gathercole, S.E. & Baddeley, A.D. 1993. *Working memory and language*. Hove: Lawrence Erlbaum Associates.

Gathercole, S.E. & Baddeley, A.D. 1989c. The role of phonological memory in vocabulary acquisition: A study of young children learning new names.

Gathercole, S.E. & Baddeley, A.D. 1989. Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Jounal of Memory and Language* 28: 200-213.

Geva, E. & Ryan, E.B. (1993). Linguistic and cognitive correlates of academic skills in first and second languages. *Language Learning*, 43, 5-42.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), Syntax and semantics: 3. Spreech acts. New York: Academic Press.

Griffiths, R. (1990a). Facilitating listening comprehension through rate-control. *RELC Journal: A Journal of Language Teaching and Research in Southeast Asia*, 21, 55-65.

Griffiths, R. (1990b). Speech rate and NNS comprehension: A preliminary study in time-benefit analysis. *Language Learning: A Journal of Applied Linguistics*, 40, 311-336.

Griffiths, R. (1991a). Language classroom speech rates: A descriptive study. TESOL Quarterly, 25, 189-194.

Griffiths, R. (1991b). Pausological research in an L2 context: A rationale, and review of selected studies. Applied Linguisitics, 12, 345-364.

Healy, A. & Bourne, L. (1998). *Foreign language learning: Psycholinguistic studies on training and retention* Mahwah, NJ: Lawrence Erlbaum Associates.

Howard, D. & Butterworth, B. (1989). Short-term memory and sentence comprehension: A reply to Vallar and Baddeley, 1987. *Cognitive Neuropsychology*, 6, 455-463.

Kintsch, W. & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. Journal of Memory and Language, 33, 149-174.

Kroll, J. F., Michael, E., & Sankaranarayanan, A. (1998). A model of bilingual representation and its implications for second language acquisition. In A. Healy & L. Bourne*, Foreign language learning: Psycholinguistic studies on training and retention* (pp. 139-164). Mahwah, NJ: Lawrence Erlbaum Associates.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387-417.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: The MIT Press.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. Cognition, 42, 1-22.

Levelt, W.J.M. (1993). Language use in normal speakers and its disorders. In G. Blanken, J. Dittmann, H. Grimm, J. Marshall, & C. Wallesch (Eds*.), Linguisitic disorders and pathologies: An international handbook* (pp. 1-15). Berlin: de Gruyter.

Levelt, W.J.M. (1995). The ability to speak: from intentions to spoken words. *European Review*, 3 (1), 13-23

Long, M. (1985). Input and second language acquisition theory. In S. Gass & C. Madden (Eds.), Input and second language acquisition (pp. 377-393). Rowley, MA: Newbury House.

Martin, R. C. (1987). Articulatory and phonological deficits in short-term memory and their relation to syntactic processing. Brain and Language, 32, 159-192.

Miyake, A. & Friedman, N.. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. Healy & L. Bourne*, Foreign* *language learning: Psycholinguistic studies on training and retention* (pp. 139-164). Mahwah, NJ: Lawrence Erlbaum Associates.

Möhle, D. (1984). A comparison of the second language speech production of different native speakers. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), Second language production (pp. 26-49). Tübingen: Narr.

Papagno, C., Valentine, T., & Baddeley, A.D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, 30, 331-347.

Poulisse, N. & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, 15, 36-57.

Poulisse, N. (1997). Language production in bilinguals. In A.M.B. de Groot & J. F. Kroll (Eds.) Tutorials in bilingualism. Mahwah, NJ: Lawrence Erlbaum Associates.

Rader, K.E. (1990). The effects of three levels of word rate on listening comprehension of third-quarter university Spanish students. Unpublished doctoral dissertation. Ohio State University, Columbus.

Renfrew, C.E. (1969). *The bus story: A test of continuous speech*. Oxford: Author.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14, 357-385.

Service, E. (1987). Applying the concept of working memory to foreign language listening comprehension. Equity in Language Learning. Proceedings of the Nordic Conference of Applied Linguistics 5th, Jyvaskyla, Finland, ERIC ED 343 392.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology*, 45A, (1), 21-50.

Service, E. & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, 16, 155-172.

Speidel, G.E. 1993. Phonological short-term memory and individual differences in learning to speak: A bilingual case study. *First Language* 13: 69-91.

Speidel, G.E. 1989. A biological basis for individual differences in learning to speak. in G.E. Speidel and K.E. Nelson (eds.): New York: Springer-Verlag.

Speidel, G.E. 1989. Imitation: A bootstrap for learning to speak? in G.E. Speidel and K.E. Nelson (eds.): New York: Springer-Verlag.

Sternberg, R. J., & Powell, J. S. (1983). Comprehending verbal comprehension. *American Psychology*, 31, 878-893.

Waters, G.S., Caplan, D., & Hildebrandt, N. (1991). On the structure of verbal short-term memory and its functional role in sentence comprehension: Evidence from neuropsychology. Cognitive Neuropsychology, 8, 81-126.

Weise, R. (1984). Language production in foreign and native languages: Same or different. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language production* (pp. 11-25). Tübingen: Narr.

Developing L2 Oral Proficiency through Synchronous CMC: Output, Working Memory, and Interlanguage Development.

On the applied side of second language acquisition (SLA) theory much of the debate over what promotes competence has focused on the role of input in language learning. It has even been argued that input is the greatest sole determiner of language acquisition (Krashen, 1985). However, there is evidence that input alone is not sufficient to obtain high levels of proficiency in a second language. Language immersion programs in Canada provide students with an input-rich learning environment, but equivalent opportunities to produce the target language are often lacking. Research on these immersion programs depicts the learners as highly developed in their receptive language skills while exhibiting weaknesses in grammatical accuracy (Harley, 1993).

Consistent with the hypothesis that output practice is important to competence, Swain (1985, 1993) and Swain and Lapkin (1995) argued that L2 output may trigger certain cognitive processes necessary for second language learning. Swain's proposal of the Output Hypothesis places an emphasis on language learners "noticing" the gaps in their linguistic knowledge as a result of external feedback (clarification requests, modeling, overt correction, etc.) or internal feedback (monitoring) of language they have produced. By becoming consciously aware of ones own language production, output can serve the metalinguistic function of helping to internalize linguistic forms, test hypotheses about the language, and increase control over previously internalized forms.

The Output Hypothesis has sparked numerous studies addressing its components. In the interactionist literature, research has found that learners test hypotheses about the target language and modify their output in response to clarification or confirmation requests by their interlocutors (Pica, Holliday, Lewis & Morgenthaler, 1989). In studying native speaker – nonnative speaker interaction, Linnell (1995) found that clarification requests resulted in more syntax modification on the part of nonnative speakers than modeling correct responses and that those modified (improved) syntactical structures were maintained over time. Findings from research of the construct of “noticing” suggest that second language learners do notice gaps in their Interlanguage knowledge (Swain & Lapkin, 1995). Further research has investigated whether learner awareness of problems in output can prompt the solicitation of additional input (Izumi, et al. 1999).

Unfortunately, process models that could suggest causal mechanisms have not guided research on the role of output in acquisition. Employing process-based working models has the distinct advantage of allowing researchers to make specific predictions about the performance of second language speakers under specific task requirements. In the L1 literature, Levelt's model of language production (1989, 1993, 1995) has received the most empirical attention and is the most widely adapted model for depicting L2 or bilingual language production processes (De Bot, 1992; De Bot & Schreuder, 1993; Poulisse & Bongaerts, 1994). In 1996, de Bot employed Levelt’s model together with Anderson’s (1982) notions of declarative and procedural knowledge as a means of analyzing the notions of the Output Hypothesis from a psycholinguistic perspective. De Bot limited his discussion to lexical access and how it relates to the shift from controlled (declarative knowledge) to automatic processing (procedural knowledge), or a process referred to as restructuring. The crux of de Bot's argument was that output plays a crucial role in the restructuring of linguistic into a procedural form that allows for automatic, efficient performance. However, according to de Bot, output does not play a role in the acquisition of declarative knowledge itself.

The purpose of the present paper, like de Bot's work, is to use Levelt's model as a basis for proposing mechanisms that influence L2 acquisition. However, I will attempt to show that Levelt's model (1989; 1995) augmented with other concepts from information processing, particularly Working Memory theory, can serve as a basis for understanding second language processes beyond those considered by de Bot. Though Levelt's model alone may prove useful for depicting second language production processes, it does not have a way of accounting for individual differences in processing capacity and how they may relate to performance on L2 production tasks.

Levelt (1989, p.21) acknowledges the importance of the short-term storage of information in language production, but this aspect of his model has not been fully developed. Working Memory theory provides researchers with models and measurement techniques for determining an individual's capacity for temporarily maintaining verbal and visual-spatial information in memory, and for performing judgement or executive functions based changing conditions in ones immediate environment. First language research has found that individual differences in Working Memory capacity are closely related to: 1.) verbal fluency (Daneman, 1991), 2.) the ability of individuals to utilize contextual clues in text for learning novel words (Daneman & Green, 1986), and 3.) maintaining a representation of language strings for "off-line" processing when language becomes to complex for "on-line" processing (Gathercole & Baddeley, 1993). Findings from second language studies suggest that verbal Working Memory capacity serves as an effective predictor of L2 vocabulary development (Gathercole & Baddeley, 1989; Papagno, Valentine, & Baddeley, 1991), second language proficiency (Service, 1992; Service & Kohonen, 1995), and it appears to play an even more crucial role in L2 than L1 acquisition (Miyake & Friedman, 1998; Geva & Ryan, 1993).

The goal of this paper is to augment Levelt’s model of language production with Working Memory theory and to use this framework for testing the hypothesis that synchronous computer mediated communication (CMC) or chatting in a second language can indirectly improve oral proficiency by developing the same cognitive mechanisms underlying spontaneous conversational speech. Within the context of this research question, what is currently known about Working Memory and the role that it plays in learning will make it possible to make a priori predictions about whose L2 development will benefit the most from the chatroom environment and why. Before reporting on this study and its results, an explanation of how Levelt’s model and Working Memory theory will support these research goals is in order.

Levelt's Language Production Model

According to Levelt's model (1989; 1995; *see Figure 1*), utterances begin as non-language specific communicative intentions in what Levelt refers to as the Conceptualizer. During production the job of the Conceptualizer is to determine the semantic content of a to-be-spoken utterance. The preverbal message generated by the Conceptualizer is maintained in Working Memory and fed into the Formulator where the lemmas or lexical items are selected that most accurately represent the semantic content of each chunk of the preverbal message. Lemmas also contain the information necessary for formulating syntax and are used to generate the surface structure of an utterance through a process called Grammatical Encoding. The second task of the formulator is to select phonological representations or lexemes for the selected lemmas. What emerges from the Formulator is the articulatory plan of an utterance. However, prior to entering the Articulator, where the vocal musculature is engaged for producing an utterance, the articulatory plan is monitored internally with the support of subvocalization. During this internal feedback loop, the articulatory plan is stored in the Articulatory Buffer (Working Memory).

The stages of Levelt's model operate in a modular and incremental fashion. That is, once the preverbal message has entered the Formulator and the lexical access process has begun, it is not possible for the Formulator to check back with the Conceptualizer to verify the intended meaning of the message. Nor is it possible for the Articulator to be alerted as to processes that are currently underway in the Formulator. When a lemma and its lexeme have been selected, that information leaves the Formulator where the first opportunity to screen output via internal monitoring is possible. The autonomy of operation (modularity) and consecutive progression is what makes parallel processing within Levelt's model possible. In other words, while one word is being uttered, the lemma and lexeme for another word are being selected, and in the Conceptualizer the speaker is still deciding what words will follow.

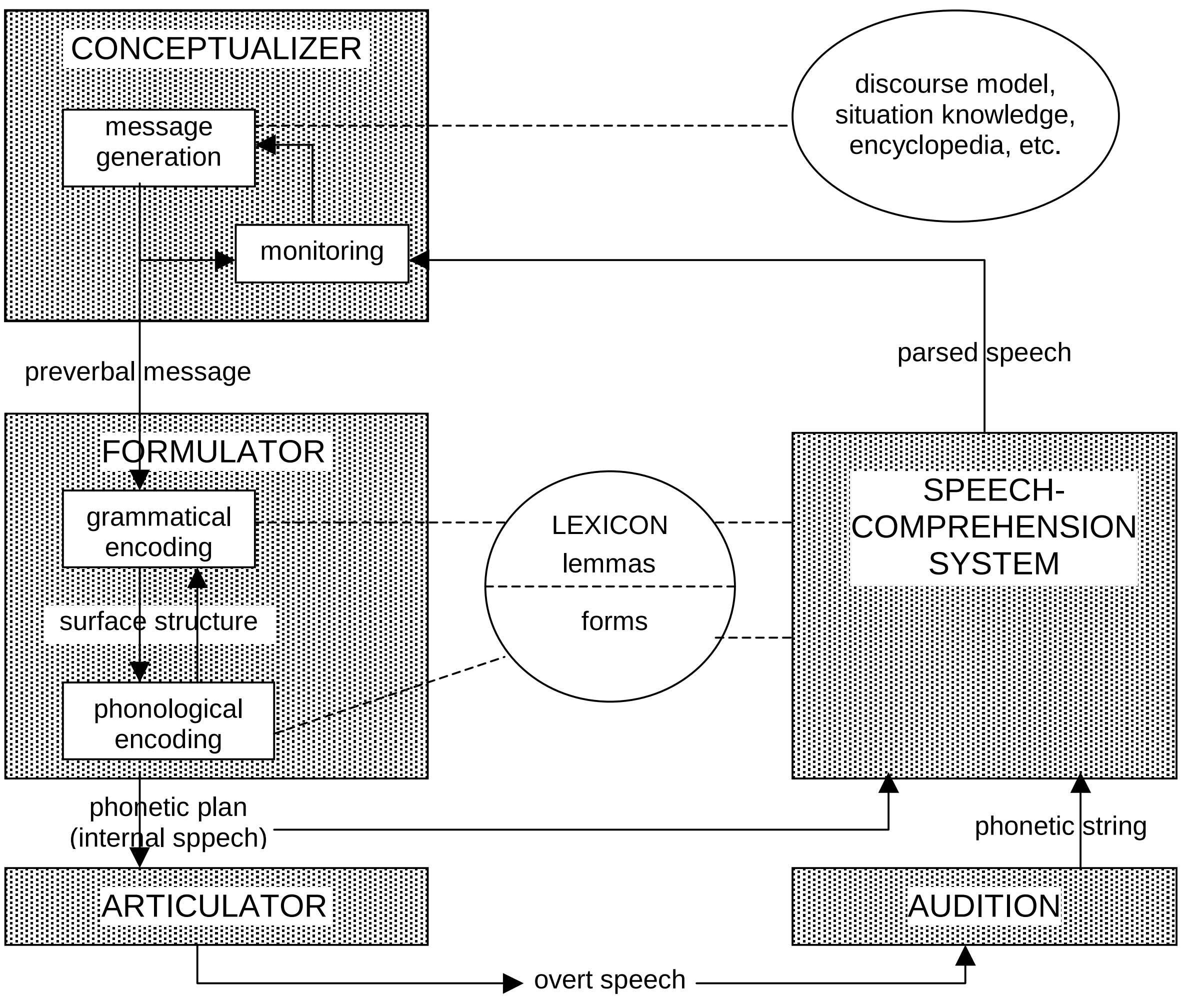

Figure 1: A blueprint for the speaker. Boxes represent processing components; circle and ellipse represent knowledge stores (Levelt, 1989, p. 9).

In fact, as speakers we often begin uttering a thought before we have even determined how we are going to end it. This is what is meant by *incremental* in the model.

Three adaptations of Levelt's model to illustrate bilingual language production processes have been proposed. De Bot (1992) augmented the model with language-specific Formulators in an attempt to explain fluent code-switching behaviors. A year later, de Bot and Schreuder (1993) introduced an additional component called the Verbalizer, located between the Conceptualizer and the Formulator, with the function of organizing information in the preverbal message into lexicalizable chunks. In a third effort, Poulisse and Bongaerts (1994) employed spreading activation theory to explain how preverbal concepts can be tagged for language. It was argued that spreading activation theory obviated the need for adding a component to the model and addressed weaknesses in de Bot's (1992) multiple Formulator approached. These modifications have been proposed to account for code-switching among bilinguals. However, as de Bot (1992) suggested, a bilingual production model must also account for cross-linguistic influences, equivalent language processing speed between mono- and multilinguals, unbalanced bilingualism, and the potential to master an unlimited number of languages. These three adaptations of Levelt's model and the additional bilingual phenomena mentioned by de Bot (1992) point to important questions for bilingual language processing research. Unfortunately, these proposals and suggestions fail to address the need to understand how individual differences in Working Memory capacity may boost or constrain the language processing capabilities of second language learners.

Working Memory and Levelt's Model

As Levelt's model suggests, lexical access and articulation in the L1 are automatic. Controlled processing in the model is limited to the Conceptualizer where communicative intentions are generated, and where internal speech is monitored (Levelt, 1989, p. 21). Second language production on the other hand, is quite different. Controlled processing appears to play a central role in lexical access and articulation in a second language, at least until a high level of proficiency has been achieved. L2 speech tends to be more hesitant with longer and more frequent pauses, consist of shorter utterances, and contain many more slips of the tongue than L1 speech (Poulisse, 1997; Weise, 1984; Möhle, 1984; Lennon, 1990). As second language speakers becomes more fluent, speech rate and length of run increase, and the number of filled and unfilled pauses decrease (Lennon, 1990). The assumption is that fluency is a direct function of automatic language processing ability. Since controlled processing implicates Working Memory, limitations in Working Memory capacity should have an impact on L2 performance and consequently acquisition. Not surprisingly, then, many of the same dependent measures used as indices of competence in L2 speech research (e.g. articulation rate, pause length, length of run, slips of the tongue) have been assessed when researchers have tested the role of Working Memory in first language development. Most of these studies have investigated the role of phonological Working Memory capacity in the spoken language and vocabulary development of young children (Adams & Gathercole, 1995, 1996; Gathercole & Baddeley, 1989). Only a few of these studies have examined second language development in children (Speidel, 1989, 1993; Service, 1992; Service, & Kohonen, 1995). Findings from this line of research show that articulation rate among children between the ages of 4-7 is directly related to their phonological Working Memory capacity. In other words, the larger the capacity for temporary storage and maintenance of sound information in memory, the faster a child at an intermediate stage in language development is able to talk. Pauses during speech have also been linked to lexical access in research with adults. These hesitations most often occur before content words and signal the speaker's need to access items from the mental lexicon with the time required to complete the search as a function of the difficulty of the content word and Working Memory capacity (Daneman & Green, 1986).

In a series of studies, the Kassel Group (Dechert, 1980, 1983; Dechert, Möhle, & Raupach, 1984; Dechert & Raupach, 1980a, 1980b, 1987; Raupach, 1980, 1984; Rehbein, 1987) examined pauses in L1 and L2 speech samples of German, French, and English. The major difference between fluent and non-fluent L2 learners of these languages is the type or level of processing that occurs during periods of hesitation. For less fluent learners, the focus is on lower levels of planning, whereas pauses in speech among fluent speakers represent integration and macroplanning processes, much like the pausal behavior of native speakers (Schmidt, 1992). These findings suggest that the demands placed on Working Memory by less fluent L2 speakers may differ qualitatively and most likely quantitatively from more fluent L2 speakers. Less fluent speakers of a second language may expend a great deal of their attentional resources on retrieving appropriate words from their mental lexicon, determining the correct surface structure or syntax, and selecting the corresponding lexemes or phonological units for the words in the utterance. If these processes are not automatic, a burden is placed on the Phonological Loop (Baddeley, 1986) to maintain the intermediate products of calculations as the speaker cycles through Levelt's model, generating communicative intentions in the Conceptualizer, mapping lexical items and their syntactical and phonological components from the preverbal message, monitoring the utterance internally, and making any needed adjustments. While the phonological loop is storing and maintaining the utterance under construction, the Central Executive (Baddeley, 1986) is making judgements about the correctness of the lemmas selected, the syntax and sound structure of the utterance, what information needs to be retrieved from long-term memory, and what new updated information needs to be put back into the phonological loop for storage. For more fluent speakers, many of these processes occur without much conscious attention, leaving attentional resources for contemplating more subtleties of expression.

Language Production, Working Memory and Synchronous CMC

Only a handful of studies have systematically examined the impact of chatroom environments on L2 performance (Warschauer, 1996; Kern, 1995;

Chun, 1994). We would expect that chatrooms could provide a useful environment for improving some L2 processes. A few studies have looked at how interlocutors resolve break-downs in communication through negotiation of meaning, suggesting that synchronous on-line environments can play a role in Interlanguage development (Linnell, 1995; Pelletierri, 1999; Blake, 1999). In general, studies of L2 chatroom use have found that the dynamics of conversational interaction are altered in an online conferencing environment. Results from these studies have indicated that: 1.) students tend to produce more complex language in chatrooms than in face-to-face conversational settings (Warschauer, 1996; Kern, 1995), 2.) participation increases on-line with "quieter" students participating as much or even more than those individuals who normally dominate classroom discussion (Warschauer, 1996; Kern, 1995; Chun, 1994), and 3.) attitudes towards the target language were reported to improve (Healy-Beauvois, 1992; Warschauer, 1996; Kern, 1995; Chun, 1994).

Given the theoretical discussion of L2 processes covered above, one effect of chatroom practice may be to automate some language production processes and thereby ease the burden on Working Memory. Unfortunately, the impact of individual differences in Working Memory in a synchronous CMC environment has not been explored. Working Memory may prove to be a useful construct for predicting what types of learners will benefit the most from synchronous CMC. There are two characteristics of L2 chatroom interaction that may have implications for Working Memory. First, the rate of conversational exchange in a chatroom is slower than face-to-face. People simply cannot type as fast as they can speak. Thus the processing demand is reduced or more precisely, the amount of language that an individual has to parse, comprehend and respond to is lower for a given time period. Also, chatroom exchanges do not have the same ephemeral quality of speech. When chatting you can refresh memory traces by re-reading comments. This is not the case in aural conversation, face-to-face or otherwise. This would suggest that learners with lower Working Memory capacities would benefit from a conversational environment where processing demands are reduced, but where the tasks and interactions are the same. Thus, another goal of the present study is to determine whether individual differences in Working Memory capacity can effectively predict the rate of L2 oral proficiency development for different types of learners in a chatroom setting.

RESEARCH QUESTIONS

Based on Levelt's model of language production, synchronous on-line conferencing in a second language should develop the same cognitive mechanisms that are needed to produce the target language in face-to-face L2 conversation. In fact, the only difference, from an information processing perspective, should be engaging the musculature for producing overt speech. Furthermore, by augmenting Levelt's model with concepts and measurement techniques from Working Memory theory, several benefits are accrued. Firstly, we can gain insight into how individual differences in processing capacity may affect oral proficiency development. Secondly, being able to make specific a priori predictions about how such limitations can impact performance and which learners may benefit from what types of instructional treatments would be of great use to researchers and curriculum developers. Hence, the following research questions:

1. Can L2 oral proficiency be indirectly developed through chatroom interaction in the target language?
2. Can individual differences in Working Memory capacity effectively predict the rate of L2 oral proficiency development for different types of learners in a chatroom setting?

Expected results are that the oral proficiency development of participants in the experimental group will be at least equivalent with the control group, and possibly even greater since the chatroom environment should reduce the burden on Working Memory, thus facilitating the development of low span participants.

METHODOLOGY

Experimental Design

The study employed a pretest, posttest quasi-experimental design with two sample groups receiving the treatment and two sample groups receiving the face-to-face instruction typical for the language program. The experimental groups participated in 2 face-to-face and 2 on-line class periods per week. A few chatroom days were cancelled during exam periods and to dedicate some extra computer lab time for familiarizing students with the on-line, collaborative research and writing tool. The experimental sections met for a total of 21 times in the chatroom during the 15-week semester. All four days of instruction were face-to-face for the control groups. The instructional content was the same for both the experimental and the control group, thus the same activities or discussions were held on-line in the chatroom that were conducted in the face-to-face classroom. Levels of the treatment could not be randomly assigned to groups due to scheduling issues for the instructors teaching the four courses. The study lasted 15 weeks (one semester). During the second week of the semester, the computerized versions of the reading span measure, nonword repetition task, and the Shipley verbal intelligence measure were administered in a computer laboratory. During the third week and the beginning of the fourth week of the study, the speaking pretest was administered with the posttest occurring during the last week of the study. These measures are described below.

Participants

Participants were 58 volunteers from four sections of third semester Spanish courses. Intact groups were used and the treatment was assigned to the groups in a manner that could accommodate the schedules of the participating instructors (since computer access for the instructors was located in one specific building: this was necessary to avoid forcing them to run back and forth across campus). Each instructor taught one experimental and one control group. Participants received extra-credit totaling a maximum of one-third of a letter grade for participating in the study.

Materials

Currently the most recognized instrument for measuring oral proficiency is the oral proficiency interview (OPI) based on the ACTFL Oral Proficiency Guidelines. This scale ranges from 0-5 with 0 representing no proficiency and 5 representing the oral proficiency of an educated native speaker. This scale was not appropriate for use in this study for two reasons: 1.) the OPI is not sensitive enough to measure changes in oral proficiency that may occur in a single semester and 2.) a significant proportion of the OPI score consists of competencies that are not addressed by this study's research questions (e.g. socio-linguistic competence). It is important to note that the term oral proficiency in this paper is a more simplified construct than is used by ACTFL. Oral proficiency in this context refers to an individual's ability to produce language that is comprehensible, employs appropriate syntax and vocabulary for the task, is grammatically accurate, and is pronounced in a manner that approximates the speech of a native speaker. Therefore, an oral proficiency instrument was developed for this study (see Appendix A). For the speaking test, participants selected one of four envelopes containing a speaking task written in English. The speaking task description was in English in order to ensure that performance on the task was not confounded by reading ability in Spanish. Participants were required to read the instructions and then speak in Spanish for approximately five minutes. If a participant ran out of things to say on a particular topic, they selected a new task and began again. The objective was to obtain a 5-minute speech sample. The role of the examiner was to listen, but not to interview the speaker. Two examiners (1 native speaker and 1 non-native speaker; both were female) administered the speaking tests. The 50-point maximum on the scale represented an educated non-native speaker of Spanish. The examiners were told to think of someone they know who is a very fluent non-native speaker of Spanish and consider that individual's language ability as a perfect score. This differs from the ACTFL scale for the reason that measuring second language learners against a standard that views the language skills of an educated native speaker as the highest rating may be effectively confounding results. The possibility of reaching the level of an educated native speaker in another language could be argued to be an unobtainable task by all but a very small number of people who ever study a foreign language. The examiners were instructed how to use the scale and compared evaluations for the first two or three speaking tests on each testing day to establish interrater reliability. For the oral proficiency pretest, interrater reliability on the 50-point scale was .86. On the posttest, interrater reliability was higher at .94. The examiners were paid $100 each for their efforts.

Working Memory measures consisted of a recognition-based nonword repetition task and a reading span measure. The nonword repetition task measures an individual's capacity to maintain phonological information in Working Memory. This is the most widely used test for measuring verbal Working Memory capacity. There are several variations of the nonword repetition task that have been reported in the literature. In the nonword repetition task developed for this study, participants listened to an audio file of 8 pseudo-words read with a one-second interval between words.

After listening to the audio file, participants clicked on a button to receive a screen containing 16 pseudo-words, 8 of which were articulated in the audio file. Students selected the 8 words they believed to have heard by clicking on the checkbox next to each word. The participants could take as much time as they needed to make their 8 selections. After clicking the submit button, the next audio clip would load, ready to be played. The nonword repetition task consisted of 3 sets of 8 pseudo-words.

The reading span test used in this study is an adaptation of Daneman and Carpenter's (1980) measure used in numerous studies of Working Memory (see Whitney and Budd, 1999). Reading span assesses two key functions of executive Working Memory: 1.) the ability to make judgements and 2.) to temporarily store the results of calculations. The reading span test is also considered a good measure of central executive capacity (Engel, Kane, & Tuholski, 1999). This version of the reading span presents participants with 15 sets of sentences, the first 3 sets containing only 2 sentences each and the final 3 sets consisting of 6 sentences. Each sentence in the set is visible for 7 seconds. While viewing the sentence, subjects are required to determine if it makes sense and to remember the last word of the sentence. Indicating whether the sentence is sensible may be done concurrently with reading the sentence by clicking on one of two radio buttons, "makes sense" or "nonsense." However, recalling the final word in the sentence is performed after the subject has seen all sentences in the set. With only two sentences in a set, combining the judgment and recall tasks is relatively easy. However, as set size increases, more memory resources must be allocated to maintaining the final words of each previous sentence in the set, making the task of judging the sensibility of a sentence being read while maintaining the last words from the sentences previously seen much more difficult.

The computer-based delivery of the reading span measure displayed one sentence after another in 7-second intervals until all of the sentences in a set had been viewed. While reading the sentences, subjects selected the radio button corresponding to their estimate of the sentence's sensibility. After all sentences in the set had been seen, the participant clicked on a button to receive a screen of words with checkboxes next to the words. For each to-be-remembered word there were two distractors (e.g. for sets containing 5 sentences there was a total of 15 words). Distractors were of two types: 1.) the same semantic category (e.g. if the target word was "girl", the distractor could be "woman") or 2.) the last words from sentences in previous sets. Subjects simply selected the words they identified as being final words by clicking the checkbox next to the word. All Working Memory tests were recognition- and web-based with a database back-end, enabling automatic scoring and calculating of the results.

Data on student grades at the conclusion of the third semester Spanish course, their overall GPAs, and verbal intelligence, as measured by the Shipley test of verbal intelligence, were also collected. This academic and verbal IQ data was used to account for other variables that could confound interpretation of results.

Treatment

One of the challenges of conducting quasi-experimental research with intact groups is the issue of unequal treatment or of a "teacher effect." To ensure that the treatment received by participants in the experimental and control conditions was equivalent, the curriculum and lesson plans for all four groups were the same. This meant that the experimental groups meeting online in the chatroom engaged in the same activities on the same days as the control groups did face-to-face. The chatroom tool designed for this project enabled the instructors to read and participate in up to four chatrooms simultaneously. During the pilot, it was determined that the optimal sized chatroom discussion group was 4-6 students. With larger groups, active participation causes the chat window to scroll too fast for students to be able to follow and process the conversation. Foreign language classes at the institution where the study was conducted typically range from 18-22 students, so using four chatroom groups per class was optimal. The same activities and group configurations were also used in the face-to-face sessions. In fact, the instructors actually printed out the task description from the chatroom interface for use in their face-to-face groups. Tasks assigned on the days when the experimental groups were online consisted of role-plays, discussions of cultural texts or video, and other communicative activities. The first two chatroom sessions were held in the Language Learning Resource Center (LLRC) giving students the opportunity to familiarize themselves with the chatroom tool and ask any questions that they might have. After these initial training sessions, most participants did not come to the LLRC, but rather accessed the chatroom from their home computers or machines in other computer labs on campus. Those participants who continued coming to the LLRC during the study either did not own a computer or lived too far from campus to return home for a one-hour class. In fact, the participants were encouraged not to be online in the same physical location as their classmates in an effort to make their online "conversation" the only form of synchronous exchange to occur in the target language during the scheduled class hour. The largest number of students seen at one time chatting in the LLRC was never more than 4 and during most sessions only 2 students were in the same 20 station lab. This location-independent design is important because it represents a significant difference from the majority of studies investigating the intersection of synchronous CMC and second language acquisition.

Another aspect of the treatment employed in this experiment was a curriculum design that sought to control for a possible Hawthorne effect related to technology use by fully integrating technology in the form of learning systems and on-line course management features into all participating groups. Both experimental and control groups used these tools and completed these assignments:

- Weekly threaded discussion as preparation for synchronous discussion.
- Weekly on-line drill-and-practice exercises with feedback.
- Weekly on-line quizzes with feedback.
- Watching the video accompanying the textbook independently.
- Collaborative research and writing project involving a multiple draft word-processed essay.

Scoring and Data Analysis

Scores for all of the instruments consisted of raw scores. The scores from the two examiners on the 50-point oral proficiency scale were averaged for both pre- and posttests. For the nonword repetition task there was three sets of 8 words with a perfect performance of 24. The reading span measure awards 1 point for the combination of correctly indicating the sensibility of a sentence and recalling its final word. A perfect score on the reading span is 60, based on a total of 60 sentences. The Shipley verbal intelligence measure has a vocabulary and an abstract reasoning score that were combined for a total raw score.

Can L2 oral proficiency be indirectly developed through chatroom interaction in the target language?

To test this hypothesis, an ANCOVA was calculated with the pretest score functioning as a covariate to factor out the participants' level of oral proficiency when they entered the course. The alpha level was set at .05. The rationale for using an ANCOVA instead of a repeated measures ANOVA is derived from the mean pre- and posttest oral proficiency scores of the four groups. Looking at the pretest means in Table 2, it is apparent that the groups were not equal at the beginning of the experiment. While the control groups and the first experimental group exhibited very similar means, the second experimental group was considerably higher. With this being the case, using a repeated measures ANOVA would not take into account these pretest differences. Using pretest a covariate statistically adjusts the groups means to account for the higher pretest mean of the second experimental group.

TABLE 2: Pretest and Posttest Mean Oral Proficiency Scores

	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Gain Mean	Gain SD	N
Control 1	18.76	4.27	28.56	5.52	9.79	6.82	17
Control 2	18.74	6.32	28.59	7.92	9.85	6.83	17
Experimental 1	18.23	5.41	32.08	5.12	13.85	4.42	13
Experimental 2	**23.64**	7.25	33.32	7.15	9.68	7.80	11

The ANCOVA results (see Table 3) show that participants in the experimental condition as a group out-performed participants in the control condition ($p \leq .05$). These findings suggest that the participants spending half of their instructional time in a synchronous online environment were advantaged in their oral proficiency development over those individuals meeting face-to-face for their full instructional treatment. The language production processes outlined in Levelt's model imply that language production, whether aurally or textually, should develop the same set of underlying cognitive mechanisms. On the basis of Levelt's model alone, the logical prediction would be an equivalent gain in oral proficiency between the control and experimental conditions. A pair of t-tests were run to test this hypothesis as well. The results indicated that both the experimental and control groups demonstrated significant improvement from pretest to posttest (control 2-tail sig. = .021; experimental 2-tail sig. = .012). The fact that the mean gain score of participants conducting half of their class time in the chatroom was significantly higher than the control condition suggests that synchronous CMC may offer some unique benefits to second language learners that may be difficult to obtain in a conventional classroom setting.

TABLE 3: ANCOVA for Treatment and Posttest with Pretest as Covariate

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
TREATMENT	135.72	1	135.72	3.96	.052
PRETEST (Covar.)	441.55	1	441.55	12.88	.001
Residual	1885.26	55	34.28		
Total	2560.12	57	44.91		

Covariate Raw Regression Coefficient
PRETEST .472

Can individual differences in Working Memory capacity effectively predict the rate of L2 oral proficiency development for different types of learners in a chatroom setting?

The first step in analyzing the data addressing this question was to run the correlations between the to-be-predicted variable, gains in oral proficiency scores on the posttest, and the various psychometric predictor variables (see Table 4). The composite Working Memory score consisted of a nonword repetition test score, measuring phonological Working Memory capacity, and the reading span measure that provides a metric for executive Working Memory function. Based on the correlation of .09 between reading span and oral proficiency gain scores, the central executive appears to have no real relationship with oral proficiency development. However, this conclusion should be considered tentative based on the present results. Looking at the histograms (figure 2 and figure 3) of the frequency distributions for the nonword repetition test and the reading span measure, one sees that the scores are much more concentrated than is customary in the production-based reading span and nonword repetition tests. There are two potential explanations for this phenomenon. First, it could be that the participants in this study were of a more homogenous group than previously thought. The fact that subjects were drawn from third semester Spanish courses may have biased the sample. It could stand to reason that higher level courses contain students who have higher cognitive abilities in general, thus causing a truncated range of scores. The second possibility is that the recognition-based tests are not as taxing on memory resources as pure production tasks are. Having to maintain only enough of a memory trace to recognize words previously seen (i.e. reading span) or heard (i.e. nonword repetition task) as opposed to reproducing the word in either a written or aural form, may reduce the memory load. Reducing the burden on Working Memory may produce a facilitating effect for low spans and result in scores concentrating more towards the upper half of the scale. The most plausible explanation may in fact be a combination of a more homogenous sample than expected and the memory load reducing nature of recognition-based tests.

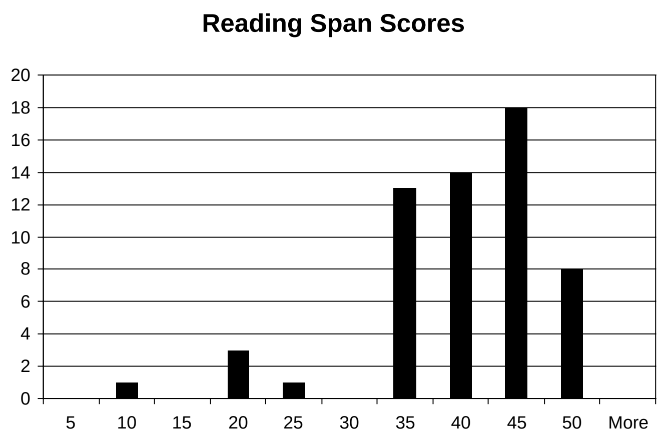

Figure 2

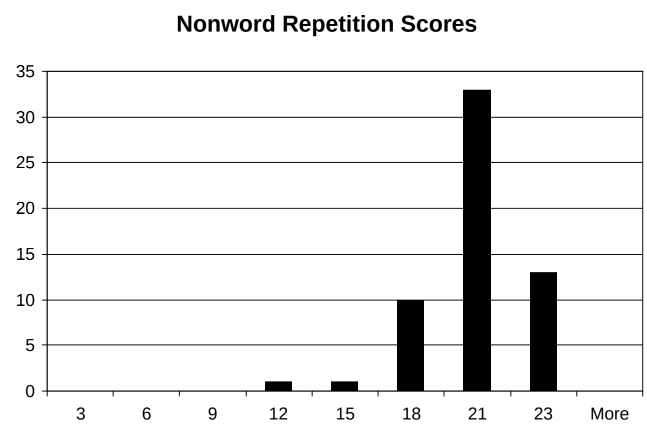

Figure 3

The relationship that stands out the most is between the nonword repetition task and the oral proficiency gain scores (r = .30). This moderate correlation suggests that phonological Working Memory capacity plays some role in oral proficiency development. The lack of a relationship between the Shipley and gains in oral proficiency suggest that it is the working memory construct measured by nonword repetition (the phonological loop) that is related to performance rather than a more global construct like general intelligence.

TABLE 4: Correlation Matrix for Predictors and Oral Proficiency Gain Scores

	OP Gain	Nonword Rep.	Rspan	Shipley
OP Gain	---	---	---	---
Nonword Rep.	.30 P= .021	---	---	---
Rspan	.09 P= .524	.33 P= .011	---	---
Shipley	.03 P= .811	.31 P= .017	.63 P= .000	---

As previously mentioned, the chatroom environment should reduce the burden on Working Memory by 1.) slowing down the pace of discussion and 2.) allowing users to refresh memory traces by rereading previous comments. The ramifications of these differences between synchronous on-line conversation and synchronous face-to-face conversation should be that learners with lower Working Memory capacity are advantaged in the chatroom setting. To test this hypothesis, correlations for the experimental and control participants were run examining the relationship between oral proficiency gain scores and nonword repetition scores. The results found that the correlation between oral proficiency gain and nonword repetition was higher for the control group (r = .33, p = .055) than for the experimental group ($r = .23$, $p = .276$). That is, the relationship between the phonological aspect of working memory and oral proficiency gains was only significant in the control group. This suggests that the learners with lower phonological buffering capacity were disadvantaged relative to others in the control group, but not so disadvantaged in the experimental group. Meeting the needs of all learners is of great importance to all second language instructors. These findings give a preliminary indication that the chatroom environment may be especially beneficial for students with lower Working Memory capacity.

DISCUSSION

The findings from this study provide evidence that L2 oral proficiency can be indirectly developed through chatroom interaction in the target language. As was suggested by Healy-Beauvois (1992) and Kern (1995), the oral proficiency gains of the experimental group indicate that a direct transfer of skills across modality from writing to speaking does occur. Based on Levelt's production model, it seemed very reasonable to expect equivalent gains on the part of control and experimental subjects. Nevertheless, the magnitude of gains on the part of the chatroom users in this study were somewhat unexpected.

It is clear that these gains are not the result of a teacher effect. In order to rule out this potential confound, an ANCOVA was run with posttest as the dependent measure, pretest as a covariate, and the independent variable of teacher with two levels: native-speaker and nonnative-speaker. The results ($p = .642$) did not provide any support for a teacher effect.

From a language instruction perspective, even equivalent levels of oral proficiency development (no significant difference) between the groups would have been a desirable outcome. Therefore, these findings suggest that processes are being activated that extend beyond the equivalence that would be predicted by Levelt's model alone. Additional qualitative data collected from this study (Payne, 1999) indicate that most of the participants in the experimental condition were conscious of their subvocalization of the language they produced in the chatroom. Of the 23 experimental participants who responded to the survey, 5 indicated that they overtly vocalized the comments they were composing and 16 said they spoke silently to themselves as they typed comments in the chatroom. When asked if they read aloud the comments others posted in the chatroom, more than 50% said at least sometimes. This qualitative data suggest that by vocalizing their own output and the input of their classmates, chatroom discourse for many participants incorporated all components of Levelt's model. This extends beyond the hypothesized equivalency that stopped short of the production of overt speech.

The question that presents itself is what are the characteristics of this form of "conversation" that appear to enhance the development in speaking skills beyond what is possible in the face-to-face setting alone? There are several qualities of chatroom discourse that might address this question. First, conversational interaction online is not subject to the turn-taking rules that apply to face-to-face discussion. In an IRC-style chatroom, where users can't see each other's comments until they have been posted, there exists a face-to-face equivalent of everyone in a discussion group talking simultaneously. In a classroom, this would be disastrous; online it works. Without having to wait for a turn, learners have a greater opportunity to produce much more language in a given on-line discussion period than is possible in most conventional classroom settings. In a 45-50 minute time period is not uncommon for students to generate 50 full-sentence comments in a lively, small-group discussion.

Language production in a chatroom is also required to be considered "present." In a classroom, students can be passive listeners and still be thought of as a participant in the discussion. In an online environment, non-participation equals non-attendance. If a student goes for more than a couple minutes without contributing to the conversation, fellow group members often inquire as to his or her whereabouts.

A third interesting difference between online and face-to-face conversation is the requirement to use language for communicating. In a classroom environment, second language learners can resort to a wide range of paralinguistic compensation strategies for getting their points across. Even if the instructional orientation of the class is towards communicative language use, once learners have understood another interlocutor's communicative intentions, the tendency is often to move ahead with the activity instead of helping their partner find the language to express his or her intentions. The necessity of using language, not pragmatics for communication in a synchronous on-line environment may push learners to experiment with the language, testing emerging hypotheses about the meaning of lexical items and the application of syntactical patterns not yet mastered (Pica et al. 1989).

The chatroom requirement of linguistic communication may also increase monitoring of one's own language and the language of others. On a five-point Likert scale, more than 50% of participants in the experimental condition reported that they focused more on grammar and the accuracy of what they say in the chatroom than in face-to-face settings (Payne, 1999). Of the participants receiving the treatment, almost two-thirds said that they noticed other people's mistakes more when conversing in the chatroom than face-to-face (Payne, 1999). Such an increased awareness may push learners to engage in more syntactic processing and to "notice" gaps in their linguistic knowledge; especially since chatroom exchanges occur in text (Swain & Lapkin, 1995).

Finally, the decreased speed of conversational exchange and the non-ephemeral nature of the medium of chatroom discourse warrant discussion. From a Working Memory perspective these two characteristics should reduce the memory load normally imposed by synchronous communication. Interlocutors can re-read comments to refresh their memory in addition to the reduced rate of exchange. The difference in the oral proficiency gain - nonword repetition correlations across the two groups in this study, suggests that this reduced memory load may benefit learners with lower phonological Working Memory capacity. Another advantage of the reduced pace of exchange in the chatroom is that students have the opportunity to engage in a limited amount of pre-task planning. The ability to plan for an oral performance task has shown to result in more fluent and syntactically complex output and increased focus on form (Ortega, 1999).

CONCLUSION

Since this is the first attempt at experimentally examining oral proficiency development as a result of synchronous CMC, these findings need to replicated with different populations and different instructional treatments. It would be beneficial to study these same variables in an online course where students rarely or never met face-to-face, but had access to pedagogically sound self-study pronunciation software. The over-arching questions that need to be addressed in light of the current push towards foreign language distance education are 1.) for what linguistic purposes is face-to-face interaction necessary for optimal second language acquisition and 2.) how can technology-mediated learning systems be used to create alternative instructional models that meet the requirements of proficiency-oriented instruction, make foreign language instruction available to a greater number of individuals, and encourage us as foreign language professionals to re-think and remain flexible in our views of what constitutes teaching and learning a second language.

As distance learning and location-independent foreign language instruction becomes more pervasive, it is important to learn how chatroom use among distributed learners differs from the computer-mediated classroom discussion (CMCD) model? In the CMCD model, students and the instructor share the same physical space (i.e. a computer lab) and interact with each other online. Understanding how the interaction between location-dependent and location-independent learners may differ is a particularly urgent question considering virtually all empirical research to date on second language chatroom use (Blake, 1999 is an exception) has been based on the CMCD model. Furthermore, almost all of these studies (Pellettieri, 1999 and Blake, 1999 are two notable exceptions) have employed the same software program, *Interchange* of the Daedelus Writing System. Since Interchange is a LAN-based technology and not a web-based or Internet Relay Chat system, using results from a location-dependent writing environment to guide pedagogical decisions about the design and implementation of location-independent instruction seems a bit precarious.

Finally, the utility of Working Memory theory for explaining the underlying mechanisms of second language acquisition clearly needs to be studied in-depth. Based on findings from this study, the connection between phonological Working Memory and second language oral proficiency warrants a closer look. The indication that learning environments can by design reduce the burden on working memory, thus producing a facilitating effect for low capacity individuals offers a new perspective on how instruction can meet the individual needs of learners.

BIBLIOGRAPHY

Adams, A.M. & Gathercole, S.E. (1996). Phonological working memory and spoken language development in young children. Quarterly Journal of Experimental Psychology, 49A (1), 216-233.

Adams, A.M. & Gathercole, S.E. (1995). Phonological working memory and speech production in preschool children. Journal of Speech and Hearing Research, Vol.38, 403-414.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369-406.

Baddeley, A.D. (1986). Working memory.

Baddeley, A.D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. Journal of Verbal Learning and Verbal Behavior, 9, 176-189.

Blake, R.J. (in press). Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning and Technology*.

Brooks, L.R. (1967). The suppression of visualisation by reading. *Quarterly Journal of Experimental Psychology*, 19, 289-299.

Chun, D.M. (1994). Using computer networking to facilitate the acquisition of interactive competence. *System*, 22 (1), 17-31.

Conoelos, T. & Olivia, M. (1993). Using computer networks to enhance foreign language/culture education, *Foreign Language Annals*, 26 (4), 527-533.

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior* , 19(4), 450-466.

Daneman, M. & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, 25, 1-18.

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research,* 20 (6), 445-464.

De Bot, K. (1992). A bilingual production model: Levelt's 'Speaking' model adapted. *Applied Linguistics*, 13 (1), 1-24.

De Bot, K. & Schreuder, R. (1993). Word production and the bilingual lexicon. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 191-214). Amsterdam: John Benjamins.

Dechert, H. W. (1980). Pauses and intonation as indicators of verbal planning in second-language speech productions: Two examples from a case study. In H. W. Dechert & M. Raupach (Eds.), Temporal variables in speech (pp. 271-285). The Hague: Mouton.

Dechert, H. W. (1983). How a story is done in a second language. In C. Faerch & G. Kaspar (Eds.), Strategies in interlanguage communication. London: Longman.

Dechert, H. W., Möhle, D., & Raupach, M. (Eds.). (1984). Second language productions. Tübingen: Narr.

Dechert, H. W., Raupach, M. (Eds.). (1980a). Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler. The Hague: Mouton.

Dechert, H. W., & Raupach, M. (Eds.). (1980b). Towards a cross-linguistic assessment of speech production. Frankfurt a.M.: Lang.

Dechert, H. W., & Raupach, M. (Eds.). (1980). Psycholinguistic models of production. Norwood, NJ: Ablex.

Engle, R., Kane, M. & Tuholski, S. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake and P. Shah (Eds.), *Models of Working Memory: mechanisms of active maintenance and executive control* (pp. 102-134). Cambridge: Cambridge University Press.

Gathercole, S.E. & Baddeley, A.D. (1993). *Working memory and language*. Hove: Lawrence Erlbaum Associates.

Gathercole, S.E. & Baddeley, A.D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. Jounal of Memory and Language, 28, 200-213.

Gathercole, S.E. & Baddeley, A.D. (1989c). The role of phonological memory in vocabulary acquisition: A study of young children learning new names.

Geva, E. & Ryan, E.B. (1993). Linguistic and cognitive correlates of academic skills in first and second languages. *Language Learning*, 43, 5-42.

Harley, B. (1993). Instructional strategies and SLA in early French immersion. *Studies in Second Language Acquisition*, 15, 245-260.

Healy-Beauvois, M. (1992). Computer assisted classroom discussion in the foreign language classroom: Conversation in slow motion. *Foreign Language Annals*, 25 (5), 455-464.

Izumi, I., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition. Studies in Second Language Acquisition, 21, 421-452.

Kelm, O.R. (1992). The use of synchronous computer networks in second language instruction: A preliminary study. *Foreign Language Annals*, 25 (5), 441-454.

Kern, R. (1995). Restructuring classroom interaction with networked computers: Effects on quality and characteristics of language production. *Modern Language Journal*, 79(4), 457-476.

Krashen, S. (1985). *The input hypothesis: issues and implications*. New York: Longman.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387-417.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: The MIT Press.

Levelt, W.J.M. (1993). Language use in normal speakers and its disorders. In G. Blanken, J. Dittmann, H. Grimm, J. Marshall, & C. Wallesch (Eds*.), Linguisitic disorders and pathologies: An international handbook* (pp. 1-15). Berlin: de Gruyter.

Levelt, W.J.M. (1995). The ability to speak: from intentions to spoken words. *European Review*, 3 (1), 13-23.

Linnell, J. (1995). Can negotiation provide a context for learning syntax in a second language? *(ERIC 393 305) Working Papers in Educational Linguistics*, 11 (2), 83-103.

Miyake, A. & Friedman, N.. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. Healy & L. Bourne*, Foreign language learning: Psycholinguistic studies on training and retention* (pp. 139-164). Mahwah, NJ: Lawrence Erlbaum Associates.

Möhle, D. (1984). A comparison of the second language speech production of different native speakers. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), Second language production (pp. 26-49). Tübingen: Narr.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21, 109-148.

Papagno, C., Valentine, T., & Baddeley, A.D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, 30, 331-347.

Payne, J.S. (1999). The psycholinguistic and pedagogical implications of integrating technology into the second language curriculum. Second Language Acquisition Institute, University of California, Davis, November 18.

Pelletierri, J. (in press). Negotiation in cyberspace: The role of chatting in the development of grammatical competence. In M. Warschauer & R. Kern (Eds.), *Network-Based Language Teaching: Concepts and Practice*. New York: Cambridge University Press.

Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*, 11, 63-90.

Poulisse, N. & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, 15, 36-57.

Poulisse, N. (1997). Language production in bilinguals. In A.M.B. de Groot & J. F. Kroll (Eds.) Tutorials in bilingualism. Mahwah, NJ: Lawrence Erlbaum.

Raupach, M. (1980). Temporal variables in first and second language speech production. In H. W. Dechert & M. Raupach (Eds.), Temporal variables in speech (pp. 271-285). The Hague: Mouton.

Raupach, M. (1984). Formulae in second language production. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.) Second language productions. Tübingen: Narr.

Rehbein, J. (1987). On fluency in second language speech. In H. W. Dechert and M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 97-105).

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14, 357-385.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology*, 45A, (1), 21-50.

Service, E. & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, 16, 155-172.

Speidel, G.E. (1989). A biological basis for individual differences in learning to speak. in G.E. Speidel and K.E. Nelson (Eds.), The many faces of imitation in language learning. New York: Springer-Verlag. pp. 199-229.

Speidel, G.E. (1993). Phonological short-term memory and individual differences in learning to speak: A bilingual case study. First Language, 13, 69-91.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass and C. Madden (Eds.), *Input and Second Language Acquisition*. Rowley, MA: Newbury House.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review*, 50, 158-164.

Swain, M. & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16(3), 371-391.

Warschauer, M. (1996). Comparing face-to-face and electronic discussion in the second language classroom. *CALICO Journal*, 13(2), 7-26.

Weise, R. (1984). Language production in foreign and native languages: Same or different. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language production* (pp. 11-25). Tübingen: Narr.

Whitney, P., & Budd, D. (1999). A separate language interpretation resource: Premature fractionation? Behavioral and Brain Sciences, 22, 113.

APPENDIX 1

Oral Production Interview Scale

Student Name: ________________________________
WSU ID#: ____________________

Comprehensibility
____ 10-9: for a native speaker: easy to understand without any confusion or difficulty.
____ 8-6: for a native speaker: can understand with minimal difficulty.
____ 5-3: for a native speaker: can understand with some difficulty.
____ 2-1: for a native speaker: can understand with great difficulty.

Fluency
____ 10-9: native-like fluency; hesitations only when appropriate.
____ 8-7: near native fluency; very few hesitations or pauses.
____ 6-5: some hesitations, pauses, but fairly continuous speech
____ 4-3: frequent hesitations and pausing, speech is more disjointed.
____ 2-1: very disjointed speech with many hesitations and pauses.

Vocabulary Usage
____ 10-9: very extensive vocabulary usage.
____ 8-7: good vocabulary usage, very few inappropriate terms.
____ 6-5: moderate vocabulary, a few inappropriate terms.
____ 4-3: limited vocabulary, some inappropriate terms used.
____ 2-1: very limited vocabulary, frequent use of inappropriate terms.

Syntax and Grammar
____ 10-9: native-like grammar and syntax; used a variety of syntax and tenses.
____ 8-7: near-native grammar and syntax; few mistakes.
____ 6-5: used few syntax structures, some grammar and syntax mistakes.
____ 4-3: very limited in syntax and grammar usage with frequent mistakes.
____ 2-1: no systematic use of grammar and syntax rules.

Pronunciation
____ 10-9: native-like pronunciation, virtually no discernable accent, no errors.
____ 8-7: near-native pronunciation, slight accent, few errors.
____ 6-5: some errors; obvious accent, but doesn't interfere with comprehension.
____ 4-3: frequent errors; strong accent; some comprehension difficulties.
____ 2-1: little effort to use Spanish pronunciation; comprehension impeded.

APPENDIX 2

Speaking Tasks

Pretest
Task 1: tell us in Spanish about a trip that you took recently.

Task 2: tell us in Spanish what you did over summer vacation.

Task 3: tell us in Spanish about your plans for Labor Day weekend.

Task 4: tell us in Spanish what you do in a normal week.

Posttest
Task 1: tell us in Spanish about a trip that you took recently.

Task 2: tell us in Spanish what you did over Thanksgiving break.

Task 3: tell us in Spanish about your plans for Christmas vacation.

Task 4: tell us in Spanish what you do in a normal week.