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

J. Scott Payne
Middlebury College

Paul J. Whitney
Washington State University

ABSTRACT

A naturalistic experiment was conducted to test the hypothesis that synchronous computer-mediated communication (CMC) can indirectly improve L2 oral proficiency by developing the same cognitive mechanisms underlying spontaneous conversational speech. The theoretical framework guiding this inquiry consists of Levelt's (1989) model of language production augmented by concepts from Working Memory theory. The findings show a significant difference between the experimental and control groups oral proficiency development with participants in the experimental condition (two of four contact hours per week were in a chatroom) scoring higher than participants in the control condition. Implications for language instruction and the second language oral proficiency development for different types of learners are discussed.

KEYWORDS

Computer-mediated Communication (CMC), Working Memory, Oral Proficiency, Computer-assisted Language Learning (CALL), Language Production, Synchronous Communication

INTRODUCTION

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 is important for developing 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 (e.g., clarification requests, modeling, and overt correction) or internal feedback (monitoring) of language they have produced. By becoming consciously aware of one's 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, Bigelow, Fujiwara, & Fearnow, 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 1992, 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)—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 forms into procedural forms allowing for automatic and 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, the authors will attempt to show that Levelt's model (1989, 1995) augmented with other concepts from cognitive psychology, 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) 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 judgment or executive functions based on changing conditions in one's immediate environment. First language research suggests that individual differences in Working Memory capacity are closely related to (a) verbal fluency (Daneman, 1991), (b) the ability of individuals to utilize contextual clues in text for learning novel words (Daneman & Green, 1986), and (c) maintaining a representation of language strings for "off-line" processing when language becomes too complex for "online" processing (Gathercole & Baddeley, 1993). Findings from second language studies indicate that verbal Working Memory capacity serves as an effective predictor of L2 vocabulary development (Gathercole & Baddeley, 1989a; 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).

Conversational exchange in a second language requires interlocutors to perform a complex set of cognitive tasks as they attempt to comprehend language input, relate it to what they know about the target language and the world, and then make decisions about whether the new information should be incorporated into their existing knowledge base in some manner. The two Working Memory concepts that appear to relate most directly to this task are executive capacity (as measured by reading span) and verbal span (as measured by nonword repetition). Executive function, or what Baddeley (1986) refers to as the Central Executive, should play a critical role in second language production and comprehension, especially in conversational exchange. Second language learners are constantly comparing what they hear and read to what they know to be true about the target language, based on their current stage of Interlanguage development. The ability to maintain a representation of target language input in memory, retrieve L2 syntactic and semantic information from long-term memory, make judgments and store the intermediate results of these calculations are all tasks of the Central Executive. Measuring an individual's capacity for executive function may provide insight into the acquisition of second language speaking skills.

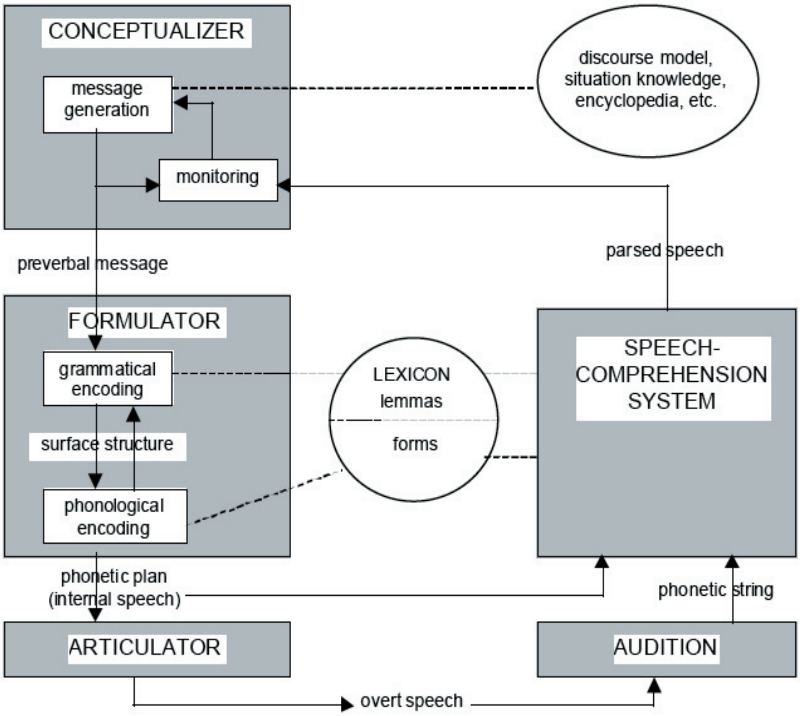
The Working Memory construct that is most intuitively associated with speaking a second language is verbal span or the ability to temporarily maintain phonological information in memory. Verb Working Memory has been the focus of most empirical studies of Working Memory and second or foreign language acquisition. Exploring the relationship between these two Working Memory constructs and second language oral proficiency development may shed light on the impact of memory limitations on such complex tasks as conversing in a foreign language.

The goal of this article 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 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), utterances begin as nonlanguage specific communicative intentions in what Levelt refers to as the Conceptualizer (see Figure 1). During production the job of the Conceptualizer is to determine the semantic content of the utterance to be spoken. 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 may be monitored internally with the support of subvocalization. During this internal feedback loop, the articulatory plan is stored in the Articulatory Buffer (Working Memory).

Figure 1
Blueprint for the Speaker



Note: Boxes represent processing components; the circle and ellipse represent knowledge stores (Levelt, 1989).

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 about 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. In fact, as speakers, we often begin uttering a sentence even before we have 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, which has 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 approach. 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). 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 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. 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, and slips of the tongue) have been employed 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, 1989b). 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 the articulation rate among children between the ages of 4 and 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 nonfluent 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, attentional resources allocated by the Central Executive (Baddeley, 1986) are required to make judgments 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 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 breakdowns in communication through negotiation of meaning, suggesting that synchronous online environments can play a role in Interlanguage development (Linnell, 1995; Pellettieri, 2000; Blake, 2000). 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 (a) students tend to produce more complex language in chatrooms than in face-to-face conversational settings (Warschauer, 1996; Kern, 1995), (b) participation increases online with "quieter" students participating as much or even more than those individuals who normally dominate classroom discussion (Warschauer, 1996; Kern, 1995; Chun, 1994), and (c) 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. To date, 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. Two characteristics of L2 chatroom interaction 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. Second, chatroom exchanges do not have the same ephemeral quality as spoken utterances. When chatting, participants can refresh memory traces by re-reading comments, which is not the case in aural conversation, face-to-face or otherwise. This characteristic 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 online 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 to produce overt speech. Furthermore, by augmenting Levelt's model with concepts and measurement techniques from Working Memory theory, two major benefits accrue. First, we can gain insight into how individual differences in processing capacity may affect oral proficiency development. Second, if we can predict which learners may benefit from what types of instructional treatments, we can use this information to provide guidance to curriculum developers. With these potential benefits in mind, the present study addresses 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?

We predicted that the oral proficiency development of participants in the experimental group would be at least equivalent to that of the control group, and possibly even greater, since the chatroom environment should reduce the burden on Working Memory, thus facilitating the development of low Working Memory 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 two face-to-face and two online 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 an online collaborative research and writing tool used as part of the course. 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 groups, thus the same activities or discussions were held online in the chatroom and 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; during the last week of the study, the posttest was administered. 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 it was necessary to avoid forcing them to run back and forth across campus.) However, each instructor taught one experimental and one control group, so the treatment was not confounded by the instructor variable. Participants in all conditions received extra credit totaling a maximum of one third of a letter grade for participating in the study.

Materials

Currently, the most widely 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: (a) the OPI is not sensitive enough to measure changes in oral proficiency that may occur in a single semester in a course meeting only four hours per week and (b) a significant proportion of the OPI score consists of competencies that are not addressed by this study's research questions (i.e., sociolinguistic competence). It is important to note that the term oral proficiency in this article is a more simplified construct than that used by ACTFL. Oral proficiency in this context refers to an individual's ability to produce language that is comprehensible with syntax and vocabulary appropriate to 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 the Oral Production Interview Scale in Appendix A). For the speaking test, participants selected one of four envelopes containing a speaking task written in English (see speaking tasks in Appendix B). The description of the speaking tasks was written in English 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 participants 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 to, but not interview, the speaker. Two examiners (one native speaker and one nonnative speaker, both female) administered the speaking tests. The examiners were told to think of someone they know who is a very fluent nonnative speaker of Spanish and consider that individual's language ability as a perfect score (the 50 point maximum). This expectation differs from that used in the ACTFL scale, which uses the language skills of an educated native speaker as the highest rating. Clearly, the very high criterion used on the ACTFL scale was inappropriate for a study examining changes in proficiency over the course of a single semester. The examiners received instruction on how to use the scale and compared evaluations for the first two speaking tests on each test day to maintain interrater reliability. For the oral proficiency pretest, interrater reliability on the 50-point scale was

.86; on the posttest, interrater reliability was .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 and is the most widely used test for measuring verbal Working Memory capacity. Several variations of the nonword repetition task have been reported in the literature. In the nonword repetition task developed for this study, participants listened to an audio file of 8 pseudowords read with a one-second interval between words. After listening to the audio file, participants clicked on a button to see a screen containing 16 pseudowords, 8 of which were articulated in the audio file. Students selected the 8 words they believed they heard by clicking on the checkbox next to each word. The participants could take as much time as they needed to make their eight selections. After clicking the submit button, the next audio clip would load, ready to be played. The complete nonword repetition task consisted of 3 sets of 8 pseudowords.

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 & Budd, 1999). Reading span assesses two key functions of executive Working Memory: the ability (a) to make judgments and (b) 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, participants are required to make a response indicating whether the sentence makes sense or not and to remember the last word of the sentence. After participants have seen all sentences in the set, they must remember all the sentence-final words. With only 2 sentences in a set, combining the judgment and memory 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 the current sentence while maintaining the last words from the previous sentences 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, participants clicked on a button to see a screen of words with checkboxes next to the words. For each word to be remembered, there were two distractors (i.e., for sets containing 5 sentences there was a total of 15 words). Distractors were of two types: (a) the same semantic category (e.g., if the target word was "girl," the distractor could be "woman") or (b) the last words from sentences in previous sets. Subjects 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 calculation of results.

Data were collected on student grades at the conclusion of the third semester Spanish course, overall GPAs, and verbal intelligence scores as measured by the Shipley test. These academic and verbal IQ data were used to account for extraneous factors that could confound the interpretation of results.

Treatment

One of the challenges of conducting research in a natural setting with intact groups is the issue of unequal treatment or of a "teacher effect." To ensure that the treatment administered to participants in the experimental and control conditions were equivalent, the curriculum and lesson plans for all four groups were the same. Thus, the students in the experimental groups meeting online in the chatroom engaged in the same activities on the same days as those in 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 phase of the project, it was determined that chatroom discussion groups of four to six students were the best. 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 to 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 foreign language computer lab 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 foreign language computer lab but, rather, accessed the chatroom from their home computers or machines in other computer labs on campus. Those participants who continued coming to the foreign language computer lab 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 foreign language computer lab was never more than four and during most sessions only two 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 online course management features into all participating groups. Both experimental and control groups used these tools and completed the following assignments:

1. weekly threaded discussion as preparation for synchronous discussion,
2. weekly online drill-and-practice exercises with feedback,
3. weekly online quizzes with feedback,
4. independent viewing of the video accompanying the textbook, and
5. a 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 were 3 sets of 8 words with a perfect performance of 24. The reading span measure awarded one point for the combination of the correct indication of the sensibility of a sentence and recall of its final word. A perfect score on the reading span was 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 the hypothesis above, an ANCOVA was calculated with the pretest score functioning as a covariate to factor out the participants' level of oral proficiency at the beginning of the course. The rationale for using an ANCOVA instead of 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 1, it is apparent that the groups were not equal at the beginning of the experiment.

Table 1
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

* $p < .05$

While the control groups and the first experimental group exhibited very similar means, that of the second experimental group was considerably higher. Because of this difference, a repeated measures ANOVA would not take this pre-existing difference into account. Using the results of the pretest as a covariate permitted a more accurate analysis of the posttest scores.

The ANCOVA results (see Table 2) showed that participants in the experimental condition as an aggregate group outperformed participants in the control condition ($p < .05$).

Table 2
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	.05
PRETEST (Covariate)	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

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 meeting face to face in the classroom. The language production processes outlined in Levelt's model imply that language production, whether aural or textual, 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. Two 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 group, $p < .05$; experimental group, $p < .05$). The fact that the mean gain score of participants conducting half of their class time in the chatroom was 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.

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 gains in oral proficiency scores on the posttest and the various psychometric predictor variables (see Table 3).

Table 3
Correlation Matrix for Predictors and Oral Proficiency Gain Scores

	Oral Proficiency Gain	Nonword Repetition	Reading Span	Shipley Verbal Intelligence
Oral Proficiency Gain	—	—	—	—
Nonword Repetition	.30*	—	—	—
Reading Span	.09***	.33*	—	—
Shipley Verbal Intelligence	.03***	.31*	.63**	—

*p < .05
**p < .001
***not significant

It should be recalled that the composite Working Memory score consisted of a nonword repetition test score, measuring phonological capacity, and the reading span measure, providing a metric for 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. A review of the histograms of the frequency distributions for the nonword repetition test and the reading span measure (see Figures 2 and 3) shows that the scores are much more concentrated than is customary in the production-based reading span and nonword repetition tests.

Figure 2
Nonword Repetition Scores

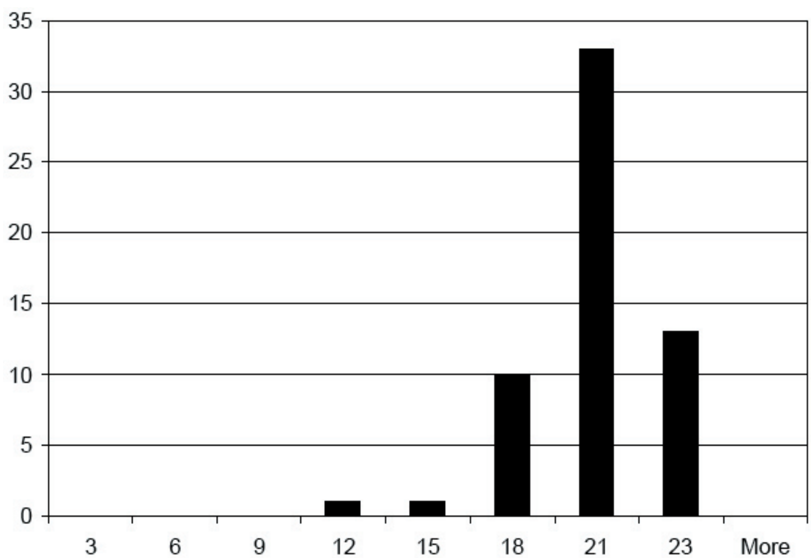
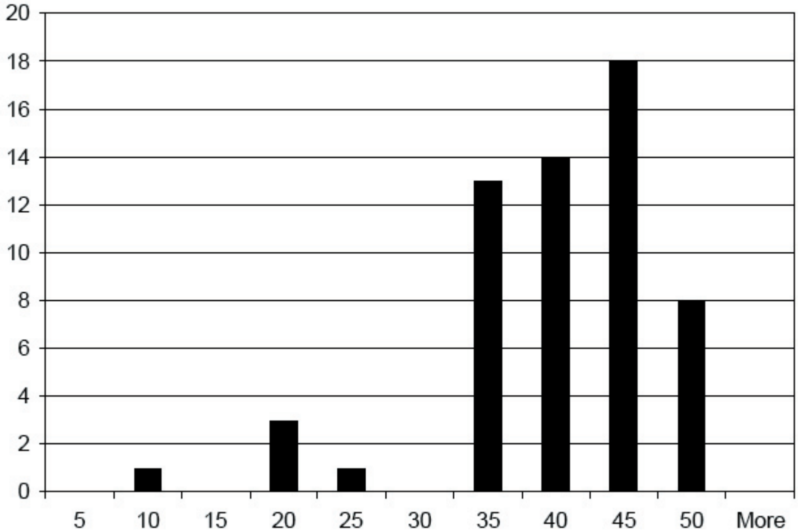


Figure 3
Reading Span Scores



There are two potential explanations for this phenomenon. First, it is possible that the participants in this study were a more homogenous group than previously thought; the fact that subjects were drawn from third semester Spanish courses may have biased the sample. It stands to reason that intermediate level foreign language courses could 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 concentrated 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.

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

As previously mentioned, the chatroom environment should reduce the burden on Working Memory by (a) slowing down the pace of discussion and (b) allowing users to refresh memory traces by re-reading previous comments. The ramifications of these differences between synchronous online 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, the relationship between oral proficiency gain scores and nonword repetition scores was examined. The correlations for the experimental and control participants between oral proficiency gain scores and nonword repetition scores in Table 3 above show that the correlation was higher for the control group ($r = .33$) than for the experimental group ($r = .23$). This finding suggests that the learners with lower phonological buffering capacity were disadvantaged relative to others in the control group but were not so disadvantaged in the experimental group. These results give a preliminary indication that the chatroom environment may be especially beneficial for students with lower ability to maintain verbal information in the Phonological Loop.

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 gain for chatroom users in this study was somewhat unexpected.

These gains are not the result of a teacher effect. An ANCOVA analysis of the posttest as the dependent measure, pretest as a covariate, and the independent variable of teacher with two levels (native speaker and nonnative speaker) showed no effect of the teacher variable ($p = .64$).

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 causal mechanisms 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 a survey questionnaire, 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 yes, at least sometimes. These 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? Several qualities of chatroom discourse can address this question. First, conversational interaction online is not subject to the turn-taking rules that apply to face-to-face discussions. In an IRC-style chatroom, where users cannot 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. This situation would be disastrous in a classroom, but it works online. Without having to wait for a turn, learners have a greater opportunity to produce much more language in online discussions than is possible in most conventional classroom settings. In a 45-50 minute online session, it 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 for a student to be considered "present." In a classroom, students can be passive listeners and still be thought of as attending the discussion. In an online environment, nonparticipation equates to nonattendance. If students go for more than a couple minutes without contributing to the conversation, fellow group members often inquire about their 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 to get their points across. Even in a classroom where students are encouraged to use the target language for communication, 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 online 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 language use may also increase students' monitoring of their 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 the grammatical correctness and the accuracy of what they said 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. Such an increased awareness may push learners to engage in more syntactic processing and "notice" gaps in their linguistic knowledge, especially since chatroom exchanges occur in written form (Swain & Lapkin, 1995).

Finally, the decreased speed of conversational exchange and the nonephemeral 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 correlation between oral proficiency gain and nonword repetition across the two groups 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 pretask 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 we are very early in experimentally examining oral proficiency development as a result of synchronous CMC, these findings need to be 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 (Donahue, 2000). It is important to emphasize that these results do not suggest that speaking skills can be developed in the absence of face-to-face conversational interaction. Clearly, the participants in the experimental groups had opportunities for face-to-face conversational interaction. It may be most useful to view the chatroom as analogous to the flight simulators used by pilots in training; the chatroom sessions may well serve as a *conversation simulator* for foreign language learners. The notion that learners can practice "speaking" in an environment where affect and rate of speech are minimized is very appealing. Possibly more important is the realization that if we as second language instructors assume that face-to-face speech is the only way to develop conversational ability, we may in fact be disadvantaging a significant portion of our students. For students who find L2 oral production an overwhelming task and tend to tune out when the linguistic data generated in face-to-face conversational settings becomes too great, the online synchronous interaction appears to give them a leg up on developing L2 oral proficiency.

The overarching questions that need to be addressed in light of the current push towards foreign language distance education are

1. When is face-to-face interaction critical for optimal second language acquisition?
2. How can technology-mediated learning systems be used to create alternative instructional models that meet the requirements of communicative language instruction, make foreign language instruction available to a greater number of individuals, and encourage us as foreign language professionals to constantly re-evaluate our own 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 has been based on the CMCD model. (Blake [2000] is an exception.) Furthermore, almost all of these studies have employed the same software program, *Interchange* of the Daedelus Writing System. (Pellettieri [2000] and Blake [2000] are two notable exceptions.) 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 in the study presented here, 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 and thereby produce a facilitating effect for low capacity individuals offers a new perspective on how instruction can meet the individual needs of learners.

REFERENCES

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.

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

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

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. (2000). Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning and Technology* [Online], *4* (1), 120-136. Available: llt.msu.edu

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.

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

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.

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.). (1987). *Psycholinguistic models of production*. Norwood, NJ: Ablex.

Donahue, S. (2000). Teaching pronunciation online. *ALN Magazine* [Online], *4* (1). Available: www.aln.org/alnweb/magazine

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 & 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. (1989a). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200-213.

Gathercole, S. E., & Baddeley, A. D. (1989b). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81* (4), 439-454.

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

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.), *Linguistic 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? *Working Papers in Educational Linguistics, 11* (2), 83-103. (ERIC Documentation Reproduction Service No. ED 393 305)

Miyake, A., & Friedman, N. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. Healy & L. Bourne (Eds.)*, 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, November). *The psycholinguistic and pedagogical implications of integrating technology into the second language curriculum*. Paper presented at the Second Language Acquisition Institute, University of California, Davis, CA.

Pellettieri, J. (2000). 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. (1997). Language production in bilinguals. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism*. Mahwah, NJ: Lawrence Erlbaum.

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

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). Norwood, NJ: Ablex.

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* (pp. 199-229). New York: Springer-Verlag.

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 aquisition*. 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 A

Oral Production Interview Scale
Student Name: ______________________________
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 B

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.

AUTHORS' BIODATA

J. Scott Payne received his Ph.D. from the Individual Interdisciplinary Doctoral Program at Washington State University in 2000. His research interests include psycholinguistics and the impact of individual differences in working memory capacity on learner behavior in computer-mediated learning environments. He is currently collaborating with faculty, conducting research and developing software under the auspices of Project 2001, an initiative funded by the Andrew W. Mellon Foundation to encourage the integration of technology into foreign language instruction in 62 liberal arts colleges.

Paul Whitney is Professor and Chair of Psychology at Washington State University. He received his Ph.D. in cognitive psychology from the University of Kansas in 1984. His research interests are in psycholinguistics, particularly in individual differences in language and memory processes. Professor Whitney has published over 30 articles dealing with the role of knowledge and ability factors in determining language processing strategies.

AUTHORS' ADDRESSES

J. Scott Payne
Instructional Technologist
Center for Educational Technology
Middlebury College
Middlebury, VT 05753
Phone: 802/443-5944
Fax: 802/443-2053
Email: jspayne@middlebury.edu

Paul J. Whitney
Department of Psychology
Washington State University
Pullman, WA 99164-4820
Phone: 509/335-8386
Fax: 509/335-5043
Email: pwhitney@mail.wsu.edu