WORKING MEMORY AND L2 ORAL FLUENCY

by

Gregory John Mizera

B.A., University of Chicago, 1983

M.S., Central Connecticut State University, 1997

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

ARTS AND SCIENCES

This dissertation was presented

by

Gregory John Mizera

It was defended on

Nov. 21, 2005

and approved by

Dr. Suzanne Curtin, Assistant Professor, University of Calgary

Dr. Robert DeKeyser, Professor, University of Maryland

Dr. Nuria Sagarra, Assistant Professor, Pennsylvania State University

Dissertation Director: Dr. Alan Juffs, Associate Professor, University of Pittsburgh

WORKING MEMORY AND L2 ORAL FLUENCY

Gregory John Mizera, M.S.

University of Pittsburgh, 2006

ABSTRACT: This main experiment in this project was designed to test the hypothesis that individual differences in working memory (WM) capacity correlate significantly with individual differences in the ability to speak a second language (L2) fluently. A pilot project was carried out to provide a set of quantifiable factors that produced a reliable description of L2 oral fluency. These factors were related to speed of delivery, pause profiles and morphosyntactic accuracy. In the main experiment, 44 native English speakers who were studying Spanish as a foreign language were tested with a set of three working memory tests, and the scores from these tests were correlated with the scores of three L2 oral fluency tests. The hypothesized strong correlations between working memory capacity and fluency were not found. Furthermore, many of the working memory scores did not correlate strongly with each other. These negative results are explained here partly by reference to the complex nature of speaking in a foreign language, which may tax other faculties more than working memory. Personal and affective variables are also mentioned as a possible explanation, as well as the relationship between working memory and long-term memory stores.

TABLE OF CONTENTS

PREFACE XI

1.0 INTRODUCTION 1

1.1 DEFINITIONS AND ASSUMPTIONS 1

1.2 CONNECTIONS AND HYPOTHESES 4

1.3 PROCEDURES 6

2.0 LITERATURE REVIEW 8

2.1 WORKING MEMORY 8

2.1.1 From short-term memory to working memory 8

2.1.2 Models of working memory 9

2.1.2.1 Baddeley's model.. 9

2.1.2.2 Alternative models of working memory. 12

2.1.3 Working Memory and Language 14

2.1.3.1 Daneman and Carpenter (1980). 14

2.1.3.2 Further research on Working Memory and Reading. 18

2.1.3.3 Phonological short-term memory and language learning. 20

2.1.3.4 Working memory and oral fluency. 22

2.1.3.5 Working Memory in Levelt's (1989) Model of Speech Production 31

2.1.3.6 Summary of Working Memory Research 34

2.2 ORAL FLUENCY 36

2.2.1 Perspectives on Fluency 36

2.2.2 Pedagogical approaches to fluency 37

2.2.3 Qualitative approaches to fluency 38

2.2.3.1 Fillmore's four dimensions of fluency 38

2.2.3.2 Communicative teaching and L2 fluency. 42

2.2.3.3 Communicative strategies and fluency.. 45
2.2.4 Quantitative studies of L2 fluency 48
2.2.5 Psycholinguistic perspectives: Fluency as automatization 55
2.2.6 Summary of working memory research 59
3.0 CHAPTER 3: EXPERIMENT 1 61
3.1 INTRODUCTION 61
3.2 METHODOLOGY 61
3.2.1 Participants 61
3.2.2 Task and Materials 62
3.2.3 Analysis 63
3.3 RESULTS 66
3.3.1 Holistic Fluency Ratings 66
3.3.2 Speed factors 67
3.3.3 Smoothness-related factors 68
3.3.4 Syntactic complexity 70
3.3.5 Morphosyntactic accuracy 70
3.3.6 Multiple regression analysis 72
3.3.7 Post-Rating Questions 73
3.3.7.1 Questions 1 and 2.. 73
3.3.7.2 Questions 3 and 4.. 74
3.3.7.3 Question 5.. 75
3.4 DISCUSSION 77
4.0 EXPERIMENT 2 80
4.1 INTRODUCTION 80
4.2 METHODOLOGY 82
4.2.1 Working memory measures 82
4.2.1.1 Speaking Span Test 82
4.2.1.2 Math Span Test.. 83
4.2.1.3 Non-Word Repetition Test. 84
4.2.2 L2 fluency tests 85
4.2.2.1 Narrative Monologue Test. 85

4.2.2.2 Word Translation Test. 87

4.2.2.3 Imitation/Grammaticality Test 87

4.2.3 Spanish Proficiency Test 89

4.3 HYPOTHESES 90

4.4 RESULTS AND ANALYSIS 93

4.4.1 Working memory scores 93

4.4.2 Fluency-related factors 97

4.4.3 Correlations between working memory and fluency variables 99

4.4.4 L2 proficiency and the working memory-fluency relationship 102

4.5 DISCUSSION 104

4.5.1 The nature of oral fluency 105

4.5.2 The influence of working memory over time 107

4.5.3 Personal and Affective Factors 109

4.5.4 The Nature of Working Memory 111

5.0 SUMMARY AND CONCLUSIONS 114

5.1 THE STUDY OF L2 ORAL FLUENCY 114

5.2 WORKING MEMORY AND LANGUAGE 115

5.3 LIMITATIONS OF THESE STUDIES 118

5.4 IDEAS FOR FURTHER RESEARCH 119

NOTES 121

APPENDIX A 128

APPENDIX B 130

APPENDIX C 132

APPENDIX D 134

APPENDIX E 136

APPENDIX F 138

APPENDIX G 139

APPENDIX H 140

APPENDIX I 146

BIBLIOGRAPHY 154

LIST OF TABLES

Table 1: Sample sentences from Waters and Caplan (1996). 12

Table 2: Sample sentences from Daneman and Carpenter's (1980) Reading Span Test 14

Table 3: Correlations between WM spam and reading skill (Daneman & Carpenter, 1980)....... 16

Table 4: Correlations between Listening Span Test (LST) scores and reading skill measures (Daneman & Carpenter, 1980)... 16

Table 5: Lexically ambiguous sentences (Miyake, Just & Carpenter, 1994) 19

Table 6: Correlations between repetition accuracy and ESL skills (Service, 1992)..................... 21

Table 7: Sample sentences from Contextual Vocabulary Production Task (Daneman & Green, 1986) 24

Table 8: Correlations between WM span and lexical fluency (Daneman & Carpenter, 1986) 24

Table 9: Correlations between WM scores and Speech Generation Task (SGT) scores (Daneman, 1991) 26

Table 10: Example stimuli from Oral Slip Task (Daneman, 1991)... 27

Table 11: Correlations between WM scores and oral fluency scores (Daneman, 1991).............. 28

Table 12: Correlations between WM and Speaking Generation Task (SGT) scores (Fortkamp, 1999) 28

Table 13: Mean scores from Fortkamp’s (1999) WM tests.. 29

Table 14: Correlations between WM scores (Fortkamp, 1999)... 30

Table 15: The University of Reading fluency rating scale 38

Table 16: Some formulaic expressions (Fillmore, 2000).. 40

Table 17: Excerpts from monologues by Brazilian ESL learners (Ejzenberg, 2000) (bold print is Ejzenberg's)... 47

Table 18: Temporal qualities of ESL speech studied by Lennon (1990) 49

Table 19: Possible fluency-related factors identified by Freed (2000) 52
Table 20: Summary of four quantitative studies of L2 oral fluency 53
Table 21: Fluency factors measured in Experiment 1 64
Table 22: Types of morphosyntactic errors counted in Experiment 1 66
Table 23: Speed-related factors (all figures are per minute except correlations) 67
Table 24: Factors related to voiced disfluencies (all figures are per minute except correlations) 68
Table 25: Factors related to unvoiced disfluencies 69
Table 26: Mean length of T-units (in words) 70
Table 27: Morphosyntactic error rate 71
Table 28: A sample of speakers' errors 71
Table 29: Maximum R-square multiple regressions analysis 72
Table 30: A sampling of responses to question 3 74
Table 31: Responses to question 5 76
Table 32: Correlations between speed-related factors and morphosyntactic accuracy 77
Table 33: Some strong correlation with holistic fluency ratings in Experiment 1 78
Table 34: Test used in Experiment 2 90
Table 35: Types of scores produced in Experiment 2 91
Table 36: Working memory scores 94
Table 37: Correlations between WM scores 95
Table 38: Scores for fluency-related variables 97
Table 39: Correlations between fluency variables 98
Table 40: Correlations between WM scores and fluency scores 100
Table 41: Correlations between Spanish Proficiency Test (SPT) scores and WM scores 101
Table 42: Correlations between Spanish Proficiency Test (SPT) scores and L2 fluency scores. 101
Table 43: Fluency scores for 20 least proficient participants 102
Table 44: Working memory scores for 20 least proficient participants 103
Table 45: Correlations between WM scores and L2 fluency scores for 20 least proficient participants 104

LIST OF FIGURES

Figure 1: Baddeley's model of working memory... 2
Figure 2: Revised model of working memory (Baddeley, 2000) 11
Figure 3: Levelt's model of speech production.. 32
Figure 4: Anderson’s (1983) ACT* model of skill acquisition 56

PREFACE

I would like to acknowledge, first and foremost, my wife, Marianne. Without her unswerving support this dissertation would never have been conceived and initiated, much less completed. Secondly, my thanks go to Gabriella, who sacrificed innumerable hours of playtime to this project with as much good grace as a 4-year-old can muster. Thank you to my adviser, Prof. Alan Juffs, who suffered through many bad drafts and shepherded this dissertation through to completion while carrying out a multitude of more important duties. I would also like to acknowledge the other members of my committee: Prof. Robert DeKeyser, Prof. Suzanne Curtin (both of whom kindly agreed to remain on my committee after leaving the University of Pittsburgh), and Prof. Nuria Sagarra. Finally, I give my heartfelt thanks to the following people, who contributed to this project in one way or another: Dr. Beatrice DeAngelis, Maeve Eberhardt, Veronica Lifrieri, Patricia Pharmakis, Guillermo Rodriguez and Armando Zavaleta. Thank you all.

1.0 INTRODUCTION

1.1 DEFINITIONS AND ASSUMPTIONS

The purpose of this project was three-fold. Firstly, and most generally, it sought to extend a quarter-century-long line of research on working memory and language into the largely unexplored area of second language (L2) speech production. The central hypothesis was that working memory plays a significant and measurable role in the production of fluent, spontaneous L2 speech. Secondly, it set out to explore how individual differences in learners' working memory capacity may affect specific skills underlying L2 fluency, including lexical retrieval efficiency and self-monitoring skill. Finally, the first experiment of the two presented here was designed to provide a manageable set of quantifiable factors that could be used to describe and assess L2 oral fluency.

"Working memory" is defined here as the ability to simultaneously maintain and manipulate information in the performance of complex cognitive tasks. This definition is derived primarily from Baddeley's (Baddeley, 2003; 1999; 1986; Gathercole & Baddeley, 1993; Baddeley & Hitch, 1974) model of working memory. Baddeley envisions a central executive component and two "slave" systems, the phonological loop and the visuo-spatial sketchpad (Fig. 1-1). As their names suggest, these slave systems are dedicated to storing auditory information and visual information respectively. The central executive is responsible for allocating attentional resources during learning and/or performance, and for suppressing irrelevant information. The most recent version of Baddeley's model (2000) includes another component, the "episodic buffer," which integrates knowledge from long-term memory and the slave systems. Though based on Baddeley's model, the definition of working memory given here is broad enough to be consistent with other views (see the "consensus" definition proposed by Miyake & Shah (1999b, p. 430)). [1]

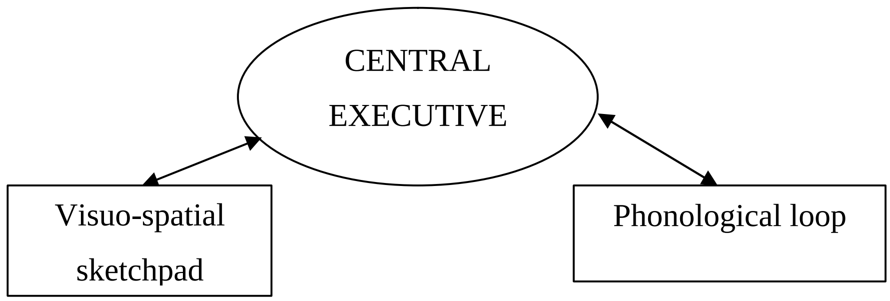

Figure 1: Baddeley's model of working memory

However, there are lively theoretical debates over the precise architecture and operation of working memory, and these differences of opinion should not be ignored. As outlined in the next section, one key question is working memory's independence from long-term memory. Ericsson and his collaborators (Ericsson & Delaney, 1999; Ericsson & Kintsch, 1995; Ericsson & Charness, 1994) propose that the efficiency of working memory, or "long-term working memory" in their parlance, depends on knowledge and skills already stored in long-term memory. If they are correct, then the importance of working memory in learning and accessing material will necessarily be reduced. Another controversial topic centers on whether verbal working memory is best conceptualized in terms of a single, capacity-based resource (Baddeley, 1999, 1986; Just & Carpenter, 1992), or in terms of multiple processing mechanisms (Roberts & Gibson, 2002; Caplan & Waters, 1996). The way in which these questions are answered has implications for how one visualizes working memory's role in language learning and production.

Several basic assumptions about working memory are relevant here. Firstly, working memory is assumed to be a limited-capacity system in that people are constrained in how many items of information they can store and manipulate at one time. This number is often put at between two and six in various memory span tests (Daneman & Carpenter, 1980; Daneman & Green, 1986; Salthouse & Babcock, 1991). Secondly, people differ in the number of items they can handle in working memory, as measured by these span tests. In Daneman & Carpenter (1980) and Daneman & Green (1986), for example, mean scores on one such test (the Reading Span Test) were around 3.0 items, with individual variations ranging from 2.0 to 4.5. Thirdly, these individual differences manifest themselves more clearly as the complexity of the cognitive task being performed increases, along with the demands placed on working memory (Miyake, Just & Carpenter, 1994). By extension, it is argued here that individual differences will be more apparent as the individual's skill relative to the task decreases. In language learning terms, this means that novice learners will be more dependent on working memory in L2 speech production than more advanced learners (Temple, 1997).

The other factor in this study, the notion of oral fluency, is also the topic of considerable debate. For the purposes of this study, fluency is defined as the ability to spontaneously speak a language quickly and comprehensibly, without an undue number of formal errors that distract listeners from the speaker’s message. This definition leaves considerable room for subjective interpretation: how rapid must speech be to be considered fluent, and how many errors are permissible before speech becomes disfluent or distracting? There are no simple, definitive answers to these questions, but the first experiment in this study was meant to provide at least a partial answer through a combination of quantitative and qualitative approaches.

It may be objected that formal accuracy should not be included under the rubric of fluency, that by combining both speed and accuracy in the definition, “fluency” becomes nothing less than a synonym for overall speaking proficiency. This objection has some validity, and the choice of the term “fluency” rather than some other term such as “oral proficiency” needs to be explained. One justification is that while formal accuracy is generally assumed in mature L1 speech, it cannot be ignored in non-native speech. An L2 learner whose speech contains so many lexical and/or morphosyntactic errors that it is incomprehensible cannot reasonably be considered “fluent,” no matter how quickly he or she speaks. In fact, there is some evidence, including evidence from the first experiment reported below, that native speaker/listeners take non-native limitations into account and judge L2 fluency differently than L1 fluency. Another reason for using "fluency" is that alternate terms such as “oral proficiency” are generally too comprehensive, as this term may be interpreted as including pragmatic factors such as knowledge of appropriate language use in various social contexts. This sort of socio-linguistic skill is certainly important but it is irrelevant to the experiments in this study, which focus on fluency as a psycholinguistic phenomenon.

At first glance, oral fluency seems to be an expression of linguistic performance, that is, of the ability to make use of underlying linguistic competence quickly and skillfully. However, it is possible to think of fluency as straddling a fuzzy borderline between competence and performance. As Leeson (2000, p. 27) notes, a speaker may be able to retrieve a desired word given enough time, but "enough time" can mean either mere seconds or several days. The same is true in the case of longer phrases and syntactic patterns: the difference between lack of knowledge and long lapses in retrieval of knowledge is, for the most part, unimportant in spontaneous speech. Conversely, an L2 speaker may possess the ability to recognize his or her own *lacunae* during ongoing discourse and to quickly employ circumlocutions or compensatory strategies. Ejzenberg (2000) describes how this linguistic "juggling act," as she calls it, disguises problems in L2 competence while maintaining an impression of fluency in the listener's mind. It is theorized that working memory efficiency may help learners successfully perform this “juggling act.”

1.2 CONNECTIONS AND HYPOTHESES

There are, in fact, good reasons to believe that these two factors, working memory capacity and L2 oral fluency, are related. In broad terms, speaking is a matter of choosing words and assembling them in sequences that are acceptable under the rules of the language and that express the speaker’s communicative intention. The speaker accesses lexical items, either words or larger “chunks” of language, from long-term memory stores, perhaps modifies them to fit the demands of his or her message, and, simultaneously or in very quick succession, puts them in appropriate sequences. This recall-modify-assemble process is outlined in Levelt’s (1989) model of speech production, the model used in much of the discussion in later chapters. But some variant of it will almost certainly be present, explicitly or otherwise, in any such model. The point is that this process, which is repeated constantly during creative, unrehearsed speech, seems to call on exactly the skills that are said to be involved in working memory, namely the ability to simultaneously maintain and manipulate bits of information. This is true to some extent of other linguistic skills besides speech, including writing, the other productive skill. But writers rarely face the temporal demands that spontaneous speech presents, including the need to maintain the flow of a conversation or narrative with an interlocutor.

Working memory capacity may be even more important in non-native speech production than in L1 speech, particularly when a speaker is at the low or intermediate level of L2 proficiency. Lexical access will be generally slower and more deliberate than in L1 speech, and there will be fewer multi-word "chunks" of prefabricated language that the speaker can employ (Lennon, 2000; Oppenheim, 2000; Segalowitz, 2000). This will leave fewer attentional resources for the morphosyntactic ordering of the accessed items, and either cause errors or else force the speaker to adopt a slow, serial mode of production, as opposed to the parallel processes that underlie fluent native-like speech (Temple, 1997). Therefore, one hypothesis that was explored in the main experiment here is that the correlation between working memory capacity and L2 oral fluency will be especially strong at lower levels of L2 expertise.

Two more specific connections between working memory capacity and fluency are proposed. Firstly, it is argued that the attentional resources of the central executive, its ability to suppress or exclude information that is irrelevant to the task at hand, fulfill an important function in lexical access in L2 speech production. This function is essentially to ward off interference from both the L1 lexicon and from similar-sounding L2 words that have been previously learned- “clang associates,” in Ellis’ (1996) term. According to the argument advanced herein, the novice learner must expend precious attentional resources suppressing these irrelevant associations, and learners who can do this better will produce more fluent speech.

A second argument outlined below is that working memory helps ensure formal accurate speech. This idea is derived from Levelt’s (1989) model of speech production, where working memory is presumed to play a role in the monitoring of one’s own speech, both internally (i.e., before articulation) and afterward. In the second experiment in this project, a new measure, the Imitation/Grammaticality Test, is introduced as a substitute for this sort of self-monitoring.

As it turned out, these arguments were not borne out by the results of this study. The significant and strong correlations between working memory capacity and L2 oral fluency were not found. Nor, for the most part, were strong correlations found between working memory capacity and lexical retrieval skill or monitoring ability. Moreover, there were relatively weak correlations between different measures of working memory, lending support to a domain-specific view of working memory. These results are explained by reference to the complexity of L2 speech, the persistence of working memory’s influence on language learning, personal and affective variables and the nature of working memory and its relationship to long-term memory.

1.3 PROCEDURES

Two experiments are described below. The first is a pilot study aimed at determining how native (English) speaker/listeners assess non-native oral fluency. The results suggest that three factors correlate most closely with fluency ratings: speed of speech production, length of inter-clausal pauses and morphosyntactic error rate. Follow-up interviews with the same native speaker/listeners indicate that they do in fact judge L2 fluency differently than L1 fluency, and that formal accuracy is inextricably bound up with their judgments about non-native speech.

The second experiment sought correlations between L2 fluency and working memory capacity. In this experiment, 38 native speakers of English who were studying Spanish as a foreign language were tested for both working memory capacity and L2 (Spanish) oral fluency. The main hypothesis was that there would a strong, positive correlation between the two factors. A secondary hypothesis was that this correlation would be stronger for low-level learners than for more advanced learners. Participants of various L2 proficiency levels were included in this experiment in order to test the second hypothesis.

Working memory capacity was measured in three ways in this experiment: a Speaking Span Test, a Math Span Test and a Non-Word Repetition Test. The aim of this three-pronged approach was to obtain separate measures for working memory as expressed in a verbal mode (Speaking Span), a non-verbal mode (Math Span) and a verbal but purely phonological mode (Non-Word Repetition). This allowed the researcher to compare these three measures with each other, in addition to correlating them with various fluency measures.

Fluency was assessed firstly by means of a Narrative Monologue Task, wherein participants had to produce a three-minute narrative of a children's picture book. This speech sample was recorded and analyzed in terms of speech rate, inter-clausal pauses and morphosyntactic error rate, the three factors found to be most significant in the first experiment. The participants were also given a Word Translation Test to test the speed of their L2 lexical access, a skill that is assumed to underlie fluency. In this task, they had to translate 40 common

English nouns (body parts, articles of clothing, etc.) into Spanish as quickly as they could; response times were recorded in milliseconds. In addition, the participants' ability to repeat and monitor L2 speech for formal errors was tested in the Imitation/Grammaticality Test mentioned above. This test called on the participants to recall short Spanish phrases they had just heard and, if necessary, to correct errors on the spot.

These two experiments approach fluency from distinct angles. The first experiment emphasizes the perspective of the native listener and how he or she judges non-native fluency. This experiment was mainly quantitative, but it was also qualitative in that the correlational data it produced was supplemented by extensive post-task interviews with the native listeners. The second experiment attempted to draw correlational connections between L2 fluency and individual differences in working memory capacity. It was quantitative in its overall approach. However, the researcher attempted to add some qualitative depth by questioning participants briefly during the testing procedures and by interviewing half a dozen participants at more length afterward. These interviews produced some insights into how language learners view fluency and how they try to maintain it during speech performance. It is hoped that, taken together, these experiments will provide some more general insights into L2 fluency, working memory and how the two might be related.

2.0 LITERATURE REVIEW

2.1 WORKING MEMORY

2.1.1 From short-term memory to working memory

Working memory traces its theoretical lineage at least as far back as James' (1890) description of "primary memory." Unlike "secondary memory," which contains vast stores of long-held knowledge more or less permanently, primary memory was said to contain information that has never been lost to consciousness. It belongs to "the rearward portion of the present space of time, and not to the genuine past" (James, 1890, p. 609). James' division of memory into remote and immediate constituents was revived and extended more than 60 years later. Miller's (1956) classic study placed the limit of the immediate component, now termed the "short-term store" or "short-tem memory," at seven items, plus or minus two, with variations depending on the individual. Broadbent (1958) further divided the immediate memory store into an "S-system" and a "P-system." The former was a very short-term, pre-attentive storage area for purely sensory information, similar to what would later be called "echoic memory" (Crowder, 1976). The P-system was where information available to conscious awareness was stored. Information had to pass through both of these systems in order to be integrated into the more stable long-term memory store.

Early research on short-term memory suggested that the best way to retain newly learned material was through vocal or subvocal ("covert") rehearsal. Peterson and Peterson's (1959) research showed that when rehearsal is prevented or hampered, information in short-term memory tends to decay rapidly and be forgotten. In Atkinson and Shiffrin's (1968) influential "modal model" of memory, rehearsal was a prime "control process" by which input from the senses was preserved. The longer that information remained in the short-term store, the more likely it was to be remembered for later retrieval. However, this emphasis on rehearsal was challenged by Craik and Lockhart (1972), who argued that the depth at which new information was processed was the key factor in retention. For example, a new word encountered while reading will probably be remembered better if it is processed at a "deep" semantic and/or phonological level than if it is processed only visually.

As Neath (1998, p. 55) notes, the various views of immediate or short-term memory developed in the 1950s and 1960s shared several basic features. The short-term component of memory was seen as having limited capacity and brief duration; it was primarily verbal in nature; and it was conceived as a buffer where information could be temporarily stored. It might be added that many researchers in that era allotted some role to the conscious manipulation of information in the transition from short-term to long-term memory stores. Whether that manipulation was envisioned as simple rehearsal or as a more elaborate, multi-layered encoding process, there seemed to be a consensus that some kind of attentional resources were important in determining what passed through the "bottleneck" to long-term memory.

2.1.2 Models of working memory

2.1.2.1 Baddeley's model. The concept of working memory developed by Baddeley and his colleagues (Baddeley, 2003; Baddeley, Gathercole & Papagno, 1996; Baddeley, 1986; Baddeley & Hitch, 1974, *inter alia*) incorporated some characteristics of earlier models while adding new elements. In their view, working memory is a "mental blackboard" or "temporary workspace" that allows humans to store and manipulate information while performing mental tasks (Baddeley, 2003, p. 672). It is a limited-capacity facility that combines the storage function of short-term memory (Miller, 1956; Broadbent, 1958) with the attentional control (Pashler, 2003; Norman & Shallice, 1986) needed to successfully utilize information. A simple arithmetic problem of the following type is commonly used to illustrate how working memory works: a man buys three candy bars at $1.50 each and gives the cashier a $10 bill. How much change should he receive? To solve this problem, most people will first perform one operation (1.50 x 3) and hold the answer in mind while performing a second operation (10 - 4.50). Working memory is also exercised in more complex mental activities such as planning chess moves and constructing sentences, and in fact it is believed to be fundamental to learning, reasoning and language comprehension (Goldman-Rakic, 1999, p. 92).

Baddeley and Hitch (1974) originally proposed that working memory consists of three main components: the central executive and two short-term memory stores, the visuospatial sketch pad and the phonological loop. The visuospatial sketch pad records visually presented sensory data. The phonological loop stores verbal information, and it seems to be adapted for language learning in that it helps keep new phonological strings active and in the correct serial order while meaning is attached to them (Baddeley, 2003, p. 673). These strings decay in about two seconds unless refreshed by subvocal rehearsal. Baddeley and Hitch's decision to separate these two slave systems, rather than have one "sensory" input channel, was motivated by experimental evidence indicating that there was little interference effect between the two channels. For example, having subjects repeat lists of digits had little or no effect on recall of visually presented items (Bower, 2000, p. 22).

The central executive is the component that does much of the work in working memory. It is "an attention and control system" (Groeger, 1997, p. 49), and it is crucial in the suppression of irrelevant information (Engle, 2001). The central executive also coordinates information flow between the two slave systems, focuses and switches attentional focus, manipulates new information and directs the retrieval of old information from long-term memory (Baddeley & Logie, 1999, pp. 28-30; Gathercole & Baddeley, 1993, pp. 4-8). This long list of duties has led to charges that Baddeley's central executive is in reality a homunculus, a theoretical stand-in for phenomena that cannot be accounted for otherwise. Baddeley (Baddeley & Logie, 1999; Baddeley, 1996) has in fact acknowledged that the precise make-up of this component has not been well specified, and that it remains open to debate and investigation. [2]

Baddeley and Hitch's model differs from earlier short-term memory models primarily in its emphasis on active manipulation of information rather than passive storage. The tools developed to measure working memory capacity differ accordingly. Rather than asking subjects to recall lists of random digits, as in Miller (1956), the memory span tests devised by Daneman and Carpenter (1980) and others required subjects to simultaneously remember items (often words) and process them (for example, by reading aloud sentences where those words appeared).

Such tasks are assumed to exercise both the storage and processing capabilities of working memory.

More recently, Baddeley (2000) has added another element, termed the "episodic buffer," to the original model. The episodic buffer helps integrate information from the two slave systems with information in long-term memory stores. This revision was made partly in response to experimental evidence and case studies indicating that individuals with short-term memory deficits (either temporary deficits induced in a laboratory or natural amnesia-related deficits) were capable of storing and manipulating information that was too complex to be held in the limited-capacity slave systems. [3] Since the central executive does not have a storage function [4], Baddeley (2000, p. 419) theorized that there must be a "back-up store" that could account for these phenomena. This modification is significant because it weakens the separation of working memory and long-term memory that had been assumed in Baddeley and Hitch's earlier model. [5]

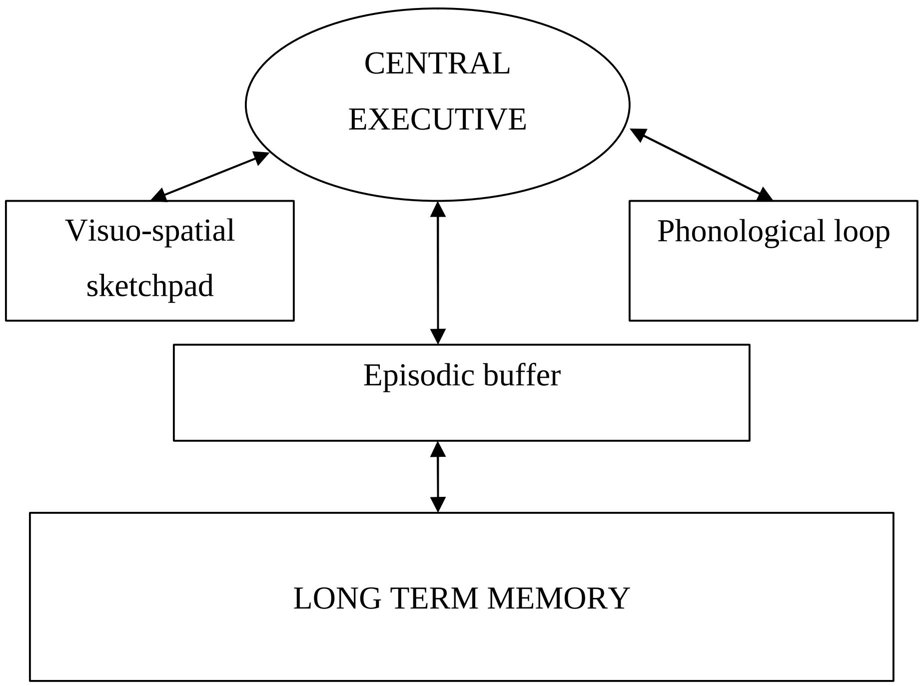

Figure 2: Revised model of working memory (Baddeley, 2000)

2.1.2.2 Alternative models of working memory. Although influential, Baddeley's model of working memory is only one of several such models. A fundamental theoretical question is whether working memory should be viewed as a single functional construct, and, if not, how it should be subdivided. Baddeley's model contains two domain-specific components, one for visual-spatial information and one for verbal information. This view has been adopted by other researchers who study span and/or loop capacity, and its influence on language (Just, Carpenter & Keller, 1996; Just & Carpenter, 1992; Daneman & Green, 1986; Daneman & Carpenter, 1980).[6] However, Engle and others (Engle, Kane & Tuholski, 1999; Conway & Engle, 1996) argue that working memory is a unitary, domain-free construct that is strongly related to general fluid intelligence. Although there are domain-specific codes and maintenance mechanisms (i.e., the phonological loop for verbal information), the core of working memory lies elsewhere, in the ability for controlled attention. Engle and his colleagues propose that individual differences in this area are more important to explaining learning outcomes than phonological loop capacity or memory span capacity.

On the other hand, other researchers divide the verbal portion of working memory more finely than Baddeley. Waters and Caplan (1999; 1996) believe that Baddeley's model and the measurement tools based on it fail to adequately account for different levels of linguistic processing, particularly syntactic processing. Based on research with both normal subjects and subjects with short-term memory deficits, they maintain that the working memory resources used in assigning syntactic structure is distinct from the resources tested in memory span tests. Waters and Caplan asked both low- and high-span subjects to listen to sentences with relative clauses and garden path sentences such as those in Table 1.

Table 1: Sample sentences from Waters and Caplan (1996).

a.	**The boy that the girl pushed kissed the baby.**
b.	**The experienced soldiers warned about the dangers conducted the midnight raid.**

They found no differences in the way the two groups of subjects processed these sentences despite earlier research (King & Just, 1991) suggesting that the low-span readers would be hindered. Roberts and Gibson (2002) take a similar view, saying that verbal working memory includes a distinct propositional element that is absent in the Daneman-Carpenter-Just approach to measuring working memory. Like Caplan and Waters, they envision multiple resource pools that are called upon in language comprehension, resources that are not fully quantified in span and loop tests.

Other models of working memory place more emphasis on processing mechanisms than on distinct and stable components. In Schneider's (1999) connectionist perspective, working memory consists of short-term patterns of activation between modular processors consisting of neuron-like units. These patterns and the strength, or "weight," of the connections within them are partly controlled by an executive that controls attentional resources. Working memory is limited not so much by span capacity, as in the Danemen-Carpenter-Just approach, but by interference effects and the limits of the executive function. These limits are gradually lessened as expertise develops in a learner, allowing for faster parallel processing instead of the slower, serial processing typical of novices. Similarly, O'Reilly et al. (1999) describe working memory in functional terms as a system that is distributed over several parts of the brain, notably the prefrontal cortex and the hippocampus. In their view, working memory is in fact the activated part of long-term memory.

This view of working memory is consistent with Ericsson's model of "long-term working memory" (Ericsson & Delaney, 1999; Ericsson, 1996; Ericsson & Charness, 1994). In this model, working memory is not entirely separate from long-term memory; it is the set of mechanisms that allow for retrieval of information from long-term memory. Encoding is paramount because it is the type and quality of encoding that determines how quickly and skillfully information can be accessed. Ericsson's work draws on the study of expert performance in diverse mental activities, including playing chess and remembering food orders in a restaurant. However, Ericsson and Kintsch (1995) argue that the same basic processes can help explain more commonplace tasks such as language learning and use.

Some of these alternative viewpoints of working memory are at least partly compatible with Baddeley's model. For example, Engle's "controlled attention" model contains a central executive component as well as the functional equivalent of Baddeley's phonological loop, even though it gives less prominence to the latter in determining learning outcomes. The "multiple resource" perspective advocated by Caplan and Waters does not seem to represent a broad theoretical challenge to Baddeley's model, so much as a more complex view of how working memory may affect language learning. It might be possible to accomodate this view by including more detailed measures of verbal working memory, including propositional memory, into the array of psychometric tools.

On the other hand, the connectionist-oriented models proposed by Schneider and by O'Reilly and colleagues, and Ericsson's long-term working memory schema, are further removed from the basic tenets of Baddeley's model. Baddeley's model is a modal one that presents working memory as a "gateway or stepping stone" to long-term memory (Miyake and Shah, 1999a, p. 14). By contrast, the connectionists and Ericsson see no sharp boundary between working memory and long-term memory. If, as they propose, working memory is essentially the attended-to or activated portion of long-term memory, then its importance in learning will be diminished relative to the Baddeley model. Other cognitive functions such as encoding, consolidation and retrieval of learned material then become more significant.

2.1.3 Working Memory and Language

2.1.3.1 Daneman and Carpenter (1980*).* Daneman and Carpenter (1980) introduced the Reading Span Test (RST) as a means of quantifying working memory capacity. They asked their subjects, 20 undergraduate students at Carnegie-Mellon University, to read aloud sets of sentences and then try to recall the final word of each sentence when cued. This was assumed to exercise both processing capacity (by requiring subjects to read the sentences) and storage capacity (by requiring them to remember the final words). Each sentence was 13 to 16 words long, as in the examples in Table 2.

Table 2: Sample sentences from Daneman and Carpenter's (1980) Reading Span Test

a.	**When at last his eyes opened, there was no gleam of triumph, no shade of anger.**
b.	**The taxi turned up Michigan Avenue where they had a clear view of the lake.**

In this two-sentence set, the subjects would attempt to recall the words “anger” and “lake.” They were presented first with three two-sentence sets, then three three-sentence sets, and so on, up to three six-sentence sets. If a subject completed the task successfully on two of three sets, he or she would continue to the next largest set. A subject’s reading span capacity was the largest number of sentences, from 2 to 6, that he or she could complete. As it turned out, the span scores for Daneman and Carpenter’s subjects ranged from 2 to 5, with a mean of 3.15 (S.D. = .93).

Daneman and Carpenter’s chief purpose was to determine whether these span scores correlated strongly with measures of L1 reading skill. They measured reading skill in three ways. The first two measures were based on a reading comprehension test wherein the subjects read 12 relatively short (approximately 140 words) narrative passages silently at their own pace. When they had finished each passage, the texts were removed from their view and they were asked two questions. The first question asked them to identify the referent of a pronoun located in the last sentence of the passage. The distance between the pronoun and its referent was systematically varied, between two and seven sentences. The second question was a general comprehension question that focused on a particular fact within the passage. Finally, Daneman and Carpenter recorded their subjects’ scores on the verbal (reading) portion of the Scholastic Aptitude Test, a standard entrance exam for college-bound students in the United States (it should be noted that these SAT scores were self-reported).

As seen in Table 3, Daneman and Carpenter found significant correlations between working memory span scores and the three measures of reading skill. The very high (.90) correlations with scores on the pronominal referent test is particularly noteworthy. Another interesting fact is that the researchers found only weak-to- moderate correlations between the reading measures and scores on a simple word span test that they also gave their subjects. Here, subjects were asked to recall sets of individual words, with the sets ranging in size from two words to seven. This was in effect a short-term memory test, as it did not require any reading or language processing beyond the word level, as the RST did.[7]

Table 3: Correlations between WM spam and reading skill (Daneman & Carpenter, 1980)

	Pronoun reference	**Fact question**	**Verbal SAT scores**
RST span	.90**	.72**	.59**
Word Span Test	.33	.37	.35

** $p < .01$

Daneman and Carpenter carried out a second experiment, using 21 Carnegie-Mellon University undergraduates as subjects. In this experiment, instead of a Reading Span Test, they administered a Listening Span Test (LST), where information was presented orally. Their results were quite similar. Their subjects' LST scores ranged from 2 to 4.5, with a mean score of 2.95 (S.D. = .72). Correlations with reading scores were again strong, as shown in Table 4.

Table 4: Correlations between Listening Span Test (LST) scores and reading skill measures (Daneman & Carpenter, 1980).

	Pronoun reference	**Fact question**	**Verbal SAT scores**
LST scores	**.72 ****	**.67 ****	**.53 ****

** $p< .01$

The results of the second experiment suggested to Daneman and Carpenter that working memory capacity was not skill-specific but rather remained fairly constant across modes of language presentation.

More generally, Daneman and Carpenter theorized that high-span readers benefit from being able to maintain more items in working memory than low-span readers. This might allow them to form more and deeper associations, or "chunks," between items, which in turn allow for greater storage and easier retrieval.

> The good reader has more functional working memory capacity available for the demands of chunking. He is more likely to have more concepts and relations from preceding parts of the text still active in working memory. Consequently, he

> should be able to detect and to note their relative importance. The good reader's chunks should be richer, and more coherent, and contain different information. The presence of different interrelationships could subsequently allow different inferences and generalizations to be drawn. (p. 464)

In this manner, Daneman and Carpenter speculated, quantitative differences in working memory capacity result in qualitative differences in reading comprehension, and help explain individual differences in reading skills.

Daneman and Carpenter's experiments were seminal in several respects. Most importantly, they established a standard measure of working memory capacity, one that, with modifications and extensions, remains widely used a quarter century later. Also, they focused on a specific reading skill, the ability to connect pronouns and referents, in addition to general comprehension skills. The connection between this particular skill and some type of short-term memory capacity seems plausible, and their results strengthen this plausibility. Finally, the results of their second experiment appear to confirm that working memory is a fundamental source of individual differences in language skill, and not a mode-dependent variable.

However, important questions about Daneman and Carpenter's approach need to be answered. One question is whether the RST they devised actually taxed their subjects' processing skills. Perhaps having college students read aloud sentences in their native language engaged very little of the processing capabilities of working memory because their reading skills had become highly automatized. Omaki and Ariji (2003) make just such an argument. They propose making some sentences in the RST ungrammatical and adding a grammaticality judgment task to the word-recall task in order to ensure that subjects are paying attention to what precedes the final word. (In their Listening Span Test, Daneman and Carpenter did add true/false questions in an effort to achieve this goal, but the results of those questions did not enter into the subject's score, which was based on final word recall.)

Another question is whether Daneman and Carpenter's pronoun reference test really tested working memory as they and Baddeley conceived it. The pronouns and their referents were often separated by several intervening sentences, and the subjects did not know in advance what the pronoun in the last sentence would be or what it would refer to. It seems likely that in many instances more than 2 seconds passed between the time when the subjects encountered the referent and the moment when they came to the pronoun and were asked about its referent (it should be remembered that the readings were self-paced). This time lag surpasses the purported limit of working memory unaided by subvocal rehearsal.[8] An alternate explanation is that, at least for long-distance referent-pronoun relations, the subjects were relying on some sort of short-term propositional memory. This explanation would be in line with the multi-component view of verbal working memory espoused by Caplan and Waters (1999, 1996) and Roberts and Gibson (2002).

2.1.3.2 Further research on Working Memory and Reading. Following the acceptance of the Reading Span Test as a measure of working memory capacity, a flurry of correlational studies were conducted on working memory and language skills. King and Just (1991) found that subjects who scored relatively high on the RST were better than low-span subjects at comprehending object relative sentences such as *The reporter that the senator attacked admitted the error*. They proposed a scheme of on-line parsing, called the "CC (Capacity Constrained) READER" model, that relies on working memory as an important source of individual differences in syntactic processing and general reading skill. In this model, as in Daneman and Carpenter's view of working memory, storage and processing components compete for cognitive resources, so that quantitative advantages in storage space free up processing abilities and lead to qualitative advantages in overall comprehension. King and Just proposed that this advantage shows up especially clearly when readers must parse difficult sentences such as object relative clauses. The advantage for high-span readers will be less obvious in subject relative sentences (i.e., *The reporter that attacked the senator admitted the error*) where there is no switch in syntactic or thematic role.

Miyake, Just and Carpenter (1994) found that RST scores coincided with another reading skill, the ability to resolve lexical ambiguities in context. They tested high-, mid- and low-span readers by having them read sentences like those in Table 5.

Table 5: Lexically ambiguous sentences (Miyake, Just & Carpenter, 1994)

a.	**Since Ken really liked the *boxer,* he took a bus to the nearest *sports* arena to see the match.**
b.	**Since Ken really liked the *boxer,* he took a bus to the nearest *pet* store to buy the animal.**

In each sentence, the meaning of the ambiguous word (here, *boxer*) is not clarified until later in the sentence, when a key word (*sports, pet*) provides a semantic clue. Miyake, Just and Carpenter found that high-span readers were better and faster than the two other groups in comprehending the ambiguous words. This was especially true when the semantic clue pointed to the less common usage of the ambiguous word (i.e., *boxer* = "dog") rather than the more widely used interpretation (*boxer* = "pugilist"). The advantage for high-span readers was also more pronounced when the distance between the ambiguous word and the semantic clue was increased. The researchers theorized that high-span readers were able to maintain both interpretations in working memory longer than their lower-span counterparts, which helped them resolve ambiguities correctly.

In summary, the research reviewed above indicates that high-span readers enjoy advantages over low-span readers at the word level, the sentence level and the discourse level. At the word level, they are better at resolving lexical ambiguities (Miyake, Just & Carpenter, 1994); at the sentence level, they are superior at interpreting syntactically complex phrases (King & Just, 1991); at the discourse level, they can connect pronouns and referents better, especially across long distances (Daneman & Carpenter, 1980). A common thread between these sub-skills appears to be the ability to select correct interpretations from among competing alternatives. This is clearest in the case of lexical ambiguity, where one form temporarily has two plausible meanings. But it is also present at the syntactic level when readers must quickly decide whether a noun is a subject/agent or object/patient. It may also exist at the discourse level when a referent and its pronoun are far apart and the text contains other possible referents.

This common thread is worth noting because the researchers in these studies explained their results within a Baddeley-type framework of working memory, where additional storage capacity was assumed to free up processing resources. However, these findings are also compatible with a model of working memory that emphasizes attentional control and inhibition of incorrect alternatives, a model more like that proposed by Engle (Engle, Kane & Tuholski, 1999; Conway & Engle, 1996). In this sort of model, storage capacity is a peripheral issue; what matters most is the speed and efficiency of executive control, the capacity to switch attention rapidly between items, to evaluate them and eliminate undesirable choices.

Another interesting question is whether the advantages that high-span readers reportedly enjoy extend to oral language. As stated in the introduction, there is reason to believe that the speed of spoken language places a premium on working memory capacity. On the other hand, there is the countervailing factor of inherent complexity. As Crystal (1997, p. 181) points out, written language tends to be more compact and intricately structured than speech. Speakers have to "think standing up," in Crystal's words, and rarely engage in complex preplanning; their output is more loosely structured than writing and contains more repetition and rephrasing. For example, center-embedded relative clauses like King and Just's (1991) example, *The reporter that the senator attacked admitted the error*, would be unremarkable in many journalistic or academic texts. But in spontaneous speech many people would probably choose a form that places less strain on their listener's processing abilities, perhaps *The reporter admitted that he made an error after the senator attacked him*. Likewise, the sort of long-distance separation of pronouns and referents that Daneman and Carpenter (1980) tested their subjects on would be unusual in everyday spoken discourse.[9] The matter of working memory and spoken language is addressed below.

2.1.3.3 Phonological short-term memory and language learning. One line of working memory research has focused on the role of the phonological loop, or phonological short-term memory (PSTM), in vocabulary learning and general language acquisition. Gathercole and Baddeley (1989) tested English-speaking 4-year-olds and found that their PSTM capacity, as measured by a non-word repetition task, was a good predictor of their subsequent L1 vocabulary growth. Service (1992) tested 41 native Finnish-speaking children who were studying English on their ability to mimic English-like pseudo-words such as "rendance" and "disajoinance." She found that the children's repetition accuracy, as measured by a native English listener, correlated strongly with their L2 skills, as measured by a three-part English test (**Table 6**).

Table 6: Correlations between repetition accuracy and ESL skills (Service, 1992)

	Corrrelation w/ repetition accuracy
Listening comprehension test	.62 ***
Reading comprehension test	.74 ***
Writing test	.58 ***
Overall ESL score	.65 ***

*** $p < .001$ (one-tailed test)

The connection between the phonological loop and vocabulary learning appears to depend on articulatory rehearsal. When subvocal rehearsal is suppressed, learners' ability to learn new words is significantly decreased (Papagno, Valentine and Baddeley, 1991).

Ellis and others (Ellis & Schmidt, 1997; Ellis & Sinclair, 1996; Ellis, 1996) have argued that the importance of the phonological loop in language learning goes beyond vocabulary learning to the acquisition of syntax. In one study (Ellis & Sinclair, 1996), English-speaking learners of Welsh who were encouraged to repeat novel utterances outperformed learners who were prevented from doing so (in this case by having them count from one to five during learning). They did better than their counterparts not only in terms of L2 vocabulary knowledge and pronunciation, but also in comprehension and translation of L2 words and phrases, metalinguistic awareness of syntactic rules and morphosyntactic accuracy. According to Ellis and Sinclair (1996, p. 247), the phonological loop influences grammatical development by allowing learners to acquire "chunks" of language, that is, longer sequences of multi-word speech, in the correct serial order. This has a cyclical effect because once these chunks are stored in long-term memory, they facilitate the recognition and acquisition of related items, effectively improving the operation of working memory for these items.

A few notes of caution are in order here. Firstly, Ellis and Sinclair's experiment was very limited both in terms of the L2 learning material and in terms of learning time. Subjects attempted to learn 10 Welsh nouns, two short phrases (the English translations are "where is ...?" and "his ...") and one morphosyntactic rule (the Welsh "soft mutation" of word-initial phonemes) in less than two hours of training. Extrapolating from these results to larger questions of syntactic acquisition and consolidation of learning over time seems risky. Also, Ellis and Sinclair's results say nothing directly about PSTM capacity and individual differences, only that learners who are prevented from practicing subvocal rehearsal suffer a disadvantage. The researchers did not compare high-span and low-span learners in this study, and the question of whether high PSTM capacity aids significantly in the language learning beyond the lexical level remains open.

2.1.3.4 Working memory and oral fluency. Although much of the research into working memory and language has focused on reading skills, a small but important part of this research has been devoted to speech production. Three studies are of particular relevance. Daneman and Green's (1986) study was doubly significant, firstly for introducing the Speaking Span Test (SST) as a measure of productive working memory, and secondly for correlating working memory scores with lexical fluency. Daneman (1991) took a more comprehensive view of oral fluency, and attempted to correlate L1 fluency with working memory capacity. Finally, Fortkamp's (1999) study of English L2 users in Brazil extended Daneman's research into L2 fluency.

In the Speaking Span Test, subjects silently read sets of words, one at a time, on a computer screen. Each word is displayed for 1 second. After the last word in the set is presented, the subject attempts to produce a grammatical sentence for each word he or she has just seen. For example, if they saw the words *quarter* and *battled*, they might make sentences such as "I put a *quarter* into the juke box" and "He *battled* to save his country" (Daneman & Green, 1986, p. 11). As in the Reading Span Test, the sets of words increase in size as the test progresses, from two up to a maximum of six words. There are five sets of words at each level (i.e., five two-word sets, five three-word sets). To complete a level successfully, a subject must produce a grammatical sentence for all of the words in at least three out of the five sets. His or her speaking span is the highest level he succeeds at, with a half point given for performing the task successfully two out of five times.

Before looking at Daneman and Green's results, it is important to note why they introduced the Speaking Span Test to supplement the Reading Span Test. The researchers state briefly that "working memory is not a general system with a unitary capacity" (p. 17), and they argue that different span tests are appropriate for reading and speaking.

> The important contribution here is to show that processing skill differences in working memory are highly task specific...(W)hen individuals have

> to produce vocabulary in context, it is their efficiency at sentence production processes, not comprehension processes, that will determine functional working memory capacity for the task. (p. 15)

This passage is interesting because it turns on its head the purported relationship between working memory and individual differences: span capacity becomes an expression of proficiency in particular skills rather than an independent determinant of general linguistic ability. A person who reads well and often may have a larger RST capacity by virtue of this fact. This accords with the task-specific view of working memory outlined by Cantor and Engle (1993), and it suggests that individual differences in working memory are dependent on long-term knowledge in specific skills.

For Daneman and Green's 34 subjects, the range of span scores on the Speaking Span Test ranged from 2 to 4.5, with a mean of 3.28 (S.D. = 0.63). Daneman and Green also recorded how many total words, out of 70, that their subjects successfully completed the task for. Here, the subjects' scores ranged from a low of 31 words to a high of 57, with a mean of 45 (S.D. = 0.7). The same subjects also took the RST and the overall scores were similar. Reading span scores ranged from 2 to 5 (mean = 3.5, S.D. = 0.89), and the total number of words successfully recalled ranged between 30 and 64 (mean = 45, S.D. = 0.8). The correlation between reading span and speaking span scores was 0.57 ($p < .01$), a moderately significant relationship. Daneman and Green interpreted these results as indicating that the two span tasks tapped into the same memory component of working memory but also exercised distinct, skill-specific processing resources.

Daneman and Green's study was also important because it focused on oral fluency, at least at the word level. In addition to the two span tests, the researchers tested their subjects on how easily they could produce synonyms for words presented in context. Each subject saw 20 sentences like those in Table 7, each with a target word at the end.

Table 7: Sample sentences from Contextual Vocabulary Production Task (Daneman & Green, 1986)

a.	**He expected the euphoric feeling to last forever, but its effects were only MOMENTARY.**
b.	**An abundance of rich food and drink were provided for the guests at the FEAST.**
c.	**The design of the electronic components is very INVOLVED.**
d.	**No one was quite sure if the statements he made were intended to confuse everyone or if they were just ABSURD.**

The subjects were directed to produce a context-appropriate synonym for the target words as quickly as they could (for example, for the first sentence, "temporary" or "ephemeral" are appropriate). Their responses were recorded and their response times were recorded in milliseconds. Mean response time was taken as a measure of each subject's lexical fluency.

Daneman and Green found that both speaking span and reading span correlated significantly with scores from this lexical fluency test. This was true whether or not subjects' errors were counted (when subjects produced an incorrect synonym). However, speaking span correlated with lexical fluency more strongly than reading span, as shown in Table 8.

Table 8: Correlations between WM span and lexical fluency (Daneman & Carpenter, 1986)

	Fluency: Correct items	**Fluency: All items**
Speaking span	.60 **	.56 **
Reading span	.41 *	.33

* $p < .05$, ** $p < .001$

As mentioned above, the two span scores correlated moderately with each other. When the effect of reading span was factored out, the correlation between speaking span and lexical fluency was still significant at 0.49. However, when speaking span was factored out statistically, the correlation between reading span and lexical fluency fell to an insignificant .09. These figures support the contention that working memory's influence on language is highly skill-specific, at least with regard to speech production and reading comprehension. [10]

While Daneman and Green's study was purposely limited to fluency at the word level, Daneman (1991) attempted to connect working memory to fluency in a larger sense. In addition to administering both the RST and SST to her subjects, Daneman gave them three oral fluency tests. The most comprehensive of these was the Speech Generation Task (SGT), wherein subjects were directed to describe in their native language (English) a picture as completely as they could in 60 seconds. The picture chosen for this task was titled "How to dispense with servants in the dining room," and it depicted a family being served by a group of robot-like mechanical devices. Daneman asked her subjects to describe the picture as fluently and originally as they could. This test produced two fluency scores. The first was a simple count of how many words each subject produced in the allotted time. The second score measured "richness" of speech and was more subjective: two native English speakers listened to recordings of the subjects' speech and rated them from 1 (repetitious and semantically empty) to 5 (creative and semantically rich). Daneman used the average score from the two judges.

In this experiment, the SST scores were recorded in two different ways. In the first, called "speaking span strict," subjects were given credit only if they recalled the target words in their exact form: if the word was "measure," they received no credit for making a sentence with "measured" or "measures." In "speaking span lenient," on the other hand, subjects did receive credit for using morphological variations of the target word. This latter, more lenient scored produced moderately strong correlations with the Speech Generation Task in both the number of words produced and in terms of richness. By contrast, the speaking span strict scores correlated much less strongly with the number of words produced and slightly less strongly with richness (see Table 9).

Table 9: Correlations between WM scores and Speech Generation Task (SGT) scores (Daneman, 1991)

	Speaking Span strict	Speaking Span lenient	Reading Span
Number of words	.24	.48 **	.14
Richness	.43 *	.47 **	.33

* p < .05, ** p < .01

Again, the weak relationship between the RST scores and both of the oral fluency measures supports the skill-specific view of working memory.

Another interesting point is the large difference that the SST scoring method made in terms of correlations with word productivity: the correlation is twice as strong when the lenient score is used as when the strict score is used. Daneman (1991, p. 457) interprets this difference as indicating that the lenient score, which gives credit for more flexible morphological constructions, is a better predictor of the fluency exercised in a relatively free-form task like the Speech Generation Task. However, it may be that this difference was in reality an artifact of methodology. It is unclear whether Daneman instructed her subjects to use the exact form of the target word, or let them decide whether this was allowable. She notes (ibid., p. 451) that a morphological change could have reflected either a subject's inability to remember the exact word, or else a conscious decision "to remove the constraint of having to utilize it in the prescribed syntactic role." That is, some subjects may have found it easier to make a sentence with *danger* rather than the target word *dangers,* and done so because they thought this was close enough. In that case, variations in the strict scores may have reflected differences in how subjects understood the task rather than actual differences in working memory capacity. [11] In any case, Daneman concluded (ibid., p. 457) that working memory does indeed have an important role in producing fluent and creative speech, not only at the word level but beyond. [12]

Daneman employed two other fluency measures in addition to the Speech Generation Task. One was the Oral Reading Task, where subjects were asked to read a 320-word passage from a work of fiction (*The Great Gatsby*) as quickly and accurately as they could.

They were judged on the speed of their reading and also on the number and types of reading mistakes they made (i.e., repetitions, false starts, mispronunciations, etc.). The other measure was called the Oral Slip Task, and it involved quickly reading word pairs that were purposely designed to cause slips of the tongue. For example, the subjects were asked to say aloud the pairs in Table 10 (Daneman, 1991, p. 454).

Table 10: Example stimuli from Oral Slip Task (Daneman, 1991)

SILVER PORT
SILLY POEM
SIP POLE
PICK SOAP

The idea was that by presenting three word pairs with an "S... P..." pattern, the experimenter might cause subjects to say "sick pope" instead of "pick soap" for the last item. Subjects were scored on the number of errors they made, with a maximum of 30.

Correlations between working memory measures and these two oral fluency scores are presented in Table 11 (adapted from Daneman, 1991, p. 456). There are negative correlations between speaking span and the two other fluency measures. On the Oral Reading Task, high-span subjects took less time to finish the *Gatsby* passage, and made fewer errors while doing so. As Daneman had predicted, the correlation was higher when the strict speaking span score was used as a measure of working memory. High-span subjects also made fewer mistakes on the Oral Slip Task, at least when the stricter scoring method was used to define "high-span." Reading span correlates significantly only with Oral Reading Task scores, again underlining the skill-specific view of working memory.

Fortkamp (1999) adapted Daneman's approach to examine the relationship between working memory and L2 oral fluency. Fortkamp used the same three fluency measures that Daneman had used, and some of the same working memory measures. However, her subjects were 16 advanced ESL speakers from Brazil. Fortkamp had her subjects take both the RST and

SST in their L1 (Portuguese) and in their L2 (English). The three fluency tests- the Speech Generation Task, the Oral Reading Task and the Oral Slip Task- were given in English alone.

Table 11: Correlations between WM scores and oral fluency scores (Daneman, 1991)

	Speaking Span strict	Speaking Span lenient	Reading span
Oral Reading Task: reading time	-.56 **	-.47 **	-.41 *
Oral Reading Task: errors	-.43 *	-.24	-.54 **
Oral Slip Task: slips	-.47 **	-.24	-.21

* p < .05, ** p < .01

Table 12: Correlations between WM and Speaking Generation Task (SGT) scores (Fortkamp, 1999)

	SST Portuguese: strict	SST Portuguese: lenient	SST English: strict	SST English: lenient
SGT words/min.	-.08	.19	.64 *	.61 **
SGT richness	NA	NA	.42	.35

* p < .05, ** p < .05

The main finding to emerge from Fortkamp's study was the weak relationship between L1 working memory span and L2 oral fluency as measured by the Speech Generation Task. [13] Indeed, as shown in Table 12, the correlation between these two factors was sometimes negative. This contrasted with the moderate correlation between L2 working memory and L2 oral fluency. Taken at face value, Fortkamp's results strongly indicate that working memory capacity is language-dependent. That is, "English working memory" has little to do with "Portuguese working memory" (GM's terms), and only the former can be expected to correlate significantly with the ability to speak English fluently.

However, some interesting questions are unresolved. All of Fortkamp's subjects were advanced L2 learners; all were earning graduate degrees in English and used English frequently in both writing and speech. It might well be that working memory capacity has a stronger influence on L2 oral production in the early stages, when the processes underlying speech are less automatized (Temple, 1997). A study with a larger body of subjects, one that included beginning and perhaps intermediate learners, would address this question.

Also, some points about Fortkamp's results are unclear. Table 13 shows that, overall, the subjects' scores on the English and Portuguese working memory tests were quite similar. Language did not seem to matter; indeed, in the SST, group scores were virtually identical (note also the relatively small standard deviations).

Table 13: Mean scores from Fortkamp's (1999) WM tests

	Mean (S.D.)
Speaking Span strict: Portuguese	21.5 (2.7)
Speaking Span strict: English	21.4 (2.8)
Speaking Span lenient: Portuguese	23.3 (2.1)
Speaking Span lenient: English	23.4 (3.5)
Reading Span: Portuguese	27.0 (4.7)
Reading Span: English	24.1 (5.1)

In light of this, it is difficult to explain the weak correlations between "English working memory" and "Portuguese working memory," as measured by the SST. These scores are shown in Table 14. Of course, these correlations represent individual scores rather than group scores, and it is possible that the similarity between the group scores masks large individual differences. However, the relatively small standard deviations in the working memory tasks suggest otherwise.

The robust correlation (r = 0.78) between English and Portuguese RST scores is interesting also. Why should reading span capacity match so well cross-linguistically while speaking span capacity does not? These apparent discrepancies are not addressed in Fortkamp's paper.

More generally, the approach to measuring oral fluency adopted by both Daneman (1991) and Fortkamp (1999) deserves attention. The Speech Generation Task is the most comprehensive of the three tasks they used, and it is a semi-naturalistic task that requires subjects to speak on a common topic as fluently and accurately as possible. They must perform all the steps that speakers usually do in producing utterances: they must plan what they will say, access the appropriate words and syntactic and intonational patterns, and articulate utterances. The two other tests, the Oral Reading Task and the Oral Slip Task, focus more specifically on execution and articulation of speech, as the content is already provided by the researcher. (Daneman chose not to use the synonym production task, which Daneman and Green (1986) had used as a gauge of lexical fluency.)

Table 14: Correlations between WM scores (Fortkamp, 1999)

	Speaking Span English: strict	**Speaking Span English: lenient**	**Reading Span: English**	**Reading Span: Portuguese**
Speaking Span Portuguese: strict	.20	.16	.16	.09
Speaking Span Port.: lenient	.11	.27	.36	.27
Reading Span: Portuguese	.13	.17	.78*	–
Reading Span: English	.33	.35	–	.78 *

* p < .05

Taken together, these three tools- the Speech Generation Task, the Oral Reading Task and the Oral Slip Task- seem to provide a fairly good general picture of fluency, at least L1 fluency. There are measures of speed and creativity, as well as articulatory accuracy (note that accuracy, at least low-level phonological accuracy, is an issue even in L1 speech). But why these tests were selected and not others (for instance, why not an interactive speaking task, or a narrative task instead of a description? why not a lexical fluency task?) is not fully explained. The question of what constitutes fluency and how it should be described in L1, let alone L2, is an open question, one that will be addressed below.

2.1.3.5 Working Memory in Levelt's (1989) Model of Speech Production. Before addressing the question of fluency directly, it is necessary to examine Levelt's (1989) model of speech production and how it accounts for fluent speech. It will be helpful to see what roles Levelt assigns to working memory in this process. Levelt's model is presented in some detail because it provides the theoretical framework for the present discussion and the experiments described below.

The blueprint of Levelt's model is shown in Figure 3. There are three main components: the Conceptualizer, the Formulator and the Articulator. Each of these is presented as a highly autonomous part of the production chain, with a characteristic input and/or output, as will be described below. The Speech Comprehension System is important in monitoring, but as it is presumed to be the same system that monitors other people's speech it is not considered part of the production chain proper. There are also two relevant long-term knowledge stores. The first is the lexicon, which includes both the lemmas (the semantic and syntactic forms) of words and the outward phonological realization of those words. The second store includes non-lexical information in long-term memory plus knowledge of the current discourse situation- what has been said before, which register of speech is being used, etc.

The Conceptualizer is responsible for producing preverbal messages in the message generator. It does this in two stages, which Levelt terms "macroplanning" and "microplanning." Macroplanning involves conceiving a communicative goal, creating subgoals and retrieving the necessary information for expression of these subgoals. Microplanning entails giving propositional shape to these subgoals and the overall message, and selecting an informational perspective so that, for example, one entity is chosen as the topic of an utterance. The message generator deposits its results in working memory, which contains "all the information currently available to the speaker...the information *attended to* by the speaker [italics Levelt's] (Levelt, 1989, p. 10)." The working memory store also contains a "small, focused part" of the discourse record, which allows speakers to keep track of what they and their interlocutors have said (ibid., p. 10).[14]

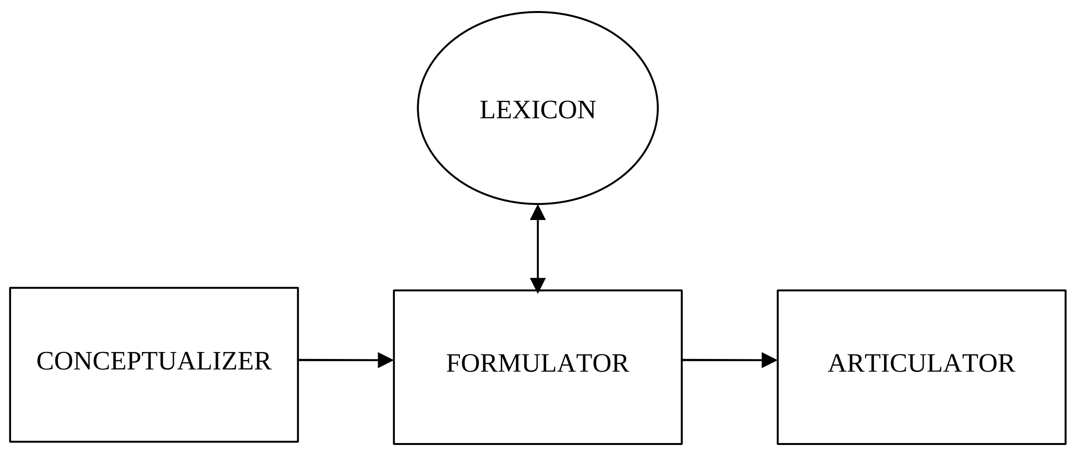

Figure 3: Levelt's model of speech production

The Formulator “translates a conceptual structure into a linguistic structure” (Levelt, 1989, p. 11). It takes the preverbal message, encodes it grammatically and phonologically, and produces a phonetic or articulatory plan, or “internal speech.”. The grammatical encoder accesses lemmas from the lexicon, builds syntactic constituents such as noun phrases and verb phrases and produces a linearized string of constituents. This interim product is deposited not in working memory but rather in a separate store which Levelt (1989, p. 12) calls the “Syntactic Buffer.” The phonological encoder accesses the formal part of the lexicon and assigns morphological and phonological form to the utterance, including prosodic features.

The Articulator’s function is to take the phonetic plan produced by the Formulator and to execute it physically by coordinating the movements of the speaker’s respiratory system and muscles. The resulting output is overt, external speech. As production of overt speech often lags behind the generation of internal speech, Levelt postulates another storage device, an “Articulatory Buffer,” to hold as-yet unuttered phonetic plans.

The Speech Comprehension System allows speakers to attend to the overt output of the Articulator, to check this output for errors of meaning and form. It also monitors parsed internal speech, which is held in working memory (Levelt, 1989, p. 13). This allows speakers to modify partly-delivered utterances or to abandon undelivered utterances, or to ignore the error and knowingly produce an imperfect utterance, perhaps because the error is not considered important enough to halt for. Levelt adds that speakers can monitor messages before they go to the Formulator, "considering whether they will have the intended effect in view of the present state of the discourse and the knowledge shared with the interlocutor(s)" (ibid., p. 14).

In Levelt's model, attention and working memory are most important in the conceptualization stage and in monitoring, less important in formulation and articulation. Creating a communicative intention clearly involves attentional control, as does checking speech (internal or external) against one's knowledge of linguistic norms and the current discourse situation.

> Message construction is controlled processing, and so is monitoring; self-corrections are hardly ever made without a touch of awareness. The speaker can *attend* [italics Levelt's] to his or her own internal speech. The limited-capacity resource in conceptualizing and monitoring is Working Memory (Levelt, 1989, p. 21).

By contrast, formulation and articulation must be largely automatic and beyond the scope of executive control. Otherwise, speech could not be produced as rapidly as it is, typically at the rate of two to three words, or 15 phonemes, per second. [15] The Formulator and Articulator have their own, separate buffers for interim products, and, unlike working memory, these buffers are not subject to attentional control.

Levelt created his model primarily to account for mature, fluent L1 speech. It has been modified and expanded in important ways by deBot (1992) and Temple (1997) to better describe bilingual speakers, including L2 learners. DeBot's modifications deal with the make-up of the main components while Temple discusses differences in storage and processing between fluent native speakers and less-advanced non-native speakers.

Based partly on research into code-switching, deBot theorizes that bilinguals have two Formulators, and that both produce phonetic plans for speech, although only one such plan proceeds to the Articulator. However, bilinguals possess one common mental lexicon, from which items for both languages are drawn. This arrangement allows non-balanced bilinguals to anticipate missing lexical items in their weaker language while they are forming preverbal messages, and then avoid or reformulate their intended message. DeBot's amendment provides a direct route for feedback from the Formulator to the Conceptualizer, a route that is absent in

Levelt's model. Of course, L2 speakers do not always make efficient use of this feedback mechanism, and the result is slow speech and disfluency. [16]

The relationship between working memory and L2 proficiency is addressed by Temple (1997). She theorizes that in fact working memory may be even more important in beginning- and intermediate-level L2 speech than in advanced speech. Essentially, Temple extends the scope of working memory into the formulation stage of L2 production, where subprocesses are usually less automatized than in native speech. Automatic retrieval of information from the lexicon is not always possible, and speakers may have to rely on explicit knowledge of L2 grammatical rules if these rules have not yet been proceduralized. Intermediate products that would normally be stored in the syntactic buffer or articulatory buffer, which are beyond attentional control, are instead routed to the limited working memory store. As a result, formulation proceeds slowly and serially rather than quickly and in parallel fashion. This shows up overtly in slower speech rate and in more hesitations, particularly mid-clause hesitations (Temple, 1997, pp. 86-87). [17]

If Temple is correct, what distinguishes fluent from non-fluent speakers is chiefly the degree to which they rely on working memory during routine formulation of messages. Becoming more fluent entails reducing this dependency on the working memory store. However, there is no obvious way to avoid this early stage of reliance. From this perspective, working memory provides a bridge to fluency, a bridge that becomes redundant once it is crossed (assuming the learner does not regress). Or, to use another metaphor, working memory is the midwife to fluency, an aide that provides a critical service and then recedes into the background once a certain, as yet unspecified, level of fluency is achieved.

2.1.3.6 Summary of Working Memory Research. Baddeley and Hitch (1974) introduced the concept of working memory as a theoretical successor to short-term memory, and their model provided a springboard for a great deal of psycholinguistic research. They depicted working memory as a limited-capacity system composed of two slave systems, the phonological loop and the visuo-spatial sketchpad, with the former being especially important in language use. A third component, the central executive, was given the role of coordinating attentional control and suppressing irrelevant information during cognitive processing. Some critics of the Baddeley-Hitch model maintain that it fails to account for a syntactic level of linguistic processing (Caplan and Waters, 1996). Other critics, such as Ericsson (Ericsson 1996; Ericsson & Kintsch, 1995; Ericsson & Delaney, 1999) and Schneider (1999) view working memory as more closely related to, even as part of, long-term memory. Baddeley (2000) revised his model to include an "episodic buffer," partly to help explain how working memory worked with long-term memory stores.

Since Daneman and Carpenter's (1980) groundbreaking study, working memory and its effects have been a prime focus of psycholinguistic research. Much of this research concerns individual differences in working memory capacity and how they may relate to differences in language learning outcomes. Several studies, including Daneman and Carpenter (1980), King and Just (1991) and Miyake, Just and Carpenter (1994), indicate that high working memory capacity gives some individuals advantages in certain aspects of reading, including connecting pronouns and referents and resolving lexical ambiguities. Individual variations in one aspect of working memory, phonological short-term memory (PSTM), has been found to correlate significantly with language learning success, especially in vocabulary learning (Service, 1992; Gathercole & Baddeley, 1990; Gathercole, Papagno & Vallar, 1988) and perhaps also in the acquisition of morphosyntax (Ellis & Sinclair, 1996).

Research on working memory and speech has been less extensive but some promising inroads have been opened. Daneman and Green (1986) found that L1 oral fluency correlated strongly with scores on the Speaking Span Test- but not on the Reading Span Test of working memory capacity. This suggested that working memory is at least partially dependent on the linguistic mode in which it is expressed. Fortkamp's (1999) study of ESL learners in Brazil suggested further that working memory capacity is language-dependent and that, at least for the advanced learners in this study, L2 (English) fluency was not related to working memory measured in L1 (Portuguese). This directly contradicts Osaka and Osaka (1992), whose study of Japanese ESL learners indicated that working memory was a language-independent faculty. This question of mode and language independence goes to the heart of theoretical discussions about the nature of working memory and its relationship to long-term memory.

How working memory may be employed in speech is an important theoretical topic. Levelt's (1989) influential model of L1 speech production assigns important responsibilities to working memory. These include storage of intermediate-stage products and monitoring of speech, both internal and external. Temple (1997) argues that for novice L2 learners, working memory may play an even more crucial role. If she is correct, one could expect to find an especially close relationship between working memory capacity and L2 output in less proficient learners.

2.2 ORAL FLUENCY

2.2.1 Perspectives on Fluency

Defining and measuring “fluency” are undeniably complex tasks. Among lay people and language specialists alike, the term is used in many ways, including some which raise more questions than they answer. For example, Harris (1969, p. 81) defines fluency as “the ease and speed and flow of speech.” But what constitutes “ease” or “flow of speech”? And how fast must speech be to be considered fluent? Also, do context and speech task matter when thinking about fluency? Presumably, they do: responding to an unexpected and complicated question taxes a speaker’s resources more than reciting a well-known children’s tale. Another point, and one that is very relevant to L2 speech, centers around the issue of formal accuracy. While basic competence is assumed in mature L1 speech, in L2 production there are often questions of comprehensibility and distractions caused by errors of pronunciation, word choice or syntax. How much these affect fluency, or perceptions of fluency, is hard to say with certainty.

A closer examination of the concept of oral fluency will help illuminate, if not definitively answer, these questions. In the following pages, diverse approaches to fluency will be outlined, beginning with some pedagogical evaluative tools. Several qualitative and theoretical perspectives will be presented, beginning with Fillmore’s (1979/2000) influential analysis. Communicative teaching approaches to fluency, notably those of Krashen (1981, 1982; Krashen &Terrell, 1983) and Brumfit (1984/2000), are also discussed, as are communicative strategies. The quantitative approaches to L2 fluency, including studies by Lennon (1990) and Riggenbach (1991), will be summarized. The focus will then shift to the psycholinguistic processes involved in developing L2 fluency, as envisioned by Schmidt (1992) and Segalowitz

(2000), among others. This review will conclude with a recapitulation and expansion of how working memory fits into the development of L2 oral fluency.

2.2.2 Pedagogical approaches to fluency

Language educators differ in whether and how they measure oral fluency in language learners. Some adopt a holistic approach wherein fluency is inextricably linked to factors such as pronunciation, vocabulary size and morphosyntactic accuracy. Others view fluency as a distinct component of speech, one that merits a separate rubric in test settings. Educators who take this approach generally link fluency to temporal factors such as speed of delivery and absence of hesitation, and sometimes to the cohesion of discourse as well.

Both the Modern Language Association and the American Council on the Teaching of Foreign Languages mention fluency in their guidelines for evaluating L2 speech. But neither group includes a separate scale for fluency. The MLA classifies "good" speakers as those who have mastered enough L2 vocabulary and grammar to speak "at normal speed with reasonably good pronunciation." "Superior" speakers can "speak fluently, approximating native speech in vocabulary, intonation and pronunciation," according to the MLA (Spolsky, 1995, pp. 187-88). As with Harris' definition quoted above, this classification leaves much to the subjective tastes of individual evaluators. The ACTFL's guidelines for rating L2 speech describe "superior" speakers as those who produce "(f)luent speech in which errors virtually never interfere with communication or disturb the native speaker" (Kenyon, 1995, p. 23). The inclusion of accuracy as a component- indeed *the* component- of fluency is worth noting. The reference to disturbing native speakers is also interesting, as it implies that even errors that do not hinder communication of meaning may have a negative effect on listeners' impressions of L2 fluency.

The U.S. Foreign Service Institute (FSI) treats fluency as a distinct component of L2 speech. The Foreign Service uses a 6-point scale that identifies fluency closely with smoothness of delivery and the absence of hesitation. For example, a rating of 1 indicates that speech is "so halting and fragmentary that conversation is virtually impossible." A rating of 3 is for speech that is "frequently hesitant and jerky," with many incomplete sentences. A top rating of 6 is reserved for speech that is "effortless and smooth as a native speaker's" (Hughes, 1989, p. 112).

The University of Reading has created a 4-point scale for rating the fluency of ESL learners. Like the FSI scale, it mentions hesitations prominently among its descriptors. But it also includes coherence, length of utterances and the ability to use "fillers" skillfully. The scale is reproduced in its entirety in Table 15 (from Weir, 1993, p. 44). The Reading system includes separate scales for pronunciation, vocabulary and sociolinguistic appropriateness.

Although this approach to fluency is more detailed than the holistic approach , it still leaves considerable leeway for individual raters to judge for themselves what constitutes "disjointed" or "hesitant" speech, or how fast speech should be at a certain level. More specific guidelines and elaborations are not provided. This is perhaps a bow to reality: in most learning situations it would be very time-consuming and costly to employ detailed temporal measures of fluency for larger numbers of students. However, that does not preclude the possibility that more specific standards might be achievable, at least for research purposes. More detailed quantitative approaches to fluency may help establish such standards.

Table 15: The University of Reading fluency rating scale

0	Utterances halting, fragmentary and incoherent
1	Utterances hesitant and often incomplete except in a few stock remarks and responses. Sentences are, for the most part, disjointed and restricted in length.
2	Signs of developing attempts at using cohesive devices, especially conjunctions. Utterances may be hesitant, but are gaining in coherence, speed and length.
3	Utterances, whilst occasionally hesitant, are characterized by an evenness and flow, hindered, very occasionally, by groping, rephrasing and circumlocutions. Inter-sentential connectors are used effectively as fillers.

2.2.3 Qualitative approaches to fluency

2.2.3.1 Fillmore's four dimensions of fluency. Fillmore's (1979/2000) identifies four basic characteristics of oral fluency. The first is "the ability to talk at length with few pauses, the ability to fill time with talk" (2000, p. 51). Fillmore mentions disk jockeys and sports announcers as examples of people who typically exhibit this sort of fluency. The second dimension is more qualitative in nature. It is “the ability to talk in coherent, reasoned, and ‘semantically dense’ sentences” (ibid., p. 51). People who show this kind of fluency have mastered the syntax and semantics of the language, and they are skilled at communicating their thoughts in a careful, compact manner. The third dimension is “the ability to have appropriate things to say in a wide range of contexts” (ibid., p. 51). This includes knowing how to speak in different social settings, including unexpected situations. Finally, there is “the ability some people have to be creative and imaginative in their language use” (ibid., p. 51), including skill at making puns and jokes, producing novel expressions, etc. Fillmore’s use of the qualifier “some” here makes clear that this last aspect of fluency is one that is not universal, perhaps not common.

Fillmore’s discussion is also notable for its balanced and contextual view of fluency. It includes both qualitative characteristics (the first and, partly, the third dimensions) as well as largely quantitative aspects (the second and fourth dimensions). The implication is that fluency ideally exhibits a tension between various qualities, none of which can be allowed to dominate speech production. This is especially true with regard to the first two dimensions, “filling time with talk” and “semantic density.” It is certainly possible for a person to talk rapidly and smoothly for some time without adding much informational content to a discourse. At the opposite extreme, a person may speak compactly and without extra verbiage, yet still speak so slowly that he annoys or disengages listeners. Neither of these speakers would display truly fluent speech, according to Fillmore’s understanding. Likewise, having a large inventory of formulaic expressions at hand may help a speaker fill time with speech. But listeners often interpret overuse of fixed expressions as a sign that the speaker lacks creativity and skill at adjusting his output to changing contexts (Fillmore, 2000, p. 53).

Throughout his article Fillmore acknowledges, the importance of considering social context when assessing fluency. Although all mature native speakers probably know how to perform common speech acts, he says, they may differ greatly in how well they’ve mastered the indirect communication that often accompany these acts. And individuals certainly differ in their narrative ability, as some people have mastered better than others typical storytelling schemata-theme development, transitions, resolutions, etc. (Fillmore, 2000, p. 56). Fillmore also distinguishes between fluency when producing monologues and when taking part in dialogues. In the former situation, a large vocabulary may help enhance a speaker’s fluency by allowing him to use unusual words. But vocabulary size is probably less important in dialogues, where an interlocutor's ability to follow what is being said is crucial, according to Fillmore (2000, p. 52). Although Fillmore doesn't say so directly, dialogue and group discussion formats presumably place more value on the creativity and flexibility than monologues do, simply because the speaker cannot control the topic or direction of discourse as completely.

Fillmore's article is also noteworthy for the importance it assigns to formulaic expressions in maintaining oral fluency. In Fillmore's view, these fixed or semi-fixed utterances are crucial in achieving fluency, particularly the first-mentioned, "filling time" type of fluency. He cites several examples of formulaic expressions in English (Table 16).

Table 16: Some formulaic expressions (Fillmore, 2000)

a.	**It's my turn.**
b.	**All in favor say "aye"!**
c.	**Speak of the devil.**
d.	**Let me be the first to congratulate you.**
e.	**Don't tell a soul.**
f.	**Anybody home?**
g.	**Plenty more where that came from.**
h.	**It takes one to know one.**
i.	**We'll hate each other in the morning.**

Fillmore notes that these expressions tend to be memorized as wholes and associated with particular communicative situations. They are not creatively "generated" in the Chomskyan (e.g., Chomsky, 1965) sense from an abstract knowledge of the language [18], and they could not be predicted by someone who "merely knew the vocabulary and grammar of the language" (Fillmore, 2000, p. 50). This is one reason why Fillmore doubts that the competence-performance distinction, which is central to generative theory, is tenable in terms of oral fluency.

Like Levelt, Fillmore is dealing exclusively with speech production by mature native speakers. However, his analysis has implications for L2 instruction and research. One implication is that L2 learners' fluency should be assessed in a variety of different formats and contexts, including perhaps semi-planned as well as unplanned production. Only such a diverse assessment scheme is likely to give a well-rounded picture of a student's capabilities. Another implication is that formulaic expressions, collocations and other "chunks" of prefabricated speech are at least as important as individual words in building fluency. Pawley and Syder (2000, p. 179) estimate that mature English speakers may have at their disposal hundreds of thousands of these conventionalized multi-word units. For the non-fluent L2 speaker, these chunks may provide what Dechert (1984, p. 184) calls "islands of reliability," points of respite in the taxing process of planning and executing speech under time pressure. Later researchers, including Oppenheim (2000) and Ejzenberg (2000) stressed the role that recurring strings of fixed or semi-fixed speech play in increasing L2 fluency.

Fillmore's description of fluency is less relevant to L2 speech in other ways, both in what it says and in what it doesn't. Fillmore's fourth dimension of fluency, the ability to use language creatively, make puns and jokes, etc., is surely a desirable trait, especially in societies that highly value verbal skill. But as many L1 speakers do not possess these talents, they are best considered luxuries in an L2 learning context. Likewise, the ability to spontaneously produce "semantically dense" sentences is probably uncharacteristic of most speakers under most circumstances. Indeed, a basic characteristic of speech, as opposed to written language, is its "looseness": writing tends to be more compact and tightly structured, while speech contains many more repetitions, rephrasings and fillers (Crystal, 1997, p. 181). Tellingly, Fillmore (2000, p. 51) mentions Noam Chomsky and William F. Buckley, two people often regarded as unusually good public speakers, as models of this second type of fluency. The point is that, however one decides to evaluate L2 fluency, the standard should not be set at a level higher than many or most native speakers attain.

Fillmore's description of fluency is also interesting for what it does not include. He does not mention speed of delivery explicitly, though this is probably subsumed under his first, temporal kind of fluency. More importantly, the question of formal accuracy is largely absent. This is understandable in a discussion of native language speech, where formal proficiency is usually assumed. But it leaves open the issue of how accuracy and fluency are related in L2 speech. It is at least theoretically possible for a non-native speaker to exhibit all of Fillmore's four dimensions of fluency and still be incomprehensible to native listeners because of severe pronunciation and/or syntactic errors. And even when errors are not so egregious as to hinder communication, they may disturb listeners and lower their perceptions of the speaker's fluency (Hammerly, 1991). If so, the problem is to determine, at least roughly, what level of accuracy a non-native speakers must achieve in order to be considered fluent by native listeners. Or, since there be no identifiable "threshold level" of accuracy, the researcher must attempt to determine how much influence (in)accuracies have on perceptions of L2 fluency.

2.2.3.2 Communicative teaching and L2 fluency. Fillmore's approach to fluency is largely communicative in nature, emphasizing as it does creativity, coherence and social appropriateness. Other observers, inspired by the communicative approaches to language teaching that took root in the 1970s and afterward, adopted a similar attitude toward L2 oral fluency. For example, Beardsmore (1974, p. 323) states that for ESL learners fluency involves the ability to produce "sustained oral production implying a certain communicative competence, as well as the unstilted spontaneous use of English 'conversational lubricants.'" These "lubricants" are the sort of semantically empty or redundant bits of language that Fillmore's ideal speaker might use to "fill time." For Sajavaara (1987, p. 62), L2 fluency consists of two factors, "linguistic acceptability and smooth continuity of speech." Sajavaara makes clear that the former means primarily "communicative acceptability," that is, whether a speaker's utterance fit well with his or her listener's expectations in a particular social context.

Two other advocates of communicative language teaching, Stephen Krashen and Christopher Brumfit, have written extensively about fluency and how it can be developed. In Krashen's Monitor Model and the Natural Approach of instruction (Krashen, 1981, 1982; Krashen & Terrell, 1983, *inter alia*), there is a strict division between consciously acquired, or "learned" language and unconsciously "acquired" language. Only the latter is available for automatic, fluent L2 use.

> Our fluency in production is thus hypothesized to come from what we have "picked up," what we have acquired, in natural communicative situations. Our "formal knowledge" of a second language, the rules we learned in class and from texts, is not responsible for fluency, but only has the function of checking and making repairs on the output of the acquired system (Krashen & Terrell, 1983, p. 30).

Krashen and Terrell appear to be using the term “fluency” in the sense of “rapid and smooth.” The additional sense of “exhibiting native-like accuracy” is evidently not included, as indicated by the reference to a separate checking and repairing faculty, elsewhere called the “Monitor.” Of course, a basic competence in the formal properties of the target language is necessary to the development of fluency. But in Krashen’s system, this emerges naturally as the learner’s innate linguistic abilities are exposed to sufficient amounts of comprehensible input. In fact, Krashen and Terrell advocate a “silent period” in the early stages of L2 learning, when the student builds up a basic competence without the pressure of being forced to use the target language productively.

In the Monitor Model, individual differences in L2 learning arise primarily from affective differences, which influence how much students use the Monitor. “Over-users” fail to produce fluent speech because they are too concerned with correctness and applying consciously learned rules. They speak hesitantly and often correct their own speech in mid-stream. There are also “under-users,” who ignore error correction and consciously learned rules. But these people are closer to, and more likely to become, “optimal Monitor users,” those skilled students who “will not be excessively concerned with applying conscious rules to performance” (Krashen & Terrell, 1983, p. 45). Given enough input, these students may eventually attain a global kind of fluency, one that includes accuracy as well as speed and smoothness.

Brumfit (1984/2000) takes a pedagogical and learner-oriented approach to L2 fluency. He defines fluency as “natural language use, whether or not it results in native-speaker-like language comprehension or production” (2000, p. 68). He adds that fluency “can be seen as the maximally effective operation of the language system so far acquired by the student” (p. 69). This definition is remarkable for its focus on the student and his/her developing L2 system rather than on L2 norms or the effect on listeners. Clearly, the sort of “natural language use” Brumfit describes here would often appear unnatural to native speakers. It is “natural” in the sense of being produced in real-life or life-like communicative contexts, and not during inauthentic, teacher-imposed learning exercises. From this perspective, fluency can only be judged in relation to the student’s current level of L2 proficiency: some students with the same rough level of proficiency will make better use of what they know than others.

To be sure, Brumfit does not advocate a total emphasis on fluency in the classroom. His main point is to distinguish between fluency-building activities and accuracy-building activities, and to insist that the latter do not encroach on the former. He advises instructors (Brumfit, 2000, p. 69) to use correction sparingly if at all during fluency-building activities, and to allow learners "the freedom to create their own grammars" (ibid., p. 72), based on input from instructors and other sources. Brumfit argues that this will help students integrate consciously "learned" L2 forms into unconsciously "acquired" communicative systems that can be employed for fluent production.

Though Krashen and Brumfit present similar approaches to L2 acquisition, Brumfit's allows a larger role for cognitive factors such as working memory. Krashen's insistence on a strict divide between learning and acquisition, as well as his emphasis on affective influences, leaves little room for other factors in explaining individual differences in L2 fluency. If "learned" material never becomes "acquired" and therefore available for fluent use, one must turn to the workings of the learner's innate system (whether that system is called Universal Grammar or something else) for explanations. These workings are difficult to plumb with any specificity at present, and being universal in nature, they do not offer a promising source of explanations for individual differences. [19] Affective variables such as anxiety may influence the efficiency of working memory, probably negatively in most cases. But this places working memory in the role of a secondary factor rather than a primary reason for different learning outcomes.

By contrast, though he adopts Krashen's "learning" and "acquisition" terminology, Brumfit leaves open the possibility that "learned" forms can be integrated into an interlanguage system capable of producing fluent speech. In that case, differences in working memory capacity or efficiency might help explain why some learners access learned material better than others during speech production. In fact, Brumfit's comment about the "maximally effective operation of the language system so far acquired by the student" supports this line of speculation. In essence, the main question is this: between two learners with identical or near-identical L2 competence, why does one make better use of what he or she knows? Why can one access and articulate words, phrases and syntactic patterns more quickly and/or correctly than the other? Assuming for the moment that affective differences cannot entirely account for this disparity, two possibilities suggest themselves: individual variation in cognitive factors such as working memory and variation in use of communicative strategies. The former was the focus of the main experiment described below; the latter will be taken up now.

2.2.3.3 Communicative strategies and fluency. A communicative strategy is, in Corder's (1983, p. 16) words, "a systematic technique employed by a speaker to express his meaning when faced with some difficulty." Among the techniques are paraphrase, circumlocution, message abandonment, avoidance of difficult forms or topics and appeals to interlocutors for assistance (see Tarone et al., 1983, for an inventory of strategies). [20] Some of these techniques can serve as fluency-enhancing strategies in that they allow a speaker to "hold the floor" and advance the discourse despite formal difficulties. A good, concise paraphrase can do just that. But some techniques are more problematic in terms of enhancing an interlocutor's perceptions of fluency. Avoidance and simplification- what Faerch and Kasper (1983) call "formal reduction strategies"- may enhance fluency but only in a strictly temporal sense. If they are overused the result will likely be what Lennon (2000, p. 28) terms "false fluency," a style of discourse that is formally and semantically impoverished or contextually inappropriate.

A different type of strategic thinking may help learners maintain fluency, or more precisely, the appearance of fluency. Faerch and Kasper (1983b) propose that awareness of "performance features" can help in this regard. These features include pauses, drawls and repetitions, all of which can be utilized to gain time during the process of speech production. The timing of pauses is especially important because pauses can indicate either normal (i.e., native-like) moments of planning in an ongoing plan-and-execute cycle (Pawley & Syder, 2000; Clark & Clark,1976; Goldman-Eisler, 1968), or else a lack of linguistic competence. Faerch and Kasper (1983b, p. 235) suggest that teaching learners (at least advanced learners) how to produce native-like filled pauses and other "floor-holding gambits" can improve their fluency. [21]

It may seem ironic, even counterproductive, to focus on pauses, drawls and repetitions as a means of increasing oral fluency. After all, these phenomena are often signs of non-fluent speech. But it is important to recognize the difference between actual fluency in the purely temporal sense and listeners' subjective perceptions of fluency. "Perceived fluency," says Lennon (2000, p. 27),

> …refers to the impression the listener has that the psycholinguistic processes of speech planning and speech production are functioning more or less easily and effortlessly. It may not exactly mirror speaker knowledge and processing skill... (T)he speaker may maintain the illusion of effortless processing, such that the listener is largely aware of processing problems.

Most L2 learners will never be able to match native speakers in terms of sheer speed of delivery, or even come very close. But skillful uses of pauses and conversational devices may allow the learner to cover up at least some deficiencies, and simulate a native-like pattern of speech planning and execution.

Ejzenberg's (2000) study of ESL learners in Brazil helps to illustrate how this kind of simulation may be achieved. Ejzenberg recorded and analyzed the L2 speech of 46 learners at various levels of English proficiency. She found that fluent speakers tended to be better than others at employing performance factors to maintain fluency. For example, while the speech of both fluent and less-fluent speakers exhibited numerous redundancies, the fluent speakers were able to use them creatively, often adding slight variations. Ejzenberg provides extracts from a low-fluency speaker and a high-fluency speaker, both of whom were discussing crime in Sao Paulo (Table 17).

For Robinson, the low-fluency speaker, repetitions often appeared within clauses, and they tended to convey the same semantic content as the original utterance. These were in effect "restarts," and the effect was one of "debilitating hesitation rather than strategic competence," according to Ejzenberg (2000, p. 302). By contrast, more of Roberto's redundancies appeared between clauses and they often added at least a little content to what went immediately before. From the listener's perspective, the effect was to "push the talk forward," advancing the discourse rather than "holding the talk back." Another, internal effect was to give Roberto time to plan and deliver his next utterance (ibid., pp. 301-302).

An important practical question is whether students can be taught how to do what Roberto apparently does naturally and use "disfluencies" to enhance their L2 fluency. In other words, can fluency be taught as a strategic skill in itself, distinct from linguistic competence? Doing so would require, firstly, that L2 teachers are trained in the characteristics of fluent speech that go beyond issues of accuracy. Secondly, it would require that students, at least at the advanced level, are open to mastering these additional aspects of oral communication. However, the rewards may well be worth the effort, as Guillot (1999, p. 61) points out. Concentrating on fluency can make learners more aware of the performance aspects of speech, help them make more discriminating use of exposure to the L2 within and outside the classroom and enable them to identify fluency-enhancing features and strategies most useful to them. It does not mean attempting to instruct learners in sophisticated conversational analysis. Rather, the aim is to "give students the initiative and confidence to reflect, at their level, on the ins-and-outs of verbal behavior," and allow them to improve their own L2 behavior (ibid., p. 64). To this may be added that there is a simple question of fairness: if students are to be assessed and graded on oral fluency, as they often are, both teachers and students should know in advance what fluent speech sounds like and how it can be learned. [22]

Table 17: Excerpts from monologues by Brazilian ESL learners (Ejzenberg, 2000) (bold print is Ejzenberg's)

Spoken by Robinson (low-fluency speaker)	
1.	**But in Sao Paulo there are happening very** fun...fun cases **in ...**fun
2.	cases**. For instance, lately there are happened uh...there are many** thieves
3.	that are stoling**...**that are stoling uh "tennis" shoes..."tennis" **uh...**
4.	expensive tennis **of the ...the people that are walking in the...in the street. And**
5.	sometimes they kill the...the persons **that not...that don't want to...to give the...the**
6.	**"tennis." I think you...**you have to...**to** take care...**to** take care **with**
7.	**your clothes and things...**
Spoken by Roberto (high-fluency speaker)	
1.	**it's a** big...big **city, and this city is uh aggressive in a certain way, so when you come**
2.	**in be careful** the moment you leave **the airport or** the moment you leave **your plane**
3.	**Well, Dick. Eh: Just to tell another story about the...how the**
4.	**city can be violent you know** that young people here ar...they're having their
5.	tennis shoes robbed**. They're walking along the street with** those Nikes **and** those
6.	other tennis shoes **that...you know that...made in USA...quite** expensive/ tennis/
7.	shoes**, and** this tennis**...they...they...**they are robbed

2.2.4 Quantitative studies of L2 fluency

Over the past 15 years several studies have been conducted on the temporal and quantitative measurement of L2 oral fluency. Lennon's (1990) experiment was perhaps the most exhaustive in terms of measurement, as it examined a dozen different factors related to speed, pausing profile, self-corrections and other types of disfluency markers. Lennon's main purpose was not to identify the factors that best identify fluency but rather the factors that improve most during a study abroad period. Nonetheless, his study provided a wealth of interesting data, as well as a point of reference for subsequent research.

Lennon's subjects were four native German-speaking women who were spending a half-year in England to improve their (already advanced) English. Lennon recorded the women near the beginning of their stay, and then again near the end. In both cases, he elicited L2 speech by asking his subjects to narrate a six-frame cartoon-like picture story (they had apparently forgotten the details of the picture story by the second recording session). Lennon had a panel of nine EFL instructors, all native English speakers, listen to both sets of recordings and rate them holistically for fluency. Lennon also analyzed the recordings in terms of 12 temporal factors (Table 18). These factors cover speed of delivery, as well as various types of potential disfluency markers such as self-corrections and repetitions and the number and location of pauses (both filled and unfilled).

In factor 1, the "unpruned" words means all the words produced by the subjects. "Pruned" words in factor 2 refers to all words minus self-corrections, repetitions (excluding repetitions used for rhetorical purposes) and "asides" unrelated to the narrative task (i.e., "*sorry, I didn't know that*," which was addressed to the experimenter). A T-unit (factors 3, 4, 5, 10, 11, 12) is a linguistic measure devised by Hunt (1970), and it is a main clause with all its subordinate clauses and non-clausal units. Coordinated clauses such as "*He went to the store and he bought a shirt*" count as two T-units (though only one if the second pronoun is deleted). Filled pauses are nonwords like "er," "erm" or "mm." For unfilled pauses, Lennon established a cut-off point of 0.2 seconds; periods of silence shorter than that were excluded from his count.

Table 18: Temporal qualities of ESL speech studied by Lennon (1990)

1.	**Words per minute (unpruned)**
2.	**Words per minute (pruned)**
3.	**Repetitions per T-unit**
4.	**Self-corrections per T-unit**
5.	**Filled pauses per T-unit**
6.	**Percentage of repeated and self-corrected words**
7.	**Unfilled pause time as percentage of total delivery time**
8.	**Filled pause time as percentage of total delivery time**
9.	**Mean length of speech "runs" between pauses**
10.	**Percentage of T-units followed by pauses (filled and unfilled)**
11.	**Percentage of pause time at T-unit boundaries (filled and unfilled)**
12.	**Mean pause time at T-unit boundaries (filled and unfilled)**

Although Lennon's subjects tended to improve their L2 speech in terms of most of these variables, they achieved statistically significant improvements in only three: pruned speech rate (factor 2), filled pauses per T-unit (factor 5) and pauses between T-units (factor 10). (For the first of these three factors, improvement meant more words per minute; for the latter two factors, smaller numbers (fewer pauses) were taken as signs of improvement.) Lennon (1990, p. 413) proposes that these three factors be considered, at least provisionally, as "core" measures of fluency assessment. Interestingly, the number of self-corrections did not seem to be related to increased fluency. In fact, three of the four subjects used more self-corrections at the end of their stay than at the beginning. Lennon suggests that self-corrections are a poor indicator of fluency, as they may be interpreted as signs of improved monitoring and reformulating abilities.

Broadly speaking then, it seems that speed and smoothness, or lack of pauses, were the best indicators of fluency, or at least fluency improvement. This has to be taken in context, however. Here, speed of speech is "pruned," so this measure takes into account some disfluency measures which by themselves did not appear to be significant. And in terms of pauses, the record was mixed, with some measures proving significant changes while others didn't. Also, the small number of subjects precludes overly broad generalizations. It is worth noting that one of Lennon's subjects, named Dorothea, improved her fluency but in a somewhat different way than the other three. Among other things, Dorothea showed insignificant gains in speech rate (factors 1 and 2) and she had a large and significant decrease (54%) in redundancies (factor 6). It might be, Lennon speculated (p. 411), that individual patterns of fluency improvement exist, reflecting perhaps different emphases or strategies on the part of learners, or different personality traits.

Lennon's study is useful for several reasons. Firstly, Lennon cast a wide net, looking at a dozen factors that could be associated with fluency, at least in the narrow, temporal sense of the word. In practical terms, this meant limiting the number of subjects and the number of speech samples to be analyzed. But this is understandable and probably unavoidable in an exploratory study. Also, he used a narrative monologue to elicit speech from his subjects rather than a dialogue or discussion format. This seems well-motivated in a study of oral fluency as it avoids confusing subjects' receptive abilities with their purely productive skills. Of course, in real-life discourse the two will be closely intertwined, and interactional factors such as knowledge of turn-taking norms in the L2 will come into play as well. But in an experimental setting where productive fluency is the focus, Lennon was right to choose the picture story elicitation device that he chose. This technique had the additional advantage of forcing subjects to describe the same stimuli and produce similar narratives. A freer interview or conversational format would have allowed subjects to switch topics and styles of discourse, and would almost certainly have produced less uniform speech samples, samples that would have been difficult to compare to each other.

Finally, Lennon decided to give his panel of nine teacher-judges a formal definition of fluency before they listened to the subjects' speech samples. In this definition, fluency has two components: a temporal component that includes "speed of delivery, for example," and "a degree of freedom from various disfluency markers such as repetitions, self-corrections, filled pauses, and the like" (p. 403). This definition encapsulates a narrow, temporal kind of fluency, based on speed and smoothness, without reference accuracy, and it is certainly a plausible definition. Another feasible approach would have been to let the judges themselves decide how much weight, if any, to give to various factors when considering what constitutes L2 fluency. This would have added at least one more factor- and a complex one, at that- to an already extensive list of factors. But it should not be ruled out in future studies of fluency. In any case, Lennon's list of factors. But it should not be ruled out in future studies of fluency. In any case, Lennon's study focused most closely on changes in potential fluency markers, and the judges played a secondary role.

Riggenbach (1991) conducted a similar study, using as her subjects six Chinese-speaking ESL learners. Three of the learners had been holistically characterized as "fluent" by a panel of ESL instructors, and the other three had been labeled "nonfluent." Like Lennon, Riggenbach micro-analyzed speech samples produced by her subjects in terms of quantifiable temporal factors. Her aim was not to see which factors changed over time, but rather to determine which factors were associated with speakers judged to be either fluent or nonfluent. She found only one factor that was unambiguously identified with fluency: the three fluent speakers produced significantly more unfilled pauses than the nonfluent speakers. The evidence on speech rate was interesting but less clear-cut: speed of delivery was a significant difference under one statistical analysis but not another.

Interestingly, one of Riggenbach's six "nonfluent" subjects was almost identical to the three "fluent" speakers in her pause profile and speaking rate. Riggenbach speculates that the ESL teacher-judges had labeled this learner as "nonfluent" because she made many grammatical errors. As in Lennon (1990), these sort of errors were not among the array of factors under study. But this individual's case indicates that they may nonetheless play an important role in listeners' perceptions of L2 fluency.

Two other quantitative studies of fluency merit attention, Freed (1995/2000) and Kinkade (1995). Freed's subjects were 30 students from the United States (3 were non-native but highly accomplished English speakers) who were studying French, half in France and half in the U.S. Freed had six native French speakers listen to the students' recorded speech samples and rate them holistically for fluency. These samples lasted three minutes and were excerpted from oral interviews conducted entirely in French. Freed provided her judges with a list of eight potential factors (Table 19), and also allowed the judges to add others.

Five of the six judges selected speed (factor 2) and factor 7, which seems to be a broad, catch-all option. Four chose better grammar (factor 6), vocabulary use (factor 4) and fewer pauses or hesitancies (factor 3). Three said they had focused on accent (factor 5).

Table 19: Possible fluency-related factors identified by Freed (2000)

1.	**Their speech was more idiomatic.**
2.	**The rate of their speech was faster; they spoke more quickly- closer to native speed.**
3.	**There were fewer pauses in their speech, fewer hesitancies.**
4.	**Their vocabulary was better.**
5.	**Their accent was better.**
6.	**Their grammar was better.**
7.	**They sounded better in general: their speech was smoother, there were fewer false starts (i.e., they stopped and started or corrected themselves less).**
8.	**There was better interaction between the student and the interviewer.**
9.	**Other.**

It should be noted that these are the factors that the judges subjectively identified as most important in their fluency judgments. It is possible that these factors did not coincide with the actual occurrence of speed, smoothness and accuracy measures in the subject's L2 speech. And there may have been other factors that the judges did not identify which may nonetheless have correlated well with their evaluations of fluency. Of course, the judges' own accounts must be given some weight, but ideally this sort of qualitative data-gathering would be combined with the sort of quantitative microanalysis that Lennon and Riggenbach used.

Kinkade (1995) recorded interviews with 17 Chinese-speaking ESL learners in her study of L2 fluency. She used three native-speaking English teachers as judges [23] and, like Lennon (1990), she correlated their holistic fluency ratings with various factors. Also like Lennon, Kinkade provided her judges with a definition of fluency in advance. She informed them in writing that fluency refers to "smoothness, ease and readiness" of speech, and she specifically directed them not to take account of grammatical or lexical accuracy. However, in her later analysis, Kinkade did look at the subjects' morphosyntactic error rate. In doing so, she acknowledged that formal accuracy might have affected her judges' fluency ratings despite her instructions to ignore it.

Kinkade identified three factors that correlated most strongly with holistic impressions of fluency: mean length of unfilled pauses, speech rate and mean length of T-units. The latter is a measure of syntactic complexity, a factor that previous studies of L2 fluency did not include. Together, these three factors are measures of smoothness, speed and complexity, respectively. As a group, these factors accounted for 86 percent of the judges' fluency judgments.

Table 20: Summary of four quantitative studies of L2 oral fluency

	Non-native speakers (L1, number)	**Native speaker judges (L2, number)**	**Type of speech samples**	**Most significant fluency factors**
Lennon (1990)	German, 4	English, 9	Picture story narrations	Speech rate Unfilled pauses Intra-clausal pauses
Riggenbach (1991)	Chinese, 6	English, 12	Interviews	Unfilled pauses Speech rate
Freed (1995)	English, 30	French, 6	Interviews	Speech rate Smoothness Grammar Vocabulary Pauses
Kinkade (1995)	Chinese, 17	English, 3	Interviews	Length of pauses Speech rate Syntactic complexity

Results of the four quantitative studies discussed above are presented in Table 20. These studies approach the notion of fluency differently in terms of aims and methodology. But some points of convergence are discernible in terms of their conclusions. Speed of delivery, whether measured in words per minute (Lennon and Riggenbach) or syllables per minute (Kinkade) was consistently found to be important. Smoothness was also found to be a significant element of fluency, though it was quantified in different ways (Riggenbach's study emphasized the number of unfilled pauses, while Kinkade found that the length of unfilled pauses was important). Filled pauses, repetitions and self-corrections- all identified by researchers as potential disfluency markers- did not turn out to be significant in any of these studies. In fact, as mentioned above, the number of self-corrections may actually increase along with fluency, according to Lennon. Formal accuracy was not an important factor in these four studies but it should not be dismissed out of hand. Both Lennon and Kinkade had their judges exclude accuracy from consideration, and the latter did so very explicitly. (Incidentally, Kinkade's decision to measure syntactic complexity turned out to be well-motivated, based on her results.) Four of Freed's judges identified better grammar or vocabulary use as factors in their fluency ratings, and the case of Riggenbach's fluent but inaccurate subject points to the possibility of a role for formal accuracy. It seems, then, that there are very good reasons to believe that formal accuracy does play a role in listeners' perceptions of fluency, that it is inextricably linked to temporal qualities such as speed and smoothness.

These studies differed widely in the number of L2 subjects and in the number of L1 fluency judges used. Understandably, Lennon and Riggenbach's studies, which produced close-up, data-rich microanalyses of L2 speech, were limited to a small number of speaker-subjects. Freed and Kinkade's studies had more subjects but less intensive data analysis. Riggenbach employed a sizeable number of L1 listener-judges (12), although more would probably be needed to ensure statistical reliability.

Finally, it should be noted that three of the four studies elicited speech samples by means of free-form, or semi-free-form, interviews. Lennon was the exception; he used a picture story narrative technique. This seems to be a good decision for reasons already cited: it keeps the topic matter and hence the relative difficulty of the speech sample constant, and it eliminates factors such as comprehension ability and turn-taking skill. Again, these are undoubtedly important in many everyday settings as they influence overall conversational ability. But they are probably less significant, and possibly distracting, when the topic of investigation is productive abilities.

2.2.5 Psycholinguistic perspectives: Fluency as automatization

The growth of L2 fluency can be viewed within the larger context of skill development, whether the skill in question is playing chess, playing a musical instrument or piloting a jet plane. This development is often described as a process of automatization, following Shiffrin and Schneider's (1977) distinction between controlled and automatic processing. Controlled processes are more typical of early learning, and they tend to be slow, effortful and serial in nature: one process in a sequence must be completed before the next can begin. They demand a relatively high degree of attention and are limited by short-term memory capacity. Automatic processes, by contrast, are fast, effortless, not limited by short-term memory limits and not under voluntary control. They demand little attention and can proceed in parallel, with two or more processes taking place simultaneously (Schmidt, 1992, p. 360). Levelt (1989, p. 2) makes clear that fluent speech demands the latter, that is, "automatic, reflex-like" processes which work simultaneously.

But how do L2 learners progress from controlled to automatic processing, from slow, deliberate speech to fluency? And how might working memory be involved? The research provides some possible answers. These involve proceduralization of rule-based L2 knowledge, suppression of unwanted lexical items and retrieval of lexicalized "chunks" of speech. Two general approaches to automatization will be briefly outlined: the "resource-based" view exemplified by Anderson's (1983) ACT* model and the "memory-based" perspective of Logan's (1988) "instance theory."

As Schmidt (1992, p. 361) notes, the shift from controlled to automatic processing has been presented, at least in part, as a process of strengthening connections between memory nodes through repetition and practice. Such strengthening creates sequences of nodes, which are automatically activated as wholes when the input demands it. The result is faster performance, and reduced attentional needs for those subprocesses that have been become activated as wholes. This view of automatization as associative strengthening can probably account for many learning situations that involve easily routinized skills. It produces what Segalowitz (2000, p. 202) calls "simple speedup" of underlying performance mechanisms.

But many theorists believe that qualitative changes in the nature of knowledge representation are necessary for the automatization of complex skills such as speech. Anderson's

(1983) ACT* model of learning (Fig. 2-4) presents this change as a movement from declarative knowledge to procedural knowledge. Declarative knowledge is essentially “knowing about”- for instance, knowing that in Spanish the conditional form of verbs is usually formed by adding certain suffixes (*-ía, -ías, -íamos*, etc.) to the infinitive. Procedural knowledge is “know-how,” the ability to apply rules to carry out functions- in this case, to produce specific conditional verbs when needed. This procedural knowledge is automatized in that it is accessed quickly, unconsciously and often simultaneously with other knowledge. A well-developed store of proceduralized knowledge is obviously important, even necessary, to automatized, fluent speech. In all likelihood, this is especially true with regard to complex morphology and syntax, where conscious application of abstract rules would be unduly time-consuming. DeKeyser’s (1997) study of learners acquiring the morphosyntax of a miniature artificial language indicates that the shift from declarative to procedural knowledge occurs early in the learning process, and is marked by relatively large initial improvements in performance.

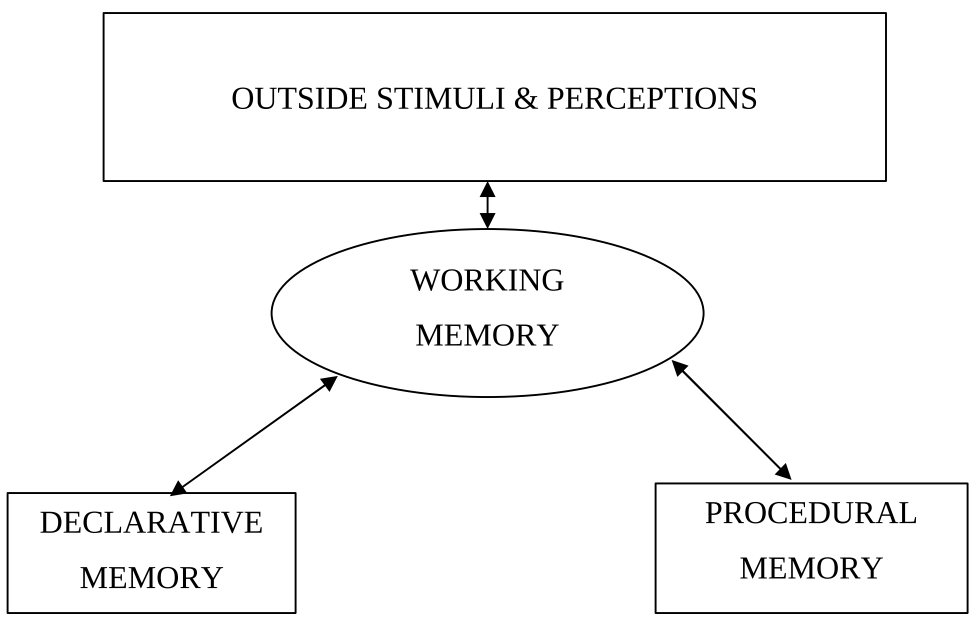

Figure 4: Anderson’s (1983) ACT* model of skill acquisition

As shown in Fig. 4, in the ACT* model working memory has a central function in coordinating declarative memory and procedural memory (“production memory” in the diagram) with perceptions from the outside world. In the context of speech production, this represents the integration of controlled and automatic linguistic processes with communicative demands originating in the ongoing discourse record [24] or task at hand (i.e., describing a picture or telling a story). The balance between controlled and automatic processes will change with the speaker's skill level and the complexity of the communicative task. But it seems likely that all but the briefest, most highly routinized utterances will require some controlled output, and hence some short-term attention-focusing mechanism. Striking the ideal balance between control and automatization falls to the central executive component of working memory.

It will be recalled that another duty of the central component is the suppression of irrelevant or distracting information. This function is germane to the process of lexical access, which Lennon (2000) suggests may be the single most important factor in oral fluency. Experimental evidence suggests that novice L2 learners face lexical interference both from their L1 and from their immature L2 lexicons. Novices tend to associate L2 words with their referents indirectly, through their L1 equivalents, which are difficult to suppress (Kroll & Curley, 1988; Potter et al., 1984). With increasing skill, the connection becomes more direct (see Kroll et al., 2002). More interestingly, word association studies show that the lexicons of novice learners are organized along phonetic lines more than semantic lines, like those of more skilled speakers (Soderman, 1993; Meara, 1984; Henning, 1974). Meara (1984) found that English-speaking French learners who were given the prime word *béton* ("concrete") produced associations such as *animal* (apparent confusion with *bête*, "beast"), *conducteur* (confusion with *baton*, "baton") and *Normandie* (confusion with *Breton*). Ellis (1996, p. 93) uses the term "clang associates" to describe these sound-based connections.

The preponderance of these clang associates suggests that, at least for low-level learners, accessing L2 words quickly and accurately is not a straightforward task. It requires not only the positive activation of correct items but also the suppression of items that are phonetically related but semantically implausible. It is unclear to what degree this suppression is conscious and subject to attentional control (see Segalowitz, 2000, pp. 207-08). But it seems plausible that the lower the speaker's L2 proficiency, the more this suppression will have to be conscious as well as frequent. If working memory is roughly equated with consciousness—as it often is in the literature (Schmidt, 1992, p.363)—it will be intimately involved in this suppressive process. It should be noted, however, that this equation is not unanimously accepted. One recent study

(Michael & MacWhinney, 2003) treats suppressive ability (as measured in the Stroop test) as separate from working memory capacity, and suggests that the former is a better predictor of early L2 vocabulary learning skill than the latter.

The ACT* model and similar perspectives on automatization (e.g., Bialystok's "control model" [Bialystok, 1989]) are, in Robinson's (1997) words, "resource-based" in that allocation of attentional resources lie at their heart. Other models are termed "memory-based" because they focus on direct retrieval of particular items from memory rather than allocation of attentional resources. Logan's (1988) "instance theory" is perhaps the best-known example. Logan envisions automatization as the end result of an accumulation of numerous instances of solutions to previously encountered problems or tasks. Specific solutions are independently stored in memory and retrieved automatically when the learner encounters stimuli similar to those in the original learning situation. There is no need in Logan's model for a transition from declarative to procedural knowledge, or for application of abstract rules to particular circumstances; particular circumstances trigger particular items, and automatization comes from repeated exposure to a variety of trigger stimuli.

The memory-based view of automatization suggests an important role for working memory, albeit a different one than in resource-based theories. It is a role that places a premium on memory capacity rather than attention. The purported connection between phonological short-term memory (PSTM) and vocabulary learning (Service, 1992; Papagno, Valentine & Baddeley, 1991; Gathercole & Baddeley, 1990, 1989) comes into play. Of particular importance is the link between PSTM capacity and the acquisition of "strings" or "chunks" of L2 speech, multi-word sequences that are learned as wholes. Pawley and Syder (2000), among others, have pointed out the frequent occurrence of collocations and conventionalized expressions in mature L1 speech, and suggested that mastery of these formulaic utterances is a key to L2 fluency. By retrieving these utterances as wholes, a speaker can minimize clause-internal encoding and devote more attention to other tasks, including planning larger discourse units. Arguing along these lines, Ellis (1996) proposes that PSTM capacity determines not only word-learning skill, but also the ability to acquire longer chunks of language and, ultimately, L2 syntax.

From this perspective, automatic speech production is less creative than most rule-based theories would have one believe. It is largely a matter of retrieving pre-assembled strings and rearranging them, sometimes in modified form. Oppenheim's (2000) study of six advanced ESL learners lends credence to this view. Her results indicate that a high proportion (66 percent on average!) of her subjects' L2 speech could be identified as "recurrent sequences." Of course, such analysis depends on how one defines and counts "recurrent sequences." Interestingly, Oppenheim's subjects relied on many utterances that were idiosyncratic and not particularly common in L1 English speech. [25] But the larger point about the importance of chunking in automatized speech production, and its relation to working memory capacity, remains valid and merits further investigation.

Finally, there is reason to believe that superior "chunking" ability helps learners improve with regard to formal accuracy as well as well as purely temporal fluency. [26] A great many errors in novice and intermediate L2 speech can be characterized as local (i.e., intra-clausal) mistakes. These are often mismatches in terms of number, gender, case or subject-verb concordance. The Chinese ESL student who says * *two book*, or the English-speaking Spanish learner who talks about * *el mujer* ("the [masc.] woman") are making agreement errors that would be avoided if they had acquired and accessed these phrases as single "instances." This connection between chunking ability and accuracy should not be exaggerated. Chunking probably does not help avoid non-local errors such as choice of verb tense, or long-distance dependencies (e.g., * *La <u>mujer</u> que habla con Gabi es muy <u>listo</u>*, "The woman [fem] who is speaking with Gabi is very smart [masc]"). Nor will it necessarily help learners accurately produce novel strings containing agreement features; here, application of rules is probably unavoidable. But it may well reduce the number of simple local errors, errors which may distract listeners from the speaker's message.

2.2.6 Summary of working memory research

This review has identified several ways in which working memory capacity may be crucial to developing L2 oral fluency. One is in the storage of intermediate products, particularly during the movement from conceptualization of messages to formulation (Levelt, 1989). Novice L2 learners may need to rely on working memory for storage more than advanced learners, especially during the formulation stage of production (Temple, 1997). A fairly large and consistent body of research (Service, 1992; Papagno, Valentine & Baddeley, 1991; Gathercole & Baddeley, 1990) implicates phonological short-term memory in the learning of lexical items for subsequent retrieval, including multi-word “chunks” of language (Ellis, 1996). It is suggested here that the central executive component of working memory has a role in suppressing interference during lexical access. This suggestion is based on the existence of phonetically-based “clang associates” in novice learners’ lexicons (Ellis, 1996). Finally, Levelt’s (1989) model of speech production gives working memory the important role of monitoring output to ensure that it matches both the speaker’s communicative intent and formal linguistic norms.

This review has also summarized and assessed a variety of perspectives on the multi-faceted nature of oral fluency. These include theoretical and qualitative views such as Fillmore’s (1979/2000) and Brumfit’s (1984/2000), as well as qualitative views such as those of Lennon (1990) and Riggenbach (1991). Although the topic of fluency is too wide to allow for universally acceptable definitions, there does seem to be a general consensus that speed and smoothness of delivery are identifying qualities of fluent speech. The question of how, and how much, formal accuracy matters in assessing L2 fluency remains open. Some preliminary ideas on how language instructors may teach fluency as a skill, at least to more advanced learners (Guillot, 1999) have been broached. Fluency has also been examined from two distinct psycholinguistic, information-processing approaches (Anderson, 1983; Logan, 1988). In both, fluency, or automatization, is linked intimately to working memory although these two approaches differ significantly in whether they emphasize the attentional aspect or the memory capacity aspect of working memory.

The two studies described below were designed to pick up and extend the issues outlined above. The main questions are these:

1. How do L1 listeners assess L2 oral fluency?
2. Does working memory capacity correlate strongly with L2 oral fluency?
3. Is this correlation stronger with low-level L2 learners than with more advanced learners?

3.0 EXPERIMENT 1

3.1 INTRODUCTION

The first experiment was a pilot study aimed at clarifying how native-language listeners judge non-native speakers' oral fluency. Twenty native English listeners listened to recordings of 20 non-native English speakers, and rated them holistically for fluency. These holistic ratings were correlated with a variety of objectively measurable factors, including factors related to speed of delivery, hesitations, syntactic complexity and morphosyntactic accuracy. The latter factor was included partly to determine whether any of the non-native speakers could be said to have exhibited "frozen fluency," that is, speech that was good in terms of temporal fluency (speed and hesitation markers) but low in terms of grammatical and/or morphological correctness. In addition, each of the native-language fluency judges was interviewed to obtain qualitative information about how they think about and rate non-native fluency. It was hoped that this first experiment would provide a set of useful and reliable correlates that could be used to measure non-native oral fluency in the second experiment.

3.2 METHODOLOGY

3.2.1 Participants

The 20 non-native speakers were recruited in the Pittsburgh community, and many were foreign students at the University of Pittsburgh. Some of these were learners at the university's English Language Institute, an intensive English as a Second Language (ESL) program. All of the latter had achieved at least an intermediate level of proficiency in English. Beginning-level learners were excluded because the researcher believed that the narrative task would be too difficult and frustrating for them. The non-native speakers came from diverse L1 backgrounds, including Chinese languages (8 participants), Spanish (6), Hungarian (2), Portuguese, Korean, German and Bosnian (1 each). They were paid for their participation in the study, as were the native English-speaking fluency judges.

The 20 judges were recruited in Pittsburgh and many were students at the University of Pittsburgh. All demonstrated native-language English ability during their interactions with the researcher. None of them were language instructors or students of foreign language education. The researcher deliberately excluded language specialists in order to obtain "naive" judges who were more representative of the larger, non-specialist community. Such non-specialists who are the primary "consumers" of non-native speech, and it is ultimately their attitudes toward fluency and non-native speech that matter most.

3.2.2 Task and Materials

As in Lennon's (1990) study, the format used to elicit non-native speech samples was a narrative monologue based on a picture book. Here, the picture book was *Frog, Where are You?* (Mayer, 1969). This small book contains 24 wordless drawings that tell the story of a young boy whose pet frog escapes from home. The boy and his dog search for the frog in a forest, and have some adventures along the way. *Frog, Where are You?* has been used previously in linguistic research, especially in the study of narrative skill development in children. For example, Berman and Slobin (1994) and their collaborators used the book as a stimuli in their cross-linguistic study of narrative development.

The researcher instructed the non-native speakers to narrate the story in *Frog, Where are You?* as fluently as they could. They were allowed to browse through the book before starting the task in order to familiarize themselves with the pictures and the plot. They were also allowed to look at the pages as they were narrating the story, and they were asked to make at least one remark about each picture. The goal was to record 3 minutes of speech for each non-native speaker. Seventeen of the 20 speakers produced at least 3 minutes of speech; three other speakers finished the story in less time, but all spoke for at least 2 minutes and 20 seconds. The recordings took place in a soundproof language laboratory recording studio.

The native English fluency judges listened to these recordings individually. They did not see the speakers, and did not know their names or L1 backgrounds. The judges listened to the recordings in a random order that was different for each judge. After listening to each recording, they rated each speaker for overall fluency on a 1-to-10 scale, with 10 being the highest category. Each judge took a short, mandatory break after listening to half of the 20 recordings; the purpose was to avoid listener fatigue.

The researcher did not provide the judges with a definition of “fluency.” Rather, they were told to use their own judgment as to what constituted fluency. This was to avoid imposing a particular view of fluency on the judges.

3.2.3 Analysis

Each recording was also analyzed in terms of 17 distinct factors that were potentially related to oral fluency. Most of these factors had been used in one or more previous studies of non-native oral fluency, including the previously cited studies by Lennon (1990), Riggenbach (1991), Freed (1995) and Kinkade (1996). These factors fall into four categories, namely those related to speed, smoothness (i.e., hesitation-related phenomena), syntactic complexity and morphosyntactic accuracy. The first two categories generally characterized “narrow” or “low” fluency, while the latter two focused on “broad” or “high” fluency (Lennon, 1990, 2000). These factors are listed in Table 21.

In the speed-related category, the term “pruned” is used as in Lennon (1990), and it excludes repetitions, corrections and asides. Thus, while factors 3 and 4 are primarily speed-related, they do incorporate an element of smoothness. The term “T-unit” refers to an independent clause, along with its subordinate clauses and related non-clausal material. It has been used as a measure of syntactic complexity in previous second language acquisition research (Halleck, 1995; Gaies, 1980).

Table 21: Fluency factors measured in Experiment 1

Speed-related factors	1. Total words/min. 2. Total syllables/min. 3. Pruned words/min. (excludes repetitions, corrections, asides) 4. Pruned syllables/min. 5. T-units/min.
Smoothness-related factors (voiced)	6. Total filled pauses/min. 7. Repetitions/min. 8. Self-corrections/min. 9. All voiced disfluencies (6,7,8)/min.
Smoothness-related factors (unvoiced)	10. Total unfilled pauses/min. 11. Mean length of unfilled pauses 12. Juncture pauses (pauses at clause boundaries)/min. 13. Mean length of juncture pauses (in msec.) 14. Non-juncture pauses (pauses within clauses)/min. 15. Mean length of non-juncture pauses (in msec.)
Syntactic complexity	16. Mean length of T-units
Morphosyntactic accuracy	17. Errors/100 words

The smoothness related factors were divided into two subcategories, those related to voiced disfluencies and those related to unvoiced disfluencies. In the former subcategory, "filled pauses" were pauses filled by hesitation syllables or other non-words such as "uh" or "mmm." Following Lennon (1990), filled pauses that occurred within repetitions or self-corrections were not counted. Repetitions included repetitions or words, parts of words or phrases. Repeated words or phrases that were plausibly used for emphasis (i.e., "very, very big") or rhetorical effect were not counted. Self-corrections included speaker-initiated corrections of words, parts of words or phrases.

In the unvoiced disfluency subcategory, silences were counted as pauses if they were at least 250 milliseconds in length. This was based on research by Goldman-Eisler (1961) and Deese (1980). Pauses were measured using PEAK LE version 2.61 speech analysis software. Because identifying hundreds of stretches of silence and measuring them accurately in milliseconds was extremely time-consuming, unfilled pauses were counted for only the first 60 seconds of each speech sample. These 60 seconds represented at least a third of each speaker's total speaking time in this experiment, and it did not appear that any of the non-native speakers changed their pause profile significantly in the latter parts of their narrative.

In the voiced disfluency subcategory, both "juncture" and "non-juncture" pauses were counted and measured because it was felt that the distinction between these two types of pauses could be important. The researcher felt, again following Lennon (1990), that non-juncture (intra-clause) pauses might be considered less natural and native-like, and might be judged more harshly by native language judges.

The researcher included factors 16 and 17 to determine whether "high" factors, namely the complexity and accuracy of speech, correlated strongly with holistic fluency ratings and with temporal factors. If they did, it could lend credence to the idea that "fluency," at least as viewed by untrained listeners, is difficult or impossible to separate from those purely temporal measurements. The mean length of T-units was chosen as the measure of syntactic complexity, although it is admittedly an imperfect tool. For instance, T-unit measurements do not account well for subject pronoun deletion in coordinated sentences (i.e., *Jason went to the bank and withdrew the money* counts as only one T-unit). Nevertheless, it seemed prudent to include at least one measure of complexity to account for any speakers who might engage in highly simple and repetitive speech.

Morphosyntactic accuracy was measured as errors per 100 (unpruned) spoken words. Counting errors is a potentially subjective process, and for that reason errors were limited to four specific types (see Table 22). Errors were not counted if the speaker self-corrected them, or if the forms in question were unusual but at least marginally acceptable to native speakers. Two native English speakers who have worked as ESL instructors counted morphosyntactic errors using these criteria. Their inter-rater reliability coefficient was .92.

Table 22: Types of morphosyntactic errors counted in Experiment 1

1. a clearly incorrect word choice (e.g., "lizard" for "frog")
2. a clearly incorrect word form (e.g., "children" for "child")
3. absence of an obligatory word (e.g., "he saw [*a, the*] frog")
4. insertion of a word that renders a clause ungrammatical (e.g., "the owl flew out to of the hole")

Many other potential disfluency markers were not used, mainly for practical reasons of time and resources. Some of these other factors, such as quality of pronunciation and intonation, or perhaps breadth of vocabulary usage, may have affected the judges' ratings. However, these factors are even more difficult to quantify with precision than syntactic complexity or morphosyntactic accuracy. Several of these factors were mentioned in the post-rating interviews.

After completing the rating task, each of the 20 native listener-judges was asked a series of five questions (see Appendix A). These questions were designed to probe more qualitatively and in more depth how the judges had gone about their task and about their views of oral fluency in general. The first two questions focused on which qualities of speech had been most important to the judges in their ratings. Questions 3 and 4 asked the judges to talk more generally about the nature of fluency, and whether "fluency" meant something different for native and non-native speakers. The final question was a semi-forced choice question where the judges were asked to list three factors that are most important to fluency.

3.3 RESULTS

3.3.1 Holistic Fluency Ratings

Holistic fluency ratings for each of the 20 non-native speakers were arrived at by taking the average of the 20 judges' ratings. These mean fluency ratings ranged from a low of 2.90 on the 10-point scale to a high of 8.85. The overall mean for all 20 speakers was 5.75 (S.D. = 2.07), and the median score was 5.0, exactly in the middle of the scale. The highest standard deviation for any individual speaker was 1.82, which represents less than a fifth of the available rating scale. Seven speakers received at least one “perfect” score of 10, and one speaker received seven such ratings. At the other end of the scale, five speakers received at least one score of 1, the lowest rating available, and one speaker received three such scores.

3.3.2 Speed factors

The 20 non-native speakers varied widely in their speed of delivery, with the fastest speakers being more than twice as fast as the slowest ones, according to most of the measurements. The fastest speaker spoke at a rate of nearly 140 words, or nearly 170 syllables, per minute. The slowest speaker delivered a bit more than 50 words per minute, or 66 syllables per minute (unpruned). Results for speed-related factors, including their correlation with the holistic fluency ratings, are shown in Table 23.

These figures clearly indicate a significant correlation between speed of delivery and the judges’ fluency ratings, however speed is quantified. The two syllable-based measures provide slightly stronger correlations than the word-based measures. The single clause-based measure (T-units per minute), though it correlated significantly with the fluency ratings, nonetheless produced the weakest correlation of these five factors.

Table 23: Speed-related factors (all figures are per minute except correlations)

	Total Words	**Pruned words**	**Total Syllables**	**Pruned syllables**	**T-units**
Low	56.7	50.9	65.8	56.0	5.1
High	139.9	139.5	169.0	168.7	14.0
Mean (S.D.)	98.7 (24.2)	93.7 (26.9)	115.5 (30.9)	109.3 (34.0)	9.63 (2.8)
Correlations w/ fluency	.729 **	.743 **	.777 **	.801 **	.623 **

** $p < .01$

However, it appears that speed must be considered in context, as one of several most relevant factors. Speaker 2, a Spanish-speaking woman, was objectively the fastest speaker in most of these five categories, yet she rated only eighth in the holistic fluency judgments. Also, speaker 15, a Portuguese speaking woman, received the highest fluency rating (8.85, with seven "perfect" scores of 10), but she was only slightly above average in terms of speed (she ranked eighth in total words per minute, for example).

3.3.3 Smoothness-related factors

Data for voiced disfluencies are shown in Table 24. The repetition measure produced the only significant correlation with the fluency ratings, at -.649 (as might be expected, the correlation here is negative). The -.450 correlation between fluency and the "total voiced disfluencies" measure is clearly the result of a carry-over effect from the repetitions factor.

Table 24: Factors related to voiced disfluencies (all figures are per minute except correlations)

	Filled pauses	**Repetitions**	**Self-corrections**	**Total voiced disfluencies**
Low	0	0	0	0.9
High	10.3	9.0	11.0	18.0
Mean (S.D.)	3.54 (3.00)	2.83 (2.51)	1.73 (2.36)	7.74 (5.47)
Correlation with fluency ratings	- .053	- .649 **	- .395	- .450 *

* $p < .05$, ** $p < .01$

The absence of a relationship between the number of filled pauses speakers produced and their fluency ratings is noteworthy. Indeed, many speakers who were judged to be highly fluent filled their narratives with "umm's" and "ah's" and other fillers. Speaker 6, a native Hungarian speaker, produced 31 filled pauses in three minutes yet she received a mean fluency rating of 8.15 and three "perfect" scores. Both speaker 15, the Portuguese speaking woman who was rated highest for fluency by the judges, and speaker 7, who ranked last of all 20 speakers, produced the same number of filled pauses (eight in three minutes).

Table 25 presents data on unfilled pauses. The most interesting correlations come from the non-juncture pauses. These two factors (number of non-juncture pauses per minute and mean length of non-juncture pauses) correlated significantly (and, again, negatively) with the fluency ratings. That is, speakers who were judged to be relatively high in fluency tended to pause less frequently within clauses, and to pause for a shorter time when they did. On the other hand, data from juncture (i.e., inter-clausal) pauses provided weak- in fact, positive- correlations with fluency ratings. The correlation produced by the total unfilled pauses appears to be a by-product of the non-juncture pause data.

Table 25: Factors related to unvoiced disfluencies

	Total unfilled pauses	**Mean length pauses**	**Juncture pauses/min.**	**Mean length juncture pauses**	**Non-juncture pauses/min.**	**Mean length non-juncture pauses**
Low	9.0	.586	7.0	.553	1.0	.322
High	32.0	1.338	21.0	1.584	25.0	.907
Mean (s.d.)	22.3 (5.2)	.830 (.193)	14.2 (3.9)	.942 (.246)	8.1 (5.3)	.557 (.165)
Correlation with fluency ratings	-.531*	.283	.153	.230	-.633**	-.606**

* p < .05, ** p < .01

3.3.4 Syntactic complexity

Individual speakers varied greatly in the syntactic complexity of their utterances, as measured by the mean length of their T-units. As shown in Table 26, these means ranged from a low of just under 6.5 words per T-unit to nearly 15 words. However, the correlation between this factor and the judges' fluency ratings was weak. It's worth noting that speaker 15, the Portuguese speaker who was highest in the fluency ratings, was lower in this particular measure (8.19 words per T-unit) than the two least fluent speakers (8.48 and 8.55 words per minute). In brief, syntactic complexity did not appear to be an important factor in the judges' ratings.

Table 26: Mean length of T-units (in words)

	Mean length of T-units (in words)
Low	6.43
High	14.80
Mean (S.D.)	9.49 (1.84)
Correlation w/ fluency ratings	.310

3.3.5 Morphosyntactic accuracy

The overall mean error rate for all 20 speakers was 3.83 errors per 100 words. There were enormous variations here, as shown in Table 27. Nine speakers had error rates that were under 2.0 errors per 100 words, and this subgroup included the seven most fluent speakers, as judged by the native listeners. The correlation between morphosyntactic accuracy and fluency was -.769, one of the strongest correlations found in this experiment.

Table 27: Morphosyntactic error rate

	Errors/100 words
Low	0.27
High	10.66
Mean (S.D.)	3.83 (3.02)
Correlation w/ fluency ratings	-.769 **

** $p < .01$

It should be pointed out that few of the errors that the speakers made were of the kind that could significantly hinder communication. They tended to be "local" rather than "global" errors in that they did not compromise the comprehensibility of the speaker's utterance. Common types of errors were lack of subject-verb agreement (often dropping third person singular -*s*), tense errors (leaving off past tense -*ed* when it was required) and missing definite or indefinite articles. A few typical mistakes are shown in Table 28.

Table 28: A sample of speakers' errors

a.	...he started to call his frog, but of course he didn't, he *didn't be* there
b.	During the night, the frog jumped out of the jar and decided to *left* the room.
c.	So the little boy *start* to look...
d.	When the boy *wake* up in the morning...
e.	A little boy...*have* got a frog...
f.	...maybe he *think it* time to go out.
g.	They're getting ready to go to bed and there is a jar with a frog in it, and [*the*] dog is curious.
h.	When they went to a place under [*a/the*] tree....

3.3.6 Multiple regression analysis

A maximum R-square multiple regression analysis was performed on the data to determine which combination of factors best accounted for the native listeners' fluency ratings. This kind of analysis begins with the single highest correlation, and then looks at various combinations of multi-variable factors. The results are shown in Table 29.

Table 29: Maximum R-square multiple regressions analysis

	Variables/Groups of variables	***R-square***
1.	Pruned syllables/min.	.6414
2a.	Repetitions/min. & Morphosyntactic error rate	.8358
2b.	Pruned syllables/min. & Morphosyntactic error rate	.8242
3.	Syllables/min. & Repetitions/min. & Morphosyntactic error rate	.8942
4.	Syllables/min. & Repetitions/min. & Non-juncture pauses/min. & Morphosyntactic error rate	.9074
5.	Syllables/min. & Filled pauses/min. & Repetitions/min. & Non-juncture pauses/min. & Morphosyntactic error rate	.9231

The results indicate that the pruned syllables per minute variable is the single best regressor. There are two different two-member groupings of variables (labeled 2a and 2b in Table 29) that of nearly equal strength: the first is repetitions per minute plus morphosyntactic error rate, the other is pruned syllables per minute plus morphosyntactic error rate. The three-, four- and five-variable groupings include various combinations of speed and smoothness factors, in addition to morphosyntactic error rate. The latter appears consistently in all multi-variable analyses.

3.3.7 Post-Rating Questions

3.3.7.1 Questions 1 and 2. After they had finished listening to and rating the non-native speakers' recordings, the 20 native listener-judges were asked five questions designed to probe more fully and qualitatively their views of oral fluency (see Appendix A). The first two questions asked the judges which specific qualities of speech made them think that a particular speaker was fluent or not. These were open-ended questions and the judges' replies were not easy to categorize, but a few themes stand out, particularly vocabulary usage and hesitations.

Vocabulary and the extent of a speaker's vocabulary were mentioned 19 times in one form or another. One judge said she looked favorably on speakers who used more sophisticated words and "more intricate adjectives." Another judge commented that some speakers used larger and more descriptive words than others. A third judge said that speakers who used the same words repeatedly seemed less fluent to her. Given the large number of mentions, it seems that some measure of vocabulary richness (perhaps a type/token ratio of distinct words/total words spoken) would have provided a strong correlation with the judges' holistic fluency ratings.

An equally common theme centered around pauses and hesitations. Nineteen judges mentioned this factor in some way, often in terms of "delays" or "fumbling around." One judge noticed "protracted pauses" in less fluent speakers, and a second judge said that the most fluent speakers were "not so hesitant" as the others. Another judge said that the speakers she regarded as more fluent "didn't say '*um*' a lot," while the less fluent speakers did. (In fact, as mentioned above, the occurrence of filled pauses did not correlate particularly well with the judges' fluency ratings.) None of the 20 judges mentioned the placement of pauses.

Three other issues that many judges brought up were grammatical accuracy, narrative skill and pronunciation. Fourteen judges mentioned grammatical mistakes, either in general or with respect to particular types of mistakes (wrong verb tenses, lack of subject-verb agreement, etc.). Eleven judges referred in some way to narrative skill, or coherence in telling the story. For example, one judge gave lower ratings to speakers who "hopped around." Others spoke about poor transitions from one scene to another. Ten judges said that pronunciation had played an important role in their ratings, and a few mentioned particular sounds. One judge said that more fluent speakers had "clear vowels," while another judge noted errors in pronouncing [l] and [r] sounds.

Only three judges mentioned speed of delivery. This seems to indicate that they were not paying attention, at least not consciously, to how fast the non-native speakers were talking. Two judges cited syntactic complexity.

3.3.7.2 Questions 3 and 4. These questions focused on the nature of fluency (question 3) and whether the judges viewed "fluency" the same when listening to non-native speakers as they did for native speakers (question 4). In response to question 3, most of the judges answered in terms of overall communicative skill, or the ability to get ideas across to listeners. Some of the judges' definitions of "fluency" and/or comments are reproduced in Table 30.

Table 30: A sampling of responses to question 3

a.	"how easily someone could relay a message to me, get their point across"
b.	"being able to get your ideas across with ease"
c.	"comfort with the language"
d.	"lucid, logical speaking"
e.	"the (fluent) person can express themselves and communicate"
f.	speakers are fluent "if I can understand them in a social setting"
g.	"a steady flow of words with confidence"
h.	a fluent speaker is "someone who knows how to describe something in the best possible way"
i.	"mastery of the language"
j.	"a confluence (of skills) that results in understandability"

Perhaps not surprisingly, the judges seemed more interested in the communicative results of speech rather than in how speakers achieved those results. This is consistent with the general nature of the question (and the fact that they had just been asked to cite specific qualities in the first two questions).

This emphasis on communicative ability may appear to suggest a relatively lax attitude toward grammatical correctness. However, seven judges volunteered that grammatical accuracy, or lack of it, was important. Again, some judges mentioned particular types of mistakes, such as choice of wrong verb tenses, while others spoke more generally. One judge said that fluency included the ability “to use tense, grammar and appropriate adjectives to describe a situation.” Only one judge brought up speed of delivery in response to question 3, though a few others said that fluent speech as “even-paced” or displayed a steady rate.

In response to question 4, almost all of the judges replied in the negative, that is, that they judge non-native speakers’ fluency by a different standard. Most said something to the effect that they held non-native speakers “to a weaker standard” or gave them “the benefit of the doubt.” Said one judge, “I usually give non-native speakers a little more slack just because the English language is so hard.” Another said that with native speakers, “I’d make different excuses for the same pauses” than with a non-native speaker. A third judge struck a similar note, saying that while native speakers hesitate “for social reasons,” non-native speakers are more likely to hesitate for “decoding” reasons. Only four of the 20 judges said they used the same standard for judging native and non-native fluency.

3.3.7.3 Question 5. In this question, the judges were given a list of six speech qualities and asked to check off three that were the most important ingredients of oral fluency. They were also allowed to write in other options. Results are shown in Table 31. Both grammatical accuracy (listed as “how many grammatical mistakes they make”) and pronunciation (listed as “how well they pronounce words”) were the top choices. It seems a bit surprising that 12 judges chose syntactic complexity, in light of the relatively low correlation between syntactic complexity and fluency ratings in the quantitative analysis. It might be that the T-unit-based measurement tool used in that analysis was not a reliable quantification of syntactic complexity. Alternately, it may be that all or most of the 20 non-native speakers had achieved some minimal “floor” level of syntactic complexity that the judges found acceptable. Or, it may be that the judges did not give as much attention to syntactic complexity during the actual rating task as they did later in the post-task interview. Other judges chose hesitation-related factors, with more of them (10) focusing on length of hesitations and fewer (4) choosing number of hesitations.

Table 31: Responses to question 5

Factor	***Number of mentions***
Number of grammatical mistakes	15
Pronunciation	15
Complexity of sentences	12
Length of hesitations	10
Number of hesitations	4
Speed of speech	2
Extent of vocabulary *	1
Use of rhetorical structures *	1

* write-in choices

Only two judges checked off the speed factor (listed as “how fast they speak”). This is also surprising, given how strongly various speed-related factors correlated with the judges’ holistic fluency ratings. It may be that speed was less salient to the judges’ ears than grammatical errors or mispronunciations; perhaps they noticed speed of delivery less consciously than they noted these formal miscues. Once more, there is an apparent mismatch between the results of the quantitative analysis and the responses produced in the qualitative questioning. It should be noted that most of the quantitative speed-related measures correlated moderately significantly with morphosyntactic accuracy, generally at the (negative) .50 level (see Table 32). That is, the non-native speakers who spoke fast also tended to speak correctly. But this does not explain why the judges gave more weight to one factor (accuracy) than the other (speed) in the interviews.

Only two judges offered write-in responses to question 5. One wrote that how a speaker uses the “rhetorical structures of speech” was important in fluency. Another cited the extent of the speaker’s vocabulary. As mentioned above, vocabulary usage was a popular choice in response to question 1. Perhaps more judges would have chosen this factor in question 5 if it had been placed on the list of ready-made options.

Table 32: Correlations between speed-related factors and morphosyntactic accuracy

Factor	***Correlation with morphosyntactic error rate***
Words/min.	-.543 *
Pruned words/min.	-.461 *
Syllables/min.	-.559 *
Pruned syllables/min.	-.498 *
T-units/min.	-.280

* p < .05

3.4 DISCUSSION

The main research question in this experiment was: how do native listeners assess fluency in non-native speech? That is, which factors do they pay most attention to? An ancillary question was whether formal accuracy and syntactic complexity are important factors in their assessments. The data reported here indicate that native listeners pay attention to many factors, including speed of delivery, morphosyntactic error rate, repetitions and non-juncture clauses. All of these factors correlated strongly with the judges' holistic fluency ratings (Table 33). Syntactic complexity, at least as measured in this experiment, did not prove to be a strong predictor of fluency judgments. However, many judges mentioned this factor in response to question 5 during the post-rating interviews, suggesting that syntactic complexity might be a factor, perhaps if it was quantified differently.

From these results, it appears that L1 listeners (at least non-language specialists) perceive "fluency" in the sense of Lennon's (1990, 2000) "high" or "broad" fluency, rather than in the "narrow" or "low" sense of the word. Formal accuracy figures highly in their assessments, as indicated by the strong correlations of the morphosyntactic error rate with the judges' fluency ratings. The regression analysis supports this view, as the morphosyntactic error rate was the single factor that appeared in every multiple-variable grouping of most significant predictors. And, it should be repeated, this emphasis on accuracy existed despite the fact that a great many errors committed by the non-native speakers did not obviously hinder comprehension. It may be that while each of these "local" errors was not itself costly in terms of overall fluency, in the aggregate they create an impression of disfluency in listeners' minds. The ACTFL guidelines that refer to errors that "disturb the native speaker" come back to mind. It seems likely that formal accuracy would be an even stronger predictor if the pool of non-native speakers was extended to low-level ESL learners.

Table 33: Some strong correlation with holistic fluency ratings in Experiment 1

General category	***Factor***	***Correlation w/ fluency ratings***
Speed	Pruned syllables/min.	.801 **
Accuracy	Morphosyntactic errors/ 100 words	-.769 **
Smoothness	Repetitions/min.	-.649 **
Smoothness	Non-juncture pauses/min.	-.633 **

** $p < .01$

It's also important to note that the error rate correlated significantly with most speed-related factors measured in this experiment. For these learners, temporal speed and accuracy seemed to have progressed together. Coupled with the above observations, this suggests that the notion of "frozen fluency," of a kind of narrow, temporal fluency co-existing with low formal accuracy, is a chimera. Furthermore, this combination would probably not be viewed as true fluency by naive native listeners (i.e, the great majority of potential listeners for most ESL learners).

This study confirmed the results of previous work (Lennon, 1990; Riggenbach, 1991) in that several factors thought to be potential disfluency markers did not in fact seem to influence the judges' ratings. These factors include filled pauses, unfilled pauses that occur at clause junctures and self-corrections. This study adds syntactic complexity to that list. This may provide some small comfort to anxious L2 students.

The limitations of this experiment should be acknowledged. Firstly, the narrative monologue task used here excluded considerations that come into play in dialogues and multi-person exchanges. These include the non-native speaker’s ability to comprehend what interlocutors say and to respond appropriately, to take conversational turns and to “keep the floor” when desired. These were purposely overlooked for methodological reasons: the researcher wanted to focus closely on speech production and keep the topic as uniform as possible. Secondly, the non-native speakers in this experiment were all intermediate- or high-level English learners. Inclusion of beginners may have produced different results by broadening the spectrum of proficiency that the judges heard. Lastly, while the non-native speakers came from many parts of the world, most of the listener-judges used in this experiment came from a fairly small region of the United States, namely western Pennsylvania. It may be that inclusion of listeners from others parts of the U.S. or the English-speaking world may have produced a group of judges with different attitudes toward non-native speech, either more or less lenient or simply different in the factors they focused on.

4.0 EXPERIMENT 2

4.1 INTRODUCTION

The chief aim of the second experiment was to determine whether, as hypothesized, significant positive correlations exist between working memory capacity and L2 oral fluency. Forty-four participants were given working memory tests in their L1 (English), as well as tests of oral fluency in their L2 (Spanish). All of the participants were studying, or had recently studied, Spanish as a foreign language in a college setting, and all had successfully completed at least one semester of study. The participants differed widely in their length of study; some were in their first year, others had two, three or more years of Spanish study. This was done so that the relationship between working memory and fluency could be studied at different level of proficiency. It was hypothesized, based on Temple (1997) and other sources, that the correlation would be stronger for novices than for more advanced learners. Each participant took a general Spanish proficiency test to establish their ability level, independent of their length of study.

Working memory capacity was measured in three ways: with a Speaking Span Test, a Math Span Test and a Non-Word Repetition Test. The first of these measures was intended to measure verbal working memory in an oral linguistic mode. The second was a largely non-linguistic type of working memory measure that used numbers and arithmetic problems as material. If working memory capacity is a general-capacity faculty, and not dependent on the mode in which it is measured, this measure should have correlated well with oral fluency measures. The Non-Word Repetition Test was included in order to tax a more specific faculty, phonological short-term memory or the phonological loop. The Speaking Span Test and Math Span Test were conducted in English to avoid confounding working memory with L2 proficiency. The Non-Word Repetition Task used Arabic words, as the purpose here was to test the participants' ability to recall and repeat unfamiliar words (none of the participants spoke or had studied Arabic).

The participants' L2 oral fluency was measured with three tests as well. The most comprehensive of these was the Narrative Monologue Test, wherein narratives were elicited with a children's picture book. Each participant narrated the story in Spanish for 3 minutes, and this speech sample was analyzed for four fluency-related factors, which were selected based on the results from Experiment 1. In addition, the participants took an English-to-Spanish Word Translation Test to measure the speed and accuracy of their L2 vocabulary recall. It was hypothesized, based on studies by Meara (1984) and others, and on Ellis' (1996) notion of "clang associates," that the suppressive, controlled attention aspect of working memory would give high-span learners an edge on this task. Finally, the participants took the Imitation/Grammaticality Test. This test was designed to simulate participants' ability to monitor speech quickly and accurately, and it was included in order to test Levelt's (1989) notion that working memory plays a role in the monitoring of one's own speech. Although the monitoring in this test was not produced by the participants themselves, it may have provided a suitable substitute. In short, the three fluency tests were designed to measure the participants' fluency in actual L2 speech performance, their L2 lexical retrieval fluency and their monitoring skill.

The 44 participants were recruited mainly from students in Spanish classes at the University of Pittsburgh, Carnegie-Mellon University and Carlow University, all located in Pittsburgh. As it turned out, all but four came from the University of Pittsburgh. All of the participants were native speakers of English who had studied Spanish for at least one semester at the college level with a (self-reported) passing grade. Thirty-four were female, and 10 were male. The participants ranged in age from 18 to 41, and their average age was 20.7 years.

All potential participants filled out a brief questionnaire that elicited information about their language learning history (see Appendix B). Only native speakers of English who were studying, or had recently studied, Spanish as a foreign language were considered. Potential participants were screened to ensure that anyone with considerably more Spanish practical speaking experience than their cohorts was excluded. For example, anyone who spoke Spanish at home with their parents or other relatives was excluded. This included "heritage learners" who spoke Spanish occasionally with grandparents or others. Also, anyone who had spent more than three months abroad in a Spanish-speaking environment was eliminated from consideration. Finally, all applicants who spoke Arabic or had studied Arabic were turned away. This latter restriction was added so that none of the participants enjoyed an advantage on the Non-Word Repetition Task, which used Arabic words.

The questionnaire also elicited data on highest level of college Spanish studied, grades received in Spanish classes and level of motivation for studying Spanish. Eighteen participants were in the first year of college-level Spanish study, 16 in the second year, seven in the third year and three in the fourth year. The average Spanish grade reported by the participants was 3.72 on a traditional 4.0 scale (A = 4.0, B = 3.0, etc.), or around a B+ grade. Participants rated their motivation for learning Spanish on a 10-point scale, with 10 representing the highest motivation possible. The mean motivation for the group was 7.93. If this self-reported data was accurate, this group of participants was probably above average in terms of classroom grades and motivation.

The participants' average score on the Spanish Proficiency Test was 26.1, out of a possible 35 points (S.D. = 3.89). Their scores ranged from a low of 14 to a high of 31.5. The proficiency scores were used to categorize the participants in terms of their overall level of Spanish language skill. It was felt that this provided a more reliable indication of their overall L2 skill level than their placement in particular classes or their self-reported grades.

4.2 METHODOLOGY

4.2.1 Working memory measures

4.2.1.1 Speaking Span Test. This test was based on Daneman and Green (1986) and Daneman (1991). Participants viewed sets of words, one word at a time, on a computer monitor. Each word remained on the screen for 1 second, as was done in those studies. After the last word in each set disappeared from view, the participants saw a prompt (a question mark) on the screen. They then had to produce original sentences for each word in the set and say those sentences aloud. The sets of words increased progressively, from two to six, and the load on working memory increased accordingly. That is, for a two-word set, the participants had to recall both words and produce two sentences; for a six-word set, they had to recall six words and produce six sentences.

A list of 100 common English words was used in this test (see Appendix C). The words were comparable in phonetic and orthographic length; all contained two syllables and seven letters. Each participant viewed all of the words, and each participant was allowed 60 seconds to produce sentences for each set. Two scores were recorded, a "span score" and a "total score." The span score was the largest set of words, from two to six, that the speaker could successfully perform the task for three out of five times. A half point was added if they completed the task two out of five times. The "total score" represented the total number of words, out of 100, that they successfully completed the task for. Each participant had five practice trials to become familiar with the task before beginning in earnest.

The participants were advised in advance that the original sentences they produced had to be grammatically and semantically acceptable. Participants were also told to avoid repeating the same sentence pattern for different words. For example, the first two-word set in the test was *kitchen, farmers*. Participants would not be given credit for producing sentences such as "I saw the kitchen" and "I saw the farmers." This restriction (which was apparently not included in Daneman and Green (1986) or Daneman's (1991) experiments) was meant to prevent participants from over-relying on rote patterns that do not truly tax the processing component of working memory.

Upon completion of the task, the researcher briefly interviewed the participants. They were asked how they had gone about the task, and specifically whether they had adopted any "tricks" or strategies to remember the target words. For instance, in the example given above, a participant might imagine farmers standing in a kitchen. This questioning was included to help determine if and how working memory capacity might be functionally extended by conscious, skillful use of strategies.

4.2.1.2 Math Span Test. This test was based on experiments conducted by Salthouse and Babcock (1991) and Roberts and Gibson (2002). The procedure was similar to that of the Speaking Span Test described above. However, instead of words, the participants saw simple addition and subtraction problems (i.e., 5 + 3 = ?, 6 - 2 = ?) on a computer monitor. Each problem was on the screen for 5 seconds, as in those studies. Participants were directed to do two things: state the answer to the problem aloud immediately, and remember the second, or “target,” digit in each problem for later recall. In the two-problem example set above, the target digits are 3 and 2. The number of problems in a set ranged from two at the outset of the test, up to six at the end. After each set of problems, participants saw a prompt (a question mark) on the screen. At this signal, they had to recall the target digits in the same order in which they had appeared.

To get credit for a set, participants had to get both parts of the task right: they had to solve the arithmetic problems and correctly recall the target digits in order. As in the Speaking Span Test, there were two scores, a “span score” and a “total score.” The span score ranged from two to six and it represented the largest number of problems that a participant could complete the task for, at least three out of five times in each set. Participants received a half-point for doing the task correctly twice in five attempts. The total score was the total number of times, up to 100, that a participant solved the arithmetic problem and recalled the corresponding target digits correctly.

As in Salthouse and Babcock (1991), the arithmetic problems were simple problems of the X + Y = ? or X - Y = ? type, where X and Y were one-digit number numbers between 1 and 9 (see Appendix D). None of the answers to the problems were negative numbers, and the target digits were never the same number for two consecutive problems. However, whereas Salthouse and Babcock provided three possible answers and asked their subjects to check off one, participants here had to provide the answers orally, as in Roberts and Gibson’s version of this task (2002). Each participant was allowed five practice trials before beginning to become accustomed to the task.

Here too, there was a brief post-task questioning. The researcher asked each participant whether he or she had used any “tricks” or strategies to remember the target numbers while solving the arithmetic problems. One such strategy would be to link the target numbers (3,2 for instance) to some number that is meaningful to the participant, perhaps an old address or the age of a relative.

4.2.1.3 Non-Word Repetition Test. This was a test of short-term phonological memory. Unlike the two span tests, there was only one task and no splitting of attention. The test used here was based on French's (2003) version of the test. French had native French-speaking children repeat 40 Arabic words as accurately as they could. The 40 words were divided into four 10-word sets, consisting firstly of words with two syllables (i.e., *habba*), then words with three syllables (*wasala*), four syllables (*tahadatta*) and finally five syllables (*mutafailun*) (see Appendix D). For the present study, a native speaker of Arabic living in Pittsburgh pronounced each word and his pronunciations were recorded on a compact disk. Three seconds of silence followed each word. The participants later listened to this recording and attempted to repeat the words as accurately as they could. (It should be noted that these words were, of course, real Arabic words; they were "non-words" only in the sense that the participants did not know them.)

Each of the participants' 40 repetitions were taped and subsequently reviewed by a judge who scored them for faithfulness to the original, native Arabic pronunciation. The judge was a native English speaker who was trained in phonetics and had practical experience in evaluating L2 (though not Arabic) pronunciation. She scored each repetition as either correct or not. Before listening to the tapes, she was told that in order to be considered correct, a repetition could neither add nor delete phonemes. Correct repetitions also had to be free of phoneme switches, or changes in syllable stress. As in French (2003), participants were not be penalized for phonetic variation (accent) unless this caused confusion with other phonemes. It was decided to use a native English speaker rather than an Arabic speaker as judge partly to make such allowances for English-influenced accents easier. A participant's score was the total number of correct repetitions, with a maximum score of 40.

4.2.2 L2 fluency tests

4.2.2.1 Narrative Monologue Test. In this task, the participants were asked to produce an original 3-minute narration in Spanish. The stimulus for this task was a children's picture book titled *Follow Carl!* (Day, 1998). This book contains 18 color drawings which depict a story about a large, friendly dog named Carl, who leads half a dozen children on a game of "follow the leader." This book was chosen because of the simplicity of the story and of the objects and scenes it depicts. The participants were instructed to speak as quickly and as accurately as they could, and to say at least one sentence about each of the pictures (see Appendix E). They were allowed to look at the book for up to 2 minutes before starting the task so that they were familiar with its content. They were not given more time in order to prevent them from memorizing in advance large parts of their narrative. Participants were allowed to look through the book as they narrated the story.

The narratives were recorded and analyzed in terms of four fluency-related factors: speed (number of syllables produced in 3 minutes), pause profile (number of intra-clausal pauses per 100 words produced), repetitions (number of repetitions of words or parts of words per 100 words) and morphosyntactic accuracy (errors per 100 words). These four factors were chosen based on results from Experiment 1. It was these factors that correlated most strongly with the judges' holistic fluency rating in that experiment, and that the judges consistently identified as potential markers of fluency during their post-task interviews.

Each of these four factors was measured as they were in Experiment 1, with the following exceptions. The figures for intraclausal pauses, repetitions and morphosyntactic errors were measured in terms of occurrence per 100 words rather than per minute, as in the first experiment. It seemed better to give an indication of frequency per amount of language actually produced instead of per time period, especially in light of the large differences in the amount of language produced in the first experiment (and, as it turned out, in the second one). Also, when counting intraclausal pauses, the minimum amount of silent time was set at 400 msec. rather than 250 msec, as in Experiment 1. This change was made to accommodate the wider range of speaker abilities in the second experiment, which resulted in more pauses. The recordings of the participants' speech were digitalized and intraclausal pauses were measured using Audacity software (downloaded from http://audacity.sourceforge.net).

As in Experiment 1, morphosyntactic accuracy was assessed by two native Spanish speakers who reviewed transcripts of the participants' monologues and identified errors. As in the first experiment, they followed strict guidelines as to what constituted actual errors, as opposed to merely unusual or informal speech. Both individuals were linguistically sophisticated and had extensive language teaching experience. Their inter-rater reliability coefficient on this task was .91.

4.2.2.2 Word Translation Test. This task was designed to test participants' lexical retrieval efficiency by measuring how quickly and accurately they could name common objects in Spanish. The English names of 34 everyday items- "pencil," "book," "eye," etc.- were shown to the participants on a computer screen (see Appendix F). [27] Participants were instructed to say the Spanish equivalent aloud as quickly as they could think of it. Each word remained on the screen until the participant could produce its Spanish name or until 5 seconds elapsed, whichever happened first. A microphone was connected to the computer, and response times were recorded in milliseconds. Response times were measured as the time between the appearance of the English word on the screen and the onset of an acceptable Spanish equivalent spoken by the participant. Two measures were derived from this test: percentage of correct answers and mean response time for correct answers (wrong answers were excluded from the latter score).

The target words for this task were chosen from the text book used by first-year Spanish students at the University of Pittsburgh, *Mosaicos Spanish as a World Language* (Castells et al., 2002). All 34 words appear in the first three-quarters of the book, and all are presented as target words that students should learn. A Spanish teaching assistant who used *Mosaicos* in a first-year class reviewed the words for level appropriateness. The first-year text books at Carnegie Mellon and Carlow universities were also checked to verify that they also presented these common words. Most of the target words do not have common English cognates (an exception is *pants-pantalones*), so the participants were not usually able to access the L2 word indirectly through its L1 counterpart. In many cases, more than one translation was acceptable. For instance, in response to the target word *eyeglasses*, participants would get credit for saying either *gafas*, *lentes* or *anteojos*. Three practice trials were carried out in order to orient the participants to the task.

4.2.2.3 Imitation/Grammaticality Test. As mentioned above, Levelt (1989) suggests that the attentional aspect of working memory is important in monitoring one's own speech. Although it is difficult to measure a speaker's attention or linguistic monitoring, elicited imitation-and-correction tests of the kind presented here may provide a workable substitute. This task was designed to test participants' ability to simultaneously retain and review brief strings of L2 speech, and correct them on the spot if necessary.

The participants listened to recordings of 40 short two-person exchanges in Spanish (see Appendix G). They were instructed to imitate the second speaker's utterances and to correct them if they detected any errors. Half of these utterances contained simple morphological or grammatical errors (i.e., * *La mesa es blanco* "The table (fem.) is white (masc.)" and half did not. Participants were told that some of the utterances contained errors, and some did not. Two native Spanish speakers, a man and a woman, made the recordings under the researcher's supervision and using texts written by the researcher. Another native speaker and teacher of Spanish reviewed the text before it was recorded. On the recording, the native speakers spoke more slowly and clearly than they would in natural speech with other native speakers. There was one error per utterance in the 20 utterances that did contain errors, and those errors always occurred in the second speaker's (the woman's) utterance. The participants were told in advance that if there was an error, it would be in the woman's utterance, and that the man's utterance was provided merely for context. They did not have to repeat or correct the man's utterance; that was included in order to provide context and help orient the participants, as explained below.

To accommodate the first-year students who took part in this experiment, the 20 errors were simple in nature, and were based on the material they had covered in class. These included mistakes in gender and number agreement, verb tenses and the *ser/estar* distinction, among others. A native Spanish-speaking teaching assistant reviewed the errors for level of difficulty. Vocabulary items used in the exchanges were chosen from the *Mosaicos* textbook, the text used by first-year students at the University of Pittsburgh.

In elicited imitation tasks, target sentences are sometimes presented in isolation (Flynn, 1986). However, in this experiment the target sentences occurred within the context of brief exchanges. The first line (the man's utterance) was a question or comment that oriented the participants to what might come next, and give them a better chance of understanding it and repeating it. For example, in one exchange, the man asks, "Dónde está el profesor?" (*Where is the professor?*), and the woman replies "*Es en la clase" (*He's in class*). Here, the third-person verb *es* should be changed to *está*. Vinther (2002, p. 67) recommends this contextualized approach, suggesting that it may encourage participants to forget that they are taking part in an experiment and adopt the role of someone in a communicative situation.

A participant's score on this test was the number of grammatically correct sentences that he or she produced, either by imitating acceptable sentences or by correcting unacceptable ones.

Responses were marked as incorrect if they included verbatim repetitions of errors (i.e., the participant repeated *es* in the above example), or if they were incorrect or partial repetitions of originally correct utterances. Utterances that were modified slightly to an acceptable form without any loss of information (as when an optional pronoun was added or deleted) were counted as correct. As with the non-word repetition task, phonetic inaccuracies due to L2 accent were not counted as wrong responses.

4.2.3 Spanish Proficiency Test

All participants were given a Spanish Proficiency Test based on the New York State Regents Exam in Spanish. The Regents Exams are state-mandated comprehensive tests that high school students in New York State must pass before being allowed to graduate. The Spanish language test has been administered since the early 1900's, and it is also used to help place incoming students in Spanish courses in the state university system. Copies of previous tests, along with answer keys, are available on-line through the state library's web site (New York State Library, 2004). The exam is designed to be taken by high school students in their third year of Spanish studies, corresponding roughly to the second year of college-level study.

An abridged version of the June, 2003, Regents exam was given to the participants in this experiment (see Appendix H). This exam included sub-tests of three skills: listening, reading and writing. A speaking section (which is included in the original Regents exam) was not used, as the participants were being tested elsewhere on their Spanish speaking skills. The exam lasted approximately 45 minutes for each participant. The listening and reading sections used a multiple-choice format, and each contained 10 items. In the writing section, participants had to compose a short letter to a fictional exchange student from Spain. These letters were later reviewed and scored on a 15-point scale by two native Spanish speakers who were experienced in language teaching (inter-rater reliability was .90). The average of their two scores was used. Each participant received an overall test score based on results from the three sub-tests. The highest possible score was 35 points for the Spanish Proficiency Test.

Table 34: Test used in Experiment 2

Working memory tests	Speaking Span Test (SST) Math Span Test (MST) Non-Word Repetition Test (NWRT)
L2 (Spanish) oral fluency tests	Narrative Monologue Test (NMT) Word Translation Test (WTT) Imitation/Grammaticality Judgment Test (IGT)
L2 proficiency test	Spanish Proficiency Test (SPT)

Altogether, each of the 44 participants completed seven tests, as shown in Table 34. Each participant was tested individually, and the total amount of time needed for completing all seven tests and tasks was approximately two hours per participant. The testing procedure was divided into two sessions of roughly equal length in order to reduce fatigue and inattentiveness. Each participant took a mandatory break of about 15 minutes during the testing. A few participants elected to do half of the testing on one day and the other half on another day, in order to accommodate their personal schedules. Most completed both halves on the same day. Participants were compensated for taking part in the experiment.

4.3 HYPOTHESES

The seven tests produced a total of 13 scores per participant, as shown in Table 35. These included five working memory scores, seven L2 fluency scores and one L2 proficiency score.

Scores were correlated both within and between groups of variables. All five working memory scores were correlated with each other, and all seven L2 fluency scores were correlated with each other. This was done to verify whether the tests, particularly the working memory tests, were measuring skills or processes that were essentially similar to each other. It was important to determine whether in fact the Word Span Test and Math Span Test were measuring an independent quality that was largely independent of the type of stimuli used (verbal or numerical). Likewise with the fluency tests: if scores from the three tests did not correlate significantly, it might indicate that at least some of the tests did not in fact measure crucial aspects underlying oral fluency.

Table 35: Types of scores produced in Experiment 2

Test	***Name of score***	***Potential range of score***
***Working memory tests*:**		
Speaking Span Test	**SST span**	**0-6**
	SST total	**0-100**
Math Span Test	**MST span**	**0-6**
	MST total	**0-100**
Non-Word Repetition Test	**NWRT correct**	**0-40**
L2 fluency tests:		
Narrative Monologue Test	**NMT speed**	**NA**
	NMT pauses	**NA**
	NMT repetitions	**NA**
	NMT errors	**NA**
Word Translation Test	**WTT correct**	**0-34**
	WTT time	**NA**
Imitation/Grammaticality Test	**IGT correct**	**0-40**
Spanish test:		
Spanish Proficiency Test	**SPT score**	**0-35**

The crux of Experiment 2 was to be found in the correlations between working memory scores and L2 oral fluency scores. The main hypothesis was that there would be a strong correlation, indicating that perhaps working memory played a significant role in the acquisition or maintenance of foreign language fluency. Each of the five working memory scores was correlated with each of the seven fluency scores obtained in this experiment, producing 35 correlations. It was predicted that moderate to strong (.4 to .7) correlations were be found between many of these variables (see Hypothesis 3 below).

Finally, working memory and L2 fluency scores for the 20 least proficient Spanish learners were isolated from the data and correlated with each other. Members of this "bottom 20" group were selected based on their scores on the Spanish Proficiency Test. It was expected, based on Temple's (1997) model, that for lower-level L2 learners the relationship between working memory and fluency would be stronger than for more advanced learners.

The hypotheses for this experiment were as follows:

1. Scores from the three working memory tests will correlate significantly with each other. This would lend credence to the idea that working memory is a real and influential capacity, one that is largely (though not necessarily entirely) independent of topic or the material in which it is measured. It was predicted that the two span tests would correlate mostly strongly (.7 or higher), and that these would correlate significantly but less strongly (.4 or higher) with the NWRT, which measured only phonological loop capacity, not processing ability.
2. Scores from the three L2 oral fluency tests will correlate moderately significantly (.4 or higher) with each other. This prediction was based on correlations from Experiment 1, including the moderately strong correlations between temporal fluency (speed and smoothness) and morphosyntactic accuracy. It was predicted that these moderate-level correlations would extend to the two additional tests used here, the Word Translation Task and the Imitation/Grammaticality Test, which were designed to measure particular components of oral fluency, namely lexical fluency and monitoring skill.
3. Working memory scores will correlate with L2 oral fluency scores at a moderate-to-high (.4 to .7) level. If true, this would support the notion that working memory plays a role in the achievement and/or maintenance of L2 fluency. Correlations with

fluency scores will be strongest (.7 or higher) for the Word Span Test because the WST is a language-based measure of working memory capacity, and one that taxes both storage and processing capacity. But correlations produced by the Math Span Test (a non-linguistic measure) and the Non-Word Repetition Task (a storage-only measure) will also be significant (.4 or higher).

4. Correlations between working memory and L2 oral fluency will be stronger for low-level learners than for more advanced learners. Again, this idea derives from Temple's (1997) proposal that working memory plays an especially large role in early L2 learning.

4.4 RESULTS AND ANALYSIS

4.4.1 Working memory scores

Data for the three working memory tests, which produced five scores, are shown in Table 36. As seen there, the scores for the Math Span Test were generally higher than the scores for the Speaking Span Test, regardless of whether scores were quantified in terms of working memory span or total items. Five participants achieved the maximum score of 6.0 on the Math Span span score, while none scored higher than 4.5 on the Speaking Span span score. Eleven participants scored 90 or better on the Math Span total score but none of the 44 participants scored better than 82.5 on the Speaking Span total score.

This difference is interesting because many participants expressed trepidation before taking the Math Span Test, saying that they were "not good at numbers" or something to that effect. None made a similar remark about the Speaking Span Test. Afterward, during the post-task questioning, these participants were asked why they did better on the Math Span Test, contrary to their expectations. Several participants noted that in the Speaking Span Test they had to create original sentences, and they said this made it very difficult for them to later recall the target words. There was no parallel requirement in the Math Span Test. There was a processing component in the latter test- participants had to solve arithmetic problems- but this did not seem to tax their abilities as much as making up new sentences. Several participants mentioned this explicitly, saying that creating sentences involved “more thinking” than solving arithmetic problems or prevented them from developing a strategy for recalling the words.

Table 36: Working memory scores

	Speaking Span- span (0-6)	**Speaking Span- total (0-100)**	**Math Span- span (0-6)**	**Math Span- total (0-100)**	**Non-Word Repetition Test (0-40)**
Mean	2.99	66.40	3.93	81.21	30.59
S.D.	0.62	6.97	1.30	12.40	4.28
Range	2.0- 4.5	55.0- 82.5	2.0- 6.0	39.0- 98.0	19.0- 40.0

A couple of participants also noted that the exposure times were different in the two tests. In the Speaking Span Test the words were on the screen for only 1 second but in the Math Span Test the problems (and target digits) were on the screen for 5 seconds. “They (the words) came so fast,” said one participant. The difference in exposure time was due partly to precedence: the researcher wanted to keep the testing procedure consistent with those of earlier researchers in order to produce comparable results. Also, and more importantly, the extra exposure time on the Math Span Test was necessary to give participants a fair chance at solving the arithmetic problem. Unlike the Speaking Span Test, the “processing” part of the task here came immediately. It may be that the differences in overall scores on these two tests, then, are due at least in part to differences in presentation rather than differences in content.

Almost all of the participants reported using the same basic strategy for remembering the target words and digits while taking these two tests. That strategy consisted of repeating the items to be remembered sub-vocally until the time for recall. There was virtually no variation on this approach. One exception was a participant who tried to rely on free recall, at least on the math test. Another exception was a participant who said he attempted to organize the target digits into “blocks” of numbers that meant something to him personally. However, he said he was unable to do this consistently as the test went on and the sets of numbers grew longer. Another participant said that on the Speaking Span Test, she tried to group words together in images. For instance, in the first two-word set, “farmer/kitchen,” she tried to imagine farmers in a kitchen. But she too was forced to abandon that strategy as the sets of words grew longer. In the end, all or almost all participants seemed to employ sub-vocal repetition as their main strategy in these tasks.

On the Non-Word Repetition Test, not surprisingly, most participants performed better in the earlier part of the test, which used shorter words. The longer, four- and five-syllable words proved more difficult to reproduce accurately in their entirety. However, unlike the two span tests, this test produced little obvious frustration in participants, probably because it required little mental processing.

Correlations between the five working memory scores are shown in Table 37. The two Speaking Span Test scores correlate strongly with each other, as do the two Math Span Test scores. This is to be expected, as the two types of score are different quantifications of working memory based on the same test. The correlation between the two Speaking Span Test scores (almost .75) is higher than the correlation between the two Math Span Test scores (around .68) but the reason for this is not clear.

Table 37: Correlations between WM scores

	SS (span)	**SS (total)**	**MS (span)**	**MS (total)**	**NWRT**
SS (span)		.749**	.265	.285	.172
SS (total)			.365*	.382*	.243
MS (span)				.683**	.225
MS (total)					.476**
NWRT					

* $p < .05$, ** $p < .01$

There is a moderately strong correlation (around .48) between the Non-Word Repetition Task scores and the Math Span Test total scores. However, three of the four scores produced by the two memory span tests failed to correlate significantly with NWRT scores. Overall, these results do not bear out the prediction made in the latter part of Hypothesis 1, namely that the scores from the two memory span tests would correlate moderately with those from the NWRT. This prediction was based on the theoretical assumption that the NWRT measured phonological short-term memory (PSTM), and that PSTM is a key component of WM. Although these results cannot by themselves be considered conclusive evidence against such a conceptualization of WM, they certainly do not support it. As Juffs (2004, pp. 206-07) notes, it is by no means clear whether PSTM is best understood as one component of a faculty that also contains a storage-and-processing module in the form of the Central Executive.

The presence of one significant correlation between the MST and the NWRT, and the lack of any such correlation between the SST and the NWRT, is noteworthy. It may be that this difference is due partly to the relative lack of semantic processing required in the NWRT and the MST. In the former test, participants were simply repeating sounds that had no semantic content (at least to these non-Arabic speakers). In the former, they were remembering digits. These digits were, of course, expressed in English and had some semantic content to the participants. But it seems fair to say that this semantic content was relatively light compared to the larger set of nouns used in the SST. That is, the "light-content" digits may have had more in common with the "no-content" Arabic words than with the "high-content" English nouns, verbs, adjectives and adverbs used in the SST. Another similarity was hinted at above: the MST required little in the way of procedural processing and interference compared to the SST, where participants had to create original sentences. In this way, it may have been more like the NWRT, where no mental procedures other than short-term memorization were carried out.

However, the most striking result here is the lack of strong correlations between scores from the two memory span tests. These correlations ranged from around .26 to about .38, and even the strongest of these correlations were only weakly significant. This is well below the .7 level of correlation predicted in Hypothesis 1. More generally, this result is troubling for the viewpoint that working memory is a real and independent faculty, one that works more or less equally well regardless of the sort of material at hand. If that were the case, one could have expected much stronger relationships here. It may be that the absence of strong correlations is partly an artifact of the testing procedures, and specifically of the methodological differences described above. However, as far as this evidence goes, it does not support the first part of Hypothesis 1.

4.4.2 Fluency-related factors

The seven fluency-related variables are shown in Table 38. As the figures there indicate, there were extremely large ranges between the participants in terms of the speed with which they spoke in the Narrative Monologue Test. One produced just over 50 syllables in 3 minutes, another produced more than eight times as many in the same amount of time. There were enormous differences in the number of pauses, repetitions and morphosyntactic errors produced by the participants as well. Large individual differences were also found in the results from the two other tests, the Word Translation Test and the Imitation/Grammaticality Test. These differences are consistent with the results obtained in Experiment 1, and they undoubtedly reflect, at least in part, the large differences in proficiency and level of Spanish study found among the participants. Whether they also reflect differences in working memory capacity remains to be seen.

Table 38: Scores for fluency-related variables

	NMT speed	**NMT pauses**	**NMT repetitions**	**NMT errors**	**WTT**	**WTT time**	**IGT correct**
Mean	177.0	20.36	2.98	17.03	22.0	1.696	21.60
S.D.	90.13	10.29	3.69	7.30	5.71	0.3271	7.94
Range	53.0-444.0	2.8- 52.8	0-19.1	6.40- 50.0	10.0-32.0	1.059-2.552	5.0- 37.0

Correlations between these seven variables are shown in Table 39. Generally speaking, these correlations are stronger than those between the working memory scores. Among the four types of scores produced by the Narrative Monologue Task, there are several moderately strong correlations in the .4 to .7 range. Not surprisingly, sheer speed (NMT speed) and rate of intraclausal pausing (NMT pause) correlated negatively and significantly. There are moderate correlation between morphosyntactic accuracy (NMT errors) on one hand and the speed and pause scores on the other. That is, participants who spoke with relatively high formal accuracy also tended to speak more quickly and with relatively few pauses. These figures reinforce the results of Experiment 1: it appears that formal accuracy progresses hand-in-hand with temporal fluency. The “odd man out” among the NMT variables is the repetitions factor, which did not correlate significantly with any of the other six fluency-related factors. It should be noted that the number of repetitions produced by a speaker was not found to be one of the most significant variables in previous studies of temporal fluency (Freed, 1995; Kinkade, 1995; Riggenbach, 1991; Lennon, 1990). In view of that fact, and its isolated status in Experiment 2, it seems likely that the results from Experiment 1 concerning repetitions were an aberration.

Table 39: Correlations between fluency variables

	NMT speed	**NMT pause**	**NMT repetitions**	**NMT errors**	**WTT correct**	**WTT time**	**IGT correct**
NMT speed		-.683**	-.043	-.418**	.614**	-.529**	.807**
NMT pause			.027	.452**	-.498**	.369*	-.572**
NMT repetitions				-.087	.110	-.001	-.049
NMT errors					-.173	.047	-.508**
WTT correct						-.575**	.494**
WTT time							-.523**
IGT correct							

Results from the other two fluency tests, the Word Translation Test and the Imitation/Grammaticality Test, also produced moderate-to-strong correlations. In the WTT, the number of correct translations correlated well (nearly - .58) with the speed of correct translations. That is, participants who translated more accurately from English to Spanish generally did so more quickly as well (it should be kept in mind that WTT time represents average time for correct translations only). The IGT produced significant correlations with every other variable except the "odd man out," the repetitions variable. It correlated especially strongly (over .8) with the speed variable. This suggests that the IGT taxed at least some of the same faculties that the Narrative Monologue Test and the Word Translation Test taxed.

One pair of weak correlation is between morphosyntactic accuracy (NMT errors) and the two scores produced by the Word Translation Test (WTT correct and WTT time). These correlations were -.173 and .047 respectively. It would seem that the ability to access L2 words quickly would be a key component of maintaining formal accuracy during actual speech (note that NMT speed factor correlates significantly with the two WTT scores). However, it may be that accessing words in isolation, as during the translation task, calls on different skills than accessing appropriate word forms and syntactic patterns during discourse. This opens the possibility that more than one kind of working memory is employed in language production, perhaps as suggested by the multi-component models of Caplan and Waters (1999, 1996) and Roberts and Gibson (2002). Clearly, more research would be needed to corroborate such conclusions.

The overall picture from the working memory tests upholds Hypothesis 2. There are consistently significant correlations between most of the fluency-related variables used in this experiment. Indeed, if one excludes the repetitions data, 13 of the 15 correlations between these factors reach the level of statistical significance.

4.4.3 Correlations between working memory and fluency variables

The correlations between the two classes of variables, working memory scores and fluency scores, are shown in Table 40. These correlations are weak across the board. Only three out of 35 pairings of variables produced a statistically significant correlation, and all three of those correlations are within the .33 to .36 range. Neither of the two scores produced by the Math Span Test correlated significantly with any of the seven fluency variables, and neither did the Non-Word Repetition Test score.

Table 40: Correlations between WM scores and fluency scores

	SST span	SST total	MST span	MST total	NWRT
NMT speed	.249	.340*	.038	.146	-.035
NMT pause	-.098	-.160	-.057	-.212	.001
NMT reps.	-.179	-.191	-.152	-.119	-.128
NMT error	-.142	-.206	-.147	-.230	-.182
WTT correct	.056	.037	-.209	-.097	-.227
WTT time	-.210	-.224	.098	.143	.098
IGT correct	.353*	.331*	.012	.106	.082

* $p < .05$

These data do not support Hypothesis 3, the central hypothesis of this experiment. There is in fact no consistent pattern of significant correlations between working memory capacity scores and L2 oral fluency scores. None of the 35 correlations produced here reach even the lowest level of significance (.4) predicted in Hypothesis 3. It was also predicted there that the Speaking Span Test would produce the strongest (.7 or higher) correlations with fluency scores. In fact, the SST produced only the three comparatively weak correlations mentioned above, and two of these correlations occurred with the same test, the IGT.

The failure of the Non-Word Repetition Test to produce significant correlations in this experiment is noteworthy in lights of several studies that connect phonological memory capacity to L2 learning (Williams & Lovatt, 2003; Ellis & Sinclair, 1996; Service, 1992; Papagno, Valentine & Baddeley, 1991). Of course, the fluency tests used here measured performance rather than learning. But if performance depends largely on learning ability and if the results of those previous studies are sound, it seems that the NWRT would have produced higher correlations. Interestingly, the NWRT scores did not correlate even moderately strongly with either of the Word Translation Test scores, WTT correct or WTT speed (in fact, the NWRT-WTT correct correlation is negative). Both the NWRT and the WTT tests focus on words in isolation and issues of syntax and larger discourse patterns are irrelevant. The main difference between the two tests is that the NWRT tests purely phonological memory for words in an unknown language, while the WTT tests memory and translation speed for learned L2 words.

Finally, correlations were drawn between scores on the Spanish Proficiency Test on one hand and scores on both the working memory and fluency tests on the other. As mentioned above, the proficiency test measured listening, reading and writing skill in Spanish. None of the five working memory scores correlated significantly with the proficiency test scores (Table 41). In fact, Non-Word Repetition Test scores produced a negative correlation. By contrast, four of the seven L2 fluency scores produced significant correlations- four of six if one excludes the aberrant repetitions factor (Table 42). (The negative correlations between SPT scores and the NMT pauses, NMT errors and WTT time variables are in the expected direction; more fluent speakers are expected to pause less, make fewer morphosyntactic errors and take less time to translate words). This pattern of generally significant correlations is unsurprising. After all, the proficiency test and the various fluency tests were designed to quantify L2 skill, albeit in different modes. However, the lack of significant correlations between proficiency scores and working memory scores would be surprising to anyone who considered working memory capacity a good predictor of L2 skill.

Table 41: Correlations between Spanish Proficiency Test (SPT) scores and WM scores

	SST span	**SST total**	**MST span**	**MST total**	**NWRT**
SPT	.269	.257	-.113	.048	-.164

Table 42: Correlations between Spanish Proficiency Test (SPT) scores and L2 fluency scores.

	NMT speed	**NMT pause**	**NMT repetitions**	**NMT errors**	**WTT correct**	**WTT time**	**IGT correct**
SPT	.559**	-.253	.033	-.270	.455**	-.372*	.631**

* $p < .05$, ** $p < .01$

4.4.4 L2 proficiency and the working memory-fluency relationship

So far, the results of Experiment 2 are in line with Fortkamp's (1999) finding that working memory is not closely related to L2 oral fluency. However, that study focused solely on advanced learners. Less advanced learners were included in the present study, and it may be that a closer look at the data will reveal that for, these learners at least, the predicted correlations were in fact present. If so, this would be in line with Temple's (1997) proposal that WM plays an especially important role in early L2 learning.

To investigate this possibility, tests results for the 20 least proficient participants were separated for analysis. This subgroup was chosen based on their scores on the Spanish Proficiency Test, which ranged from 14.0 to 26.0. As Table 43shows, this group also performed below the whole-group norm on five of the seven fluency-related variables tested in this experiment. As a group, they spoke Spanish more slowly, with slightly more pauses and errors in the Narrative Monologue Test. They did, however, perform as well as the group in terms of repetitions (which, as has been seen, appears to have been an aberrant factor) and, more surprisingly, did slightly better in the WTT correct category.

Table 43: Fluency scores for 20 least proficient participants

	Mean (S.D.)	*Mean for all participants (n =44)*
NMT speed	132.45 (48.84)	177.0
NMT pauses	22.03 (9.28)	20.36
NMT repetitions	2.98 (4.34)	2.98
NMT errors	19.10 (8.59)	17.03
WTT correct	22.60 (7.01)	22.0
WTT time	1.783 (0.3438)	1.696
IGT correct	17.20 (5.78)	21.60

It should be noted that as a group this “bottom 20” group also scored lower on most measures of working memory capacity, though the differences are small. The relevant figures are shown in Table 44. (Note that the “bottom 20” actually scored slightly higher on the NWRT test.) This raises the question of whether their relatively poor performance on the Spanish Proficiency Test, as well as the fluency tests, were due in some degree to sub-par working memory capacity.

Table 44: Working memory scores for 20 least proficient participants

	Mean (S.D.)	***Mean for all participants***
SST span	2.8 (.70)	2.99
SST total	64.13 (6.43)	66.40
MST span	3.90 (1.37)	3.93
MST total	79.95 (15.37)	81.21
NWRT correct	31.30 (4.39)	30.59

To test this idea, correlations between all five working memory scores and all seven fluency scores were correlated for this subgroup, just as they had been done for the entire set of participants. The results are shown in Table 45. Only four of the 35 variable pairings produced that reached the level of statistical significance, and three of these correlations involved one test, the SST. Both types of MST score (span and total) and the NWRT scores correlated negatively with sheer speed (NMT speed), as measured in syllables produced in three minutes of L2 speech, and also with translation accuracy (WTT correct). Taken together, these correlations are only slightly stronger than the working memory-fluency correlations produced by all 44 participants. They do not support Hypothesis 4, which predicted an especially strong correlation between working memory capacity and fluency among low-level learners.

Table 45: Correlations between WM scores and L2 fluency scores for 20 least proficient participants

	SST span	SST total	MST span	MST total	NWRT
NMT speed	.501 *	.465 *	-.082	-.161	-.054
NMT pauses	-.054	-.042	.062	.010	.221
NMT reps.	-.130	-.225	-.171	.084	.011
NMT errors	-.021	-.091	-.148	-.153	-.079
WTT correct	-.006	-.041	-.262	-.404	-.367
WTT time	-.131	-.222	.304	.484 *	.048
IGT correct	.462 *	.340	-.200	-.127	.000

4.5 DISCUSSION

The central hypothesis of this experiment, that significant correlations would be found between working memory capacity and L2 oral fluency, was not confirmed. Neither were two other key hypotheses: that the various working memory measurements employed here would correlate significantly with each other, and that for the lowest-level learners, the correlation between working memory and fluency would be greater than for the group as a whole. The only prediction that was borne out was Hypothesis 2, which stated that the various fluency measurements used in this study would correlate significantly with each other.

Why did the hypothesized relationship between working memory and fluency not materialize? This negative result must be explained in the context of previous research which implicates working memory capacity, particularly phonological short-term memory, in the acquisition of L2 vocabulary and syntax (see Ellis, 1996, for a review). Four possible explanations are provided here. These deal with the complex nature of speaking fluently in a foreign language, which may call on faculties other than working memory, the persistence of working memory's influence over the course of language learning, personal and affective factors and, finally, the nature of working memory itself.

4.5.1 The nature of oral fluency

The sort of L2 performance that participants in this study were called upon to produce was different from those in previous studies. Much of that older research focused on particular parts of language learning, notably vocabulary, with little reference to productive language use. Papagno, Valentine and Baddeley (1991) connected PSTM to long-term L2 vocabulary learning, but their measures of vocabulary learning did not include spontaneous, productive use of words and phrases. Gathercole and Baddeley's (1990) study of children indicated a relationship between children's PSTM capacity and their skill at learning new words. But there again, the linguistic skill under study was the ability to learn new words in isolation. Service's (1992) study did find significant correlations between PSTM capacity and long-term L2 learning outcomes in her study of Finnish children studying English as a foreign language. Service also found more specific correlations between PSTM capacity and specific English language skills, including listening comprehension and written production. But there was no L2 speech component in her experiment.

Even if, as these studies indicate, PSTM aids in the learning of new vocabulary words, it does not necessarily follow that it also helps learners retrieve and use words in actual speech performance- which is certainly one of the hallmarks of language proficiency. Phonological encoding skill does not guarantee lexical retrieval skill, and in a complex activity such as L2 speech there may be other variables that are more important. Craik and Lockhart (1972) proposed that the "depth" of encoding- including encoding at the semantic level as well as the phonological level- may be crucial to the retention, and presumably the retrieval, of information. Their proposal is relevant here. A word that is learned in multiple contexts or processed at a "deeper" semantic level may produce richer semantic associations in a learner's mind than one that is learned in isolation or in a single context. This rich network of associations could make retrieval easier during spontaneous speech, and may be more important than the initial phonetic encoding of the word.

It is worth noting again that the Word Translation Task used in this experiment did not produce significant correlations with the three working memory measures, including PSTM as measured by the Non-Word Repetition Task. The Word Translation Task was not a learning test; it was a test of how well participants could retrieve common L2 words that they had already encountered. Furthermore, the WTT did not call on participants to use L2 words productively, merely to recall them in response to an L1 equivalent. If this test of retrieval of isolated common words is not significantly affected by PSTM capacity, it seems unlikely that lexical retrieval in the course of L2 speech performance would be any more dependent on PSTM.

Of course, speaking fluently involves more than retrieving words. It also requires the rapid and accurate use of syntactic patterns, and the research linking working memory to syntactic learning must be acknowledged. But it must also be pointed out that this research is fairly limited in scope. Three studies are of particular importance. One is Daneman and Case's (1981) examination of morphosyntactic learning, which is cited by Ellis (1996, p. 106) as evidence of the working memory-syntax connection. But in that study the target forms were limited to a few short words and sequences of new language. The same is true of the other two studies, Ellis and Sinclair (1996) and Williams and Lovatt (2003). In the former, the researchers found significant correlations between working memory capacity and the ability to learn Welsh phrases. But, as noted in the literature review above, the "syntax" involved in this study was limited to two simple phrases and one morphosyntactic rule. In the latter study, morphosyntactic learning and testing focused on a small part of Italian morphosyntax related to gender and determiner selection. This is a far cry from the sort of complex application of syntax that L2 speakers must perform in naturalistic communication. These three studies did not make a convincing case for working memory's role in the long-term acquisition of, much less the productive use of, L2 syntax.

4.5.2 The influence of working memory over time

A second explanation focuses on the persistence of working memory's influence on L2 learning. It is certainly conceivable that working memory, whether presented as PSTM or in the more complex storage-and-processing form measured in span tests, is a component of language aptitude. This is precisely what much of the research cited in Chapter 2 indicates. Furthermore, Temple (1997) presents a solid theoretical argument for the influence of working memory in the early stage of L2 learning. But one must ask just how early this "early" stage is, and how long it lasts. The results produced by the "bottom third" of the participants in this experiment suggest that it may not last long at all. Most of these low-proficiency participants were first-year Spanish learners with only one full semester of study behind them. Yet even within this group the correlations between working memory capacity in any form and L2 fluency were generally weak.

One plausible explanation is that even for these novice learners other factors had already surpassed working memory in influence, at least as far as oral production is concerned. These factors might include other aspects of language aptitude such as language analytical ability, one of the components of the MLAT test (Carroll & Sapon, 1959). They might also include personal study habits and the quantity and quality of L2 speaking practice that learners engage in. Indeed, according to Stevick (1976, p. 60), interaction in the target language probably affects oral fluency more than other aspects of L2 performance. That would help explain why working memory played a larger role in a small-scale learning experiment like Ellis and Sinclair's (1996) than they did here, where learners had to apply knowledge they had acquired over a period of months or years. In a very short laboratory experiment, where testing follows learning almost immediately, working memory may be more influential simply because these other variables have not had time to make their presence felt.

Another possibility is that working memory is most helpful to L2 learners at later stages of acquisition. This runs counter to Temple's (1997) model, and appears to contradict the results of Ellis and Sinclair's (1996) and Williams and Lovatt's (2003) studies of working memory and short-term L2 learning. On the other hand, it is consistent with findings reported by Mackey et al. (2002), who speculate that high working memory capacity may improve L2 acquisition over the long term. Based on a study of interaction and noticing involving both low- and high-capacity ESL learners, the researchers speculated that their high-capacity participants "had gleaned more data to process and consolidated this over time, compared to low WM capacity learners who could not 'hold on' to data with great accuracy" (Mackey et al., 2002, p. 204). In their view, working memory may have contributed to superior learning, but only after enough learning had occurred to produce an organized, schematic L2 knowledge system.

However, there is a problem in attributing the cumulative effects of long-term L2 learning to variations in working memory capacity. This is especially true if the short- or medium-term benefits of high working memory capacity are not clearly established- and they were not clearly established in either the present study or in Mackey et al.'s (2002) study. A plausible alternate explanation is that superior analytical skills produce better organized and more complete L2 schema over the long term. Better schema, whether they are lexical, syntactic or pragmatic in nature, make L2 knowledge more accessible and promote automatic retrieval and fluency. They may also promote noticing of new L2 input by providing ready-made categories or networks of L2 items that learners can relate novel items to. This scenario is more compatible with the "long-term working memory" model advocated by Ericsson and Kintsch (1995) than with the "working memory as aptitude" view of Miyake and Friedman (1998), Ellis (1996) and others.

It is also possible that this latter scenario describes phonological L2 learning as well as lexical, syntactic or pragmatic learning. According to Adams and Gathercole (1995), native English speakers display a higher memory span for English-like non-words than non-words that are phonologically un-English-like. As Baddeley (1996, p. 22) states, this makes clear that "even the phonological loop is not a *tabula rasa*, but, rather, a system that has developed on the basis of the phonological experience of the rememberer." Presumably, a learner's "schema" of L2 phonology is organized differently than his or her lexical/syntactic/pragmatic knowledge, simply because it is devoid of semantic content. Still, the main point remains valid: even at this level, previously acquired knowledge affects working memory capacity. [28]

Aptitude for language learning is not a monolithic construct. Not only does aptitude contain multiple components, but these components may have more or less influence at different stages of acquisition and performance. This is an issue that calls for further study. But a logical starting point for investigation would be the following hypotheses. The first is that as a learner's knowledge of the target language increases, analytical abilities, that is, the ability to make connections between what has already been mastered, become more important while memory-based aptitude (including perhaps working memory capacity), or the ability to acquire more target language forms, recedes in influence. The second hypothesis is that as proficiency increases, the overall effect of aptitude decreases compared to other factors such as practice time, motivation and quality of instruction. These latter factors, especially the latter two, are difficult to accurately quantify, but that does not mean that their influence is any less significant.

4.5.3 Personal and Affective Factors

Another explanation for the absence of the hypothesized correlations has to do with personal and affective factors, including the effects of anxiety. For language learners, situational anxiety is often produced by the need to communicate in an L2, the desire to make positive impressions on interlocutors and, in the classroom, by testing situations (Brown, 1994, p. 142). Presumably, it is also produced in laboratory studies like the present experiment which call upon learners to communicate in an as-yet unmastered language. Research indicates that anxiety plays a detrimental role in the efficiency of working memory during language performance (Miyake & Shah, 1999b, p. 469). Many of the participants in this study expressed doubts about their ability to speak in Spanish for any length of time, and many were obviously anxious about their Spanish speaking abilities. It is possible that this anxiety negatively affected the participants' working memory efficiency and/or fluency during the tests conducted for this experiment. This may in turn have skewed the results and disguised a relationship between working memory and fluency that would have become apparent under less stressful circumstances.

So far, this explanation is not very convincing. The fact is that anxiety is likely to be present in many situations where language learners must speak spontaneously in their target language. Indeed, real-life communicative situations would produce considerably more anxiety in learners than having to narrate a children's story in an experimental situation where they were not being graded (and being paid for taking part). If anxiety was a significant factor here, it would very probably be even more so under more naturalistic circumstances.

On the other hand, other personal factors, some of which are plausibly related to anxiety, may be important. After the testing procedures were complete, half a dozen of the 44 participants were invited back to talk to the researcher at length about fluency in general and about their experiences in the narrative monologue task used in this experiment. Among other things, they were asked whether they thought consciously about maintaining fluency while they were speaking Spanish. They were also asked what problems they faced in trying to speak fluently in Spanish. Although this sampling of participants was not large, it nonetheless provided some insights that were not obvious from the quantitative data alone.

What was most noticeable was that, even within this small group, there were differences in how the participants approached L2 speaking tasks, and how much emphasis they placed on maintaining fluency. Five of the six replied that they were conscious of their fluency, or lack of fluency, while they were speaking Spanish, and four said they often became anxious or nervous because of this. But they differed noticeably in how much weight they placed on keeping up a steady stream of L2 speech and in how they handled potential breakdowns in fluency.

For example, one participant, a 26-year-old female, said she made a conscious effort to avoid using “ah” or other filled pauses because she believed listeners reacted negatively to them. This participant, who scored very high on all the fluency variables during the testing procedure, added that she kept speaking even when she couldn’t come up with the exact Spanish word she wanted, or when she made minor errors. “I think what I try to do is just keep going,” she said. This participant said that when talking to native Spanish speakers, she often relied on their willingness to overlook inaccuracies. Several other interviewees made similar statements.

However, another participant, a 22-year-old male who described himself as a very highly motivated (10 out of 10 on the questionnaire) Spanish learner, struck a different tone. Although he acknowledged that speaking Spanish sometimes made him nervous, he did not display much outward anxiety during the narrative task. At one point during the post-task interview, this participant said that he was determined not to fill his Spanish speech with meaningless chatter just to keep a conversation going. Using words that echoed Fillmore (1979/2000) and Lennon (2000), he criticized some other students in his Spanish class who created “an illusion of fluency” while contributing nothing of substance or repeating themselves. He also expressed a desire to avoid overly general words and to be as precise as possible even when a desired L2 word did not come to him immediately.

The point is that at least some learners make conscious decisions about fluency and their personal styles of L2 speech, decisions that may mitigate or outweigh factors such as working memory. It has been proposed (Faerch & Kasper, 1983; Seliger, 1980) that in learners can be characterized as “planners” or “correctors” in terms of L2 speech. “Planners” pause more frequently and use these pauses to plan their next utterance, while “correctors” are more likely to proceed without a fully formed speech plan and rely on their ability to change and correct their output. Presumably, there are many learners who fit somewhere between these two poles. Exactly where an individual fits may be rooted in his or her anxiety level and willingness to let their listeners endure pauses or stretches of slowed-down speech. One learner may see L2 interactions as a grueling task and wish to save face by attempting to maintain fluency at the cost of content. Another may view L2 interactions as an opportunity to exercise his or her L2 knowledge to the fullest, regardless of slight discomfort or inconveniences to their interlocutors. Such personal preferences may exist completely independently of language aptitude.

4.5.4 The Nature of Working Memory

Finally, the absence of significant correlations between the two working memory span tests deserves attention. As noted in the literature review, there is a lively debate about the nature of working memory, and whether it operates independently of the type of material used and of previous learning. Ericsson and others (Ericsson & Delaney, 1999; Ericsson, 1996; Schneider, 1999) contend that it is not an independent faculty at all, that “working memory” is highly dependent on content and on long-term knowledge stores. The weakness of the correlations between these two span tests seems to support this view. Working memory as measured linguistically in the Speaking Span Test did not match up particularly well with working memory as measured in the Math Span Test, regardless of whether one looks at span scores or total scores. This disconnect may be an artifact of the methodological differences in the two tests, differences that were discussed above. But another possibility is that working memory capacity depends to a large extent on the kind of material being stored and manipulated.

If this is so, it would explain easily why Math Span Test scores did not correlate strongly with L2 fluency scores. Numerical knowledge, it can be argued, is quite different from linguistic knowledge. The weaker-than-expected correlations between the Speaking Span Test scores and the various fluency measures are more difficult to explain- until one recalls that the SST was administered in English, the participants’ native language. The participants were not given an

SST in Spanish to avoid confounding L2 proficiency with L2 working memory. Also, it was felt, that a Spanish-language SST would be too difficult for the novice Spanish learners among the participants. If, as Fortkamp (1999) concluded, there is little relation between working memory measured in L1 and working memory measured in L2, one would not necessarily expect strong correlations between L1 verbal working memory and L2 oral fluency. In fact, no such correlations were found in the main study here.

It could be argued that content-dependence and language-dependence is reconcilable with Baddeley's (2000) revised model of working memory. This revised model, it will be recalled, contains an "episodic buffer" that coordinates information flow between the two slave systems and long-term memory stores. Previous knowledge of language and/or specific languages might then influence the efficiency of working memory. However, this arrangement is virtually indistinguishable from the "long-term working memory" scenario, and it will be unsatisfactory to anyone who wishes to preserve the integrity of working memory as a viable construct, one that plays an important and distinct role in language learning and production.

An alternative view holds that the efficiency of working memory is largely determined by background knowledge and familiarity with the matter under study- in short, with proficiency. This view is consistent with studies of expertise in non-linguistic domains such as chess. DeGroot's (1965) study of master and non-master chess players indicated that the masters were superior at assessing novel positions within seconds of seeing them. He attributed this to their ability to quickly create "experiential linkings," that is, to relate the new positions to thousands of similar positions they had encountered before.

> If it is correct that the most important difference between master and non-master lies in the differentiation and scope of the system of experiential linkings and if consequently the master can start thinking from a higher level, then this class difference should come out clearly in the first minutes, nay seconds, of the perceptual and thought process….(T)he master should distinguish himself during *the perceptual process* [italics DeGroot's] (DeGroot, 1965, pp. 321-22).

DeGroot's view was supported by subsequent studies of memory for chess positions. For example, Chase and Simon (1973) found that masters were better than non-masters at reconstructing positions they were allowed to view for 5 seconds. But this advantage existed only when the positions were like those they had seen in actual games. When pieces were placed randomly on the board, masters performed no better than novices. Studies of other domains of expertise, including electronic circuitry, computer programming and radiology have produced parallel findings (Committee on Developments in the Science of Learning, 2000, p. 33).

These non-linguistic studies argue for the importance of long-term knowledge stores in developing mastery, and against assigning too large a role to short-term memory skills. In essence, the participants in these studies remembered well because their memory stores were in some way superior to those of participants who performed at lower levels. The present study indicates that these same arguments can be applied to second language acquisition. This is not to deny that short-term memory skills have some impact on individual differences in L2 learning. Rather it is to suggest that this impact is comparatively weak in the long run, at least as oral fluency is concerned.

5.0 SUMMARY AND CONCLUSIONS

It is hoped that this project will make some useful contributions to the study of L2 oral fluency and to how working memory may influence language performance. Some of these possible contributions are summarized below. This section concludes with some observations on the limits of this project and some directions for further research.

5.1 THE STUDY OF L2 ORAL FLUENCY

The first experiment in this project produced a relatively small set of quantifiable factors that can be used to measure and assess speech production. These are speed of delivery, rate of intraclausal pausing, morphosyntactic accuracy rate and, perhaps, repetition rate. These were chosen for use in the main experiment based on their high correlations with holistic fluency ratings by native speakers of the target language in the pilot study. This pilot study provided several other insights into fluency and how it is perceived. For instance, it was found that speakers' interclausal pause rate was not as significant in listeners' assessment of their fluency as intraclausal pause rate. This is line with studies of L1 speech (Goldman-Eisler, 1968; Clark & Clark, 1976) which found that interclausal pausing was part of the normal plan-and-execute pattern of speech production, and did not necessarily detract from fluency.

It is not claimed that the four-factor set of factors used here is the only grouping of variables that can describe fluency. Indeed, inclusion of the repetition factor seems problematic, and this factor might well be dropped in future studies. But this set of factors does have two advantages. Firstly, it is small and manageable enough to use in quantitative research on speech production. It is much less unwieldy than the large inventories of factors that were measured in Experiment 1 here, or used in other exploratory studies, notably Lennon's (1990). Secondly, and more importantly, it is based on empirical research and may be an improvement on previous studies of fluency which relied on instinctive conceptualizations of fluency, generally as total word production plus "richness" of expression, a rather subjective quality.

The first experiment also underlined the importance of morphosyntactic accuracy in fluency assessment. Inclusion of an accuracy variable among purely temporal factors in fluency assessment may be controversial to experts who prefer to view fluency and overall speaking proficiency as strictly separate entities. But, judging from both the quantitative and qualitative results of Experiment 1, it seems unlikely that naive listeners make the same separation. Rather, it seems more likely that temporal fluency and accuracy are closely intertwined and that learners' progress in both areas goes hand-in-hand.

In Experiment 2, the narrative task was supplemented by two other measures of oral fluency, the Word Translation Test and the Imitation/Grammaticality Test. The former was a simple lexical translation task and unremarkable in its design, but it should be noted that it had a time-pressure element to it. Participants had to come up with a suitable translation within 5 seconds. This was done in order to test not only vocabulary knowledge but also the ability to access and produce known L2 vocabulary words quickly, as in spontaneous speech. Some previous research into working memory and vocabulary acquisition (Ellis & Sinclair, 1996; Service, 1992) apparently did not contain this kind of time-pressure factor.

The Imitation/Grammaticality Test used in Experiment 2 combined two sorts of tests used previously in linguistic research, sentence repetition and grammaticality judgments. The aim was to create a test that mimicked a speaker's ability to monitor his or her own internal speech. This is obviously important to the production of accurate, comprehensible L2 output. Internal speech is largely inaccessible, of course, but this test may have provided a reliable and realistic substitute, one that could be used in future studies.

5.2 WORKING MEMORY AND LANGUAGE

The study of working memory and its effects on language learning and use is by now a well-established subfield of psycholinguistics, one that has been worked on vigorously for a quarter of a century. Nevertheless, there are areas that have been received less attention than others, and the second experiment in this project was devoted to one such area, namely L2 speech. It differed from several previous studies (e.g., Williams & Lovatt, 2003; Ellis & Sinclair, 1996) in that it looked at the potential effects of working memory capacity on people who were learning a language in a classroom setting. They did not learn the target forms in a laboratory shortly before being tested on their retention of those items. This was a disadvantage in one respect because there was a high level of "contamination" in this experiment compared to the laboratory studies. The participants in Experiment 2, even the novices who had only one full semester of Spanish study behind them, had surely been influenced by many factors besides working memory, some related to their personal experiences or study habits. The results of the experiment must be considered in that light.

But, from another perspective, this contamination of variables was an advantage in that it simulated real-life learning situations. Even if working memory has a positive influence on initial learning, one must investigate how far this influence extends and when other variables become more important. The indication here is that over the long term the influence of individual differences in working memory capacity is negligible. Working memory capacity may help in the immediate acquisition of new knowledge, but the results of this study and other studies call into question whether it helps maintain, consolidate and make accessible that knowledge.

This study also differs from many previous studies of working memory and language in its focus and in its findings. The focus here was on L2 oral performance and in that sense it was unlike most previous psycholinguistic studies of working memory capacity, which focused on reading. All of these studies, including Daneman and Carpenter (1980), King and Just (1991) and Miyake, Just and Carpenter (1994), reported significant correlations between working memory span and measures of (L1) reading skill. Only Fortkamp (1999) looked in depth at working memory and L2 speaking, and her conclusions were broadly consistent with those of the present study. It may be that the purported advantages of a large working memory span show up more clearly in reading than in speaking, though why this should be is unclear. Indeed, as mentioned in Chapter 1, the rapid, spontaneous nature of oral communication would seem to place a premium on working memory.

The findings of this study appear to conflict with those of Service's (1992) study of Finnish children learning English. Service found strong correlations between phonological short-term memory capacity and L2 acquisition (see Table 2-6). Unlike Ellis and Sinclair's (1996) study but like the present study, it measured the results of long-term classroom learning rather than short-term laboratory learning. Two facts might help explain why the present findings are different from Service's. One has already been mentioned: although the Finnish children were tested on their L2 listening, reading and writing skills, they were not tested on their L2 speech skills, the main focus of this study. Secondly, it should be remembered that Service tested the children on their ability to repeat English-like words, words like "rendance" and "disajoinance" that were phonologically similar to words of their target language. It may be that what Service actually measured in her repetition test was, to some extent, the children's existing knowledge of English phonology, whether or not the words were real. Her test differed from the Non-Word Repetition Task used here in this important aspect. In the NWRT employed in Experiment 2, the target forms were real (Arabic) words but were completely unfamiliar phonologically and morphologically to the participants, none of whom had studied Arabic.

The results of this study are in accord with a relatively small but growing body of research that throws some doubt on the importance of working memory capacity in second language acquisition. This includes Juffs' (2003, 2004) study of working memory and comprehension of hard-to-parse sentences and Michael and MacWhinney's (2003) examination of working memory, attention and early vocabulary learning. Juffs found that WM as measured by a Reading Span Test was not a good predictor of ESL learners' ability to process English garden path sentences. His findings suggest that WM does not play a role in split second, on-line parsing decisions. Although Juffs was looking at L2 comprehension, it seems likely that the same or similar on-line processing abilities are equally crucial (perhaps even more crucial) to the rapid and fluent production of L2 speech. [29] Michael and MacWhinney (2003) found that WM was not as good a predictor of vocabulary acquisition as attention, as measured by a Stroop task. These studies present a challenge to the working memory-as-aptitude view espoused by Ellis (1996), Miyake and Friedman (1998), Service (1992) and others.

5.3 LIMITATIONS OF THESE STUDIES

Although this project attempted to be comprehensive, it inevitably had limitations, and these should be acknowledged. For example, the process of speech production undoubtedly contains a preverbal stage, a time when intended messages exist only as ideas to which words and syntactic structures have not yet been assigned. Levelt (1989) terms this stage "conceptualization." It may well be that some individuals exhibit more “ideational fluency,” and are capable of producing or connecting these preverbal ideas faster than other people. This may in turn influence how quickly and/or how well they produce linguistic messages, that is, their oral fluency. However, it is only these audible messages that researchers can record and analyze; what precedes them is simply inaccessible to researchers given present methods, and perhaps to the speakers themselves.

Also, this project examined working memory as a purely cognitive faculty, and it looked at fluency in the semi-artificial context of monologic speech. The researcher deliberately set up these restrictions in order to eliminate the complications that arise from interaction with other speakers. To this end, the researcher avoided introducing topics or themes that could arouse strong feelings in the participants; the main stimulus for eliciting L2 speech was a children's picture book. This procedure did not necessarily reflect natural speech behavior, where people often talk about what interests them most, and where their like or dislike for a topic, or their degree of familiarity with it, affects their production.

Most speech acts involve an interlocutor who offers questions and comments, gives feedback through linguistic and non-linguistic channels and is involved in the give-and-take of turn-taking. This was not the case in these experiments. Rather, the participants' speech samples were limited to narrative monologues and responses to recorded items. This is an important caveat, as interactional competence may be considered part of overall speech proficiency. It involves the ability to attend to what one’s interlocutor is saying and to “latch onto” themes he or she introduces, to use listening time to plan upcoming utterances and to maintain coherence in one’s speech. These are all important conversational skills that could affect fluency in conversational settings.

5.4 IDEAS FOR FURTHER RESEARCH

The first suggestion for future research into working memory and language is purely methodological in nature. It is to make the various span tests more uniform procedurally. In Experiment 2, there were differences between the Speaking Span Test and the Math Span Test in how the material was presented and recalled. In the SST, participants were asked to recall target words and produce new sentences after viewing the entire set of words. In the MST, on the other hand, participants had to first solve arithmetic problems and later, after all the problems had been viewed, recall the target digits. This difference could conceivably have affected the participants' memory abilities and made comparisons between the two tests problematic. One way to avoid this problem would be to have participants in the SST make new sentences immediately and recall the target words later.

Future researchers might also do well to focus more closely on the suppressive effects of attention, as distinct from working memory span. This was the approach of Michael and MacWhinney (2003), who found that suppressive ability as measured by a Stroop Test correlated better with vocabulary learning skill than working memory span. As presented in the Baddeley model of working memory, the central executive is responsible for inhibitory control of irrelevant material, as well as the manipulation of items stored in the slave systems and (with the episodic buffer added) for coordinating information flow with long-term memory stores. This is an impressive list of duties, and one may ask whether this sort of all-encompassing faculty can be fruitfully analyzed as a whole. It is not at all clear that standard span tests such as the Speaking Span Test or the Math Span Test provide measures of attentional control. If they do not, other measures such as the Stroop Test or a language-based equivalent of it may be useful.

In his study of chess players, DeGroot (1965) made extensive use of think-aloud protocols, and this may be a good approach for future work in speech production and memory. DeGroot had players explain why they were considering, and then making, certain moves as a way of getting at their on-line thought processes. In language studies, this sort of protocol would involve recording learners producing L2 speech and then immediately playing the recording back to them. The learners would comment (in their L1) on their L2 speech while it was fresh in their minds. Researchers could prod them to describe what they were thinking at particular moments, for example, moments of either disfluency. This approach might reveal the conscious use of fluency-enhancing strategies of the kind described by Dechert (1983) and Faerch and Kasper (1983). It could also indicate if and how learners use prefabricated “chunks” of L2 speech to maintain fluency (Pawley & Syder, 2000; Oppenheim, 2000). This sort of information is not readily available from quantitative studies alone.

This approach may also help language educators understand how they can help learners improve their L2 oral fluency. The study of how good language learners get around lexical deficits and other obstacles to fluency, how they employ creative repetition and expansion (as in Ejzenberg, 2000), how they pace their plan-and-execute cycles may provide ideas that can be taught to other learners. If fluency can be taught, as Guillot (1999) suggests, this sort of analysis of speech performance will undoubtedly be of great value.

NOTES

1 Miyake and Shah's "consensus" definition of working memory (p. 450) is reproduced here in full:

> Working memory is those mechanisms or processes that are involved in the control, regulation, and active maintenance of task-relevant information in the service of complex cognition, including novel as well as familiar, skilled tasks. It consists of a set of processes and mechanisms and is not a fixed "place" or "box" in the cognitive architecture. It is not a completely unitary system in the sense that it involves multiple representational codes and/or different subsystems. Its capacity limits reflect multiple factors and may even be an emergent property of the multiple processes and mechanisms involved. Working memory is closely linked to LTM, and its contents consist primarily of currently activated LTM representations, but can also extend to LTM memory representations that are closely linked to activated retrieval cues and, hence, can be quickly reactivated.

Although this definition is broad enough to include Baddeley's conception of working memory, its mention of "currently activated LTM representations" is more in line with Ericcson and Kintsch's (1995) view of "long-term working memory."

2 "...we can at present see no reason for taking a strong view on whether the central executive will ultimately prove to be a system within which a range of equally important control processes interact in a quasiautonomous way, with overall control forming an emergent feature, or whether there is a hierarchy of such processes with one dominant controller. In short, we leave open to investigation the question of whether the central executive resembles an organization run by a single chairperson or one governed by the collective wisdom of a committee of equals" (Baddeley & Logie, 1999, p. 40).

3 Baddeley (2000, p. 420) cites the case of a densely amnesiac patient who reportedly continued to play contract bridge well despite his memory deficit. The patient was able to keep track of the "contract," as well as which cards had already been played. Baddeley comments: "Once again, we appear to have evidence for a temporary store that is capable of holding complex information, manipulating it and utilizing it over a time scale far beyond the assumed capacity of the slave systems of WM."

4 In the original (Baddeley-Hitch, 1974) model, the central executive did have a storage function. Baddeley later eliminated this function, partly in order to avoid mimicking the storage functions of the slave systems and long-term memory. Also, he felt it made the central executive too powerful to investigate fruitfully (Baddeley & Logie, 1999, pp. 37-38).

5 One important and unresolved question is whether this episodic buffer has any inherent capacity limits. Without some testable limit, it is difficult to see how this buffer is significantly differently from long-term memory. An unlimited episodic buffer would make Baddeley's model more like connectionist models or Ericsson's long-term working memory model, where working memory is a subset of long-term memory.

6 Terminology can be confusing in regard to the unitary/non-unitary nature of working memory, as Miyake & Shah (1999a, p. 3, footnote) note. Baddeley has called his model the "multiple-component model" (Baddeley & Logie, 1999). However, a similar model adopted by Daneman, Carpenter and Just, which contains separate components for visuo-spatial and verbal processing, is sometimes referred to as a "single-resource" or "single pool" model. The relevant point here is that in terms of its verbal processing alone, this model is more of a single-resource or unitary model than the model advocated by Caplan and Waters.

7 The relatively weak correlations between scores on the word span test and the reading skill scores is interesting in light of subsequent claims that phonological short-term memory (PSTM) is heavily implicated in language skills (i.e., Ellis & Sinclair, 1996). In Daneman and Carpenter's experiment, the words in the word span test were presented orally (rather than visually as in the RST), so this test would seem to be a pretty good measure of PSTM.

8 A similar criticism could be leveled at the RST itself. In the longer sets, where subjects have to read five or six sentences, it seems that more than 2 seconds will pass between the time when they read the final words of some sentences and the time when they have to recall them (this observation came from Alan Juffs). In this case, it could be argued that working memory is still involved because subjects are presumably engaging in subvocal rehearsal to try to retain the words in working memory. But it is possible that at least some subjects do not, and prefer instead to rely on free recall. Subjects in these span tests are not directed to use subvocal rehearsal, and it cannot be assumed that all of them do.

9 Here is one reading passage, from Daneman & Carpenter (1980, p. 455), with the referent and its pronoun separated by five full intervening sentences (italics added):

> Sitting with Richie, Archie, Walter and the rest of my gang in the Grill yesterday, I began to feel uneasy. Robbie had put a dime in the juke box. It was blaring one of the latest "Rock and Roll" favorites. I was studying, in horror, the reactions of my friends to the music. I was especially perturbed by the expression on my best friend's face. *Wayne* looked intense and was pounding the table furiously to the beat. Now, I like most of the things other teenage boys like. I like girls with soft blonde hair, girls with dark curly hair, in fact all girls. I like milkshakes, football

games and beach parties. I like denim jeans, fancy T-shirts and sneakers. It is not that I dislike rock music but I think it is supposed to be fun and not taken too seriously. And here *he* was, "all shook up" and serious over the crazy music.

After introducing four other possible referents into the discourse, it would be unusual in everyday speech to expect a listener to be able to connect the last-named referent with a pronoun that appears several utterances later.

10 Daneman and Green report within the text (p. 14) that reaction times differed according to speaking span scores. Low-span speakers took an average of 2464 msec. to produce synonyms for the target words, intermediate-span speakers took an average of 2354 msec., and high-span subjects took an average of 1516 msec. However, it's difficult to know how much weight to attach to these figures because the authors do not report how many subjects fell into each of these three categories, nor what criteria were used to classify subjects as low-, intermediate- or high-span. There were 34 subjects in the experiment, so if Daneman and Green divided them into roughly equal groups, there would have been 11 or 12 subjects in each category.

11 In the Speaking Span Test administered in the main experimenter in this project, subjects were told that they should use the exact form of the word if they remember it. In trials of the SST, before this was done, it was clear that some subjects changed some word forms consciously, apparently to make sentences more easily, even though they recalled the target form correctly.

12 Daneman mentions in the text (p. 457) that subjects with low speaking spans produced on average 115 words per minute on the SGT, while intermediate-span speakers produced 137 words and high-span speakers produced an average of 155 words. A similar pattern held for the "richness" scale: low-span subjects averaged 2.1 out of a possible 5.0, intermediate-span subjects scored 2.5 on average and high-span subjects averaged 2.9. But, as in Daneman and Green (1986), there is no information on what criteria was used to categorize subjects as low-, intermediate- or high-span. Neither does the reader know how many subjects were placed in each category. As there were 29 subjects altogether, it seems unlikely that any of the three groups was very large.

13 The table includes information from both Fortkamp's Table III (p. 279), where the word count data from the SGT is presented, and from a portion of the text (pp. 285-86) where data from the "richness" scores are discussed. Throughout her analysis, Fortkamp emphasizes the word count as the main measure of fluency, evidently because it is more objective than the richness scale. She does not provide correlations between the richness scores and the Portuguese SST scores. Fortkamp mentions in passing (p. 286) that the two fluency measures (word count and richness) correlated with each other significantly ($r = 0.54$, $p < .0305$).

14 Levelt leaves open the question of exactly how small the portion of the discourse record stored in working memory is. "An interesting but little-studied issue is which aspects of discourse lead to deep encoding (i.e., to long-term storage) and which aspects are transient (i.e., kept in working memory for only short periods of time)." (p. 111). Since working memory is

severely limited relative to longer-term storage, and it must also store products of the speaker's own output, this part of the discourse record is presumably very small.

15 Levelt qualifies this strict division between controlled and automatic processing in speech production: "There may be marginal forms of executive control..." (p. 22). This could account for the fact that speakers can quickly stop talking when they detect errors, even in the middle of a syallable, he says.

16 This is a necessarily simplified overview of deBot's revised model. DeBot proposes that the Conceptualizer is partly language-specific, *contra* Levelt, and that the Articulator is non-language specific. Also, deBot's view of the lexicon is more complex than this description suggests. He believes that while both L1 and L2 items co-exist in one lexicon, the lexicon contains different "subsets" of items that may overlap to varying degrees, depending on how closely related the two languages are in general, whether two particular words are cognates and the speaker's degree of L2 proficiency.

17 Temple compared the speech of 11 intermediate-to-advanced students of French with the speech of native French speakers. She found that native speakers averaged 4.21 syllables/sec. while the learners averaged 2.34 syllables/sec. Also, he native speakers produced 5.5 pauses (pause = silence of .15 seconds or more) and 1.0 self-repairs, while the learners averaged 15.8 pauses and 2.2 self-repairs (p. 83).

18 Leeson (1975) provides what might be called a generative conceptualization of fluency. For him, a fluent speaker is one who can produce an infinite number of accurate sentences in a language "on the basis of a finite exposure to the finite corpus of that language" (p. 136). This view is notable for its emphasis on the acquisition process and its suggestion of an ability to abstract linguistic principles from little input. Of course, it is also an idealized and untestable definition, one that says nothing about temporal qualities of speech such as speed and hesitations or, for that matter, semantic content. Brumfit (1984/2000) criticizes Leeson's view as overly theoretical and detached from actual learning processes.

19 Of course, this is not meant to dismiss or diminish the importance or research on Universal Grammar and second language acquisition conducted by Lydia White (1989) and others. The point here is that this research is by its nature focused on universal properties of the mind and linguistic competence, not on individual differences or linguistic performance.

20 Tarone et al. mention (1983, pp. 8-9) use of "prefabricated patterns" as an L2 communicative strategy. However, they mean something different from Fillmore's "formulaic expressions" in that the speaker uses these prefabricated patterns as unanalyzed wholes, without understanding their underlying syntactic structure. Tarone et al. cite the English "do-you" question form as one such pattern, adding that ESL learners may use it to produce incorrect utterances such as "What do you doing?" Use of these patterns may increase temporal fluency but also introduce distracting errors into the discourse. This example shows the difficulty of addressing L2 fluency without some reference to morphosyntactic accuracy.

21 Whether conscious use of these "performance features" constitutes a kind of "communicative strategy" in Corder's sense of the latter term is open to debate. In some cases, these performance features do not in themselves convey meaning by alternate means. Rather, they buy time until the preferred means become available, or they serve to reformulate communication in a more native-like manner (in the case of pausing and particularly location of pauses). This sort of "fluency-enhancing strategy" is dealt with here under the heading of "communicative strategy" to avoid multiplication of labels.

22 Some critics may object that such a "strategic" approach to fluency consists of nothing more than a bag of time-gaining conversational ploys or "tricks," and that it is better to spend classroom time on improving formal competence. Guillot (1999, p. 43) quotes an anonymous French instructor who opined that "you don't need to teach students to '*heu*' and '*eh ben*,' they manage it quite well on their own." True enough, many students will pick up common L2 speech devices, and some will over-use them. But the point is to encourage students to use these and other devices well and sparingly in order to come closer to native-like patterns of fluency.

The kind of fluency-conscious instruction that is advocated here and by Guillot may seem difficult to implement. However, Guillot provides a host of creative approaches and exercises, most aimed at higher-level learners. Some of these involve careful listening and analysis of native speech with an ear toward noting fluency-building devices. These devices include creative repetitions and "intensifying paraphrases" (i.e., "*c'est vraiment leur inquietude/c'est effectivement leur plus grosse inquietude*..."), and latching on to phrases just used by one's interlocutor (p. 75). They also include skillful use of "formulaic platforms" and common adverbial qualifiers (i.e., *peut-etre, quelquefois, de temps en temps*, etc.) (p. 79). (As this partial list shows, there is more than adding an occasional "*heu*" and "*eh ben*") One exercise focuses on encouraging students to use these devices "to create thinking space and support the construction of utterance" (p. 117). The instructor gives students a one-line prompt such as this mock news bulletin: "*Le Front de liberation des nains de jardins a encore frappe*." In pairs, students continue orally, composing the bulletin without advance preparation while using the fluency-building devices they have learned from close listening of native French speech. Guillot recommends a post-task analysis and repetitions of the task in order to give students the chance to think about their performance and how to improve it.

The specifics of constructing fluency-conscious L2 speaking curricula go beyond the scope of the present work, but two points should be made. Firstly, in writing instruction, it is common to study the use of coherency-building devices and common formulas (i.e., "Dear Sir/Madam..." and "Sincerely, ..." in formal letter writing). It is assumed that instructors will know these features and be able to teach them. As Guillot (p. 96) asks, "why devote time to the study of written texts, and confine the spoken word to practice, when both the written and the oral medium are at least equally relevant to students' concerns...?"

Secondly, the idea of explicitly teaching L2 oral fluency is hardly new, and this part of the curriculum need not be built from scratch. In the ESL field, for instance, the textbook *Strategies for Speaking* (Rost, 1998) adopts a consciously pro-fluency attitude. It includes formulaic expressions and functional gambits that are associated with a variety of communicative situations, as well as "fluency units" every four or five chapters which give learners the opportunity to practice the use of these devices in communicative situations. Another textbook, *Talk It Over!* (Kozyrev, 2002, pp. 10-13), encourages students to focus on and imitate native-like pause patterns, emphasizing the inter-clausal location of most pauses (or between "thought groups," as the author puts it). An earlier textbook, *Building Fluency in English: Authentic Speech* (Hieke & Dunbar, 1985) is outstanding in its treatment of performance features such as pauses, cohesive devices and use of collocations. It integrates close listening of native English discourse (on cassettes, with accompanying text) with exercises for productive practice. Hieke and Dunbar's text provides a good model for the development of future fluency-focused materials.

23 Kinkade also used three native Chinese-speaking judges in order to determine how non-native English speakers judged English L2 fluency. Those ratings are not reported here because the present focus is on how native speakers assess fluency.

24 Levelt (1989, p. 111, see also footnote 13 above) assigns to working memory the task of keeping track of the immediately past discourse record, before parts of it are transferred to long-term memory. Here is another area where working memory capacity may be important to oral fluency, albeit indirectly. Speakers not only have to produce grammatically acceptable utterances but make sure that these are context-appropriate in light of what one's interlocutor just said. This function of working memory will not be addressed here as it involves comprehension more than production, but it should be kept in mind if one considers fluency to include more than merely temporal phenomena.

25 Here is how one learner in Oppenheim's study, a 37-year-old Korean woman, used a recurrent sequence (p. 230):

a. "My weakest processing mechanism is linguistic."
b. "That is my strongest processing mechanism."
c. "Friends say that is my strongest processing mechanism is linguistic."

The fact that the phrase *processing mechanism* is not an English idiom or common collocation is probably unimportant from a psycholinguistic perspective. It could still function as one of the woman's "islands of reliability" (Dechert, 1983) by allowing her to execute a phrase automatically while formulating the next phrase. The third, qualitative experiment in the present project is aimed partly at identifying these non-obvious, idiosyncratic chunks of L2 speech.

26 The terms "automaticity" and "fluency" are used interchangeably here; the latter may be considered a particular (verbal) form of the former. But it is interesting to note that in the literature on non-linguistic automaticity, the matter of accuracy is rarely addressed. The unwritten assumption seems to be that automatized behavior is correct behavior (i.e., the right musical note has been played, the right button or switch pushed). Accuracy and its relationship to temporal fluency is more of an issue in the literature on linguistic fluency. No doubt this is due largely to the complex nature of language, and the difficulty of identifying "right" and "wrong" behavior in many circumstances. However, it does not follow that the question of accuracy is completely separable from the matter of fluency. The relationship between the two is one of the main areas of investigation in the first study reported here.

27 There were actually 40 items in the Word Translation Test. However, 6 of these items were retroactively removed and not counted in the scoring data, leaving 34 items in the data analysis. It was pointed out that some of the low-level learners who took part in Experiment 2 may not have been exposed enough to the Spanish equivalents of these 6 words to be tested on them. Thanks to Nuria Sagarra for pointing out this fact..

28 While acknowledging that previous knowledge affects phonological loop capacity, Baddeley (1996, p. 22) rejects the notion that short-term memory is simply the activated portion of long-term memory. Such as a view is “so general as to be theoretically sterile, unless an attempt is made to specify in detail the processes involved,” he concludes. It is undoubtedly true that specifying the relationship between long-term memory stores and short-term or working memory is complex, and does not lend itself to close scrutiny. But of course this in itself is no reason to reject the “long-term working memory” model.

29 Interestingly, Juffs’ study suggests that simple word span rather than WM may play a role in these on-line parsing decisions. He found that low-word span participants took more time to process garden path sentences than high-word span learners.

APPENDIX A

POST-TASK QUESTIONS FOR RATERS IN EXPERIMENT 1

(The researcher asked the first 4 questions orally, then handed the paper to the raters so they could mark and/or write their answers to Question 5.)

1. When you were listening to these speakers, was there anything specific about anyone's way of speaking that made you think, "That's a fluent speaker?"

2. Was there anything specific that made you think, "That's not a fluent speaker?"

3. What does the term "fluency" mean to you as applied to spoken language?

4. Does it mean the same thing when you're listening to a non-native speaker of English as it does when you're listening to a native speaker?

5. Here is a list of ways to describe spoken speech. Please choose three (3) factors that are most important in rating a speaker's fluency. (Please choose only three.)

    ___ how well they pronounce words

    ___ how fast they speak

    ___ how complex their sentences are

    ___ how many times they hesitate

    ___ how long they hesitate

    ___ how many grammatical mistakes they make

___ other (specify here) ____

APPENDIX B

QUESTIONNAIRE FOR PARTICIPANTS IN EXPERIMENT 2

1. Name:

2. Level of Spanish class (1^{st} year, 2^{nd} year, etc.):

3. Age:

4. Sex (M/F):

5. What is your native language(s)?

6. Do you (or did you) speak Spanish at home with your parents, siblings or other people? If so, with whom and how often?

7. Have you lived in a Spanish-speaking place for any length of time (not counting short vacations)? If so, where and how long?

8. How would you rate your own ability to speak Spanish? (very poor, poor, fair, good, excellent)

9. What grades do (did) you usually receive in Spanish? (A, B, C, D, F):

10. Did you study Spanish before college? If so, where and for how long?

11. Have you studied other foreign languages? If so, which ones and for how long?

12. How would you rate your motivation for studying Spanish, compared to other students, on a 1-to-10 scale (1 is hardly motivated, 10 is very highly motivated)?

APPENDIX C

WORDS USED IN THE SPEAKING SPAN TEST

<u>Trial sets</u>
a) pumpkin, balance
b) fingers, noticed
c) machine, results, compass

<u>2-word sets</u>
a) kitchen, farmers
b) signals, thirsty
c) perfume, giraffe
d) healthy, rewards
e) biscuit, shampoo

<u>3-word sets</u>
a) pencils, observe, journey
b) nervous, quickly, younger
c) trumpet, windows, believe
d) earning, dentist, tallest
e) parking, succeed, whisper

<u>4-word sets</u>
a) butcher, wrinkle, ceiling, glasses
b) certain, warning, mittens, husband
c) diapers, special, instant, plastic
d) explain, stylish, garbage, request
e) trouble, bending, advance, roasted

<u>5-word sets</u>
a) teacher, stomach, foreign, cousins, quarter
b) jealous, monthly, arrange, sweater, treated
c) growing, surfing, ashamed, lettuce, cushion
d) damaged, respect, private, clearly, witness

e) useless, helping, passive, buttons, tonight

<u>6-word sets</u>
a) student, careful, reduced, vandals, orchard, ignored
b) morning, village, traffic, islands, handles, patient
c) chimney, achieve, cookies, explode, feather, address
d) knuckle, chicken, working, storage, injured, playful
e) lawyers, mailbox, freezer, release, lightly, fragile

APPENDIX D

ARITHMETIC PROBLEMS USED IN THE MATH SPAN TEST

<u>Trial sets</u>
P1. 6 - 4 = ?, 9 + 2 = ?
P2. 7 + 1 = ?, 8 - 5 = ?
P3. 2 + 3 = ?, 5 - 4 = ?
P4. 7 - 2 = ?, 8 + 6 = ?
P5. 5 + 4 = ?, 6 - 3 = ?, 4 + 7 = ?

<u>Two-problem sets</u>
a. 5 + 3 = ?, 4 - 1 =?
b.7 - 2 = ?, 6 + 5 =?
c. 2 + 6 = ?, 5 - 4 = ?
d. 3 - 1 = ?, 2 + 3 = ?
e. 7 + 8 = ?, 7 - 5 = ?

<u>Three-problem sets</u>
a. 9 + 3 = ?, 5 - 4 = ?, 7 + 2 = ?
b. 7 - 1 = ?, 8 + 2 = ?, 2 + 7 = ?
c. 2 + 6 = ?, 9 - 3 = ?, 6 + 4 = ?
d. 4 - 2 = ?, 5 + 1 = ?, 8 - 7 = ?
e. 8 - 3 = ?, 6 + 2 = ?, 4 + 9 = ?

<u>Four-problem sets</u>
a. 5 - 2 = ?, 4 + 4 = ?, 9 + 3 = ?, 7 - 5 = ?
b. 3 + 1 = ?, 8 + 3 = ?, 5 - 2 = ?, 6 - 4 = ?
c. 7 + 1 = ?, 4 - 3 = ?, 6 + 2 = ?, 8 + 1 = ?
d. 7 -3 = ?, 5 + 5 = ?,6 - 2 = ?, 3 - 1 = ?
e. 9 + 7 = ?, 5 - 3 = ?, 2 + 2 = ?, 6 - 3 = ?

<u>Five-problem sets</u>
a. 7 + 2 = ?, 4 + 9 = ?, 8 - 3 = ?, 1 + 6 = ?, 9 - 7 = ?
b. 6 - 3 = ?, 5 - 4 = ?, 7 + 3 = ?, 8 + 5 = ?, 6 - 2 = ?

c. 1 + 2 = ?, 9 - 3 = ?, 5 + 7 = ?, 6 - 4 = ?, 2 + 5 = ?
d. 3 + 3 = ?, 6 - 5 = ?, 7 - 2 = ?, 4 + 3 = ?, 1 + 5 = ?
e. 4 + 7 = ?, 8 - 6 = ?, 2 + 3 = ?, 8 + 9 = ?, 9 - 5 = ?

<u>Six-problem sets</u>
a. 9 + 3 = ?, 6 - 2 = ?, 4 + 5 = ?, 3 + 7 = ?, 5 - 2 = ?, 3 + 4 = ?
b. 1 + 7 = ?, 3 + 5 = ?,6 - 2 = ?, 7 + 4 = ?, 4 - 2 = ?, 3 + 8 = ?
c. 5 + 6 = ?, 2 - 1 = ?, 7 + 5 = ?, 4 - 3 = ?, 9 + 6 = ?, 2 + 9 = ?
d. 8 + 2 = ?, 8 - 6 = ?, 4 + 7 = ?, 7 - 3 = ?, 6 + 5 = ?, 9 - 4 = ?
e. 2 + 3 = ?, 8 - 4 = ?, 7 - 2 = ?, 4 + 7 = ?, 5 + 4 = ?, 9 - 1 = ?

APPENDIX E

ARABIC WORDS USED IN THE NON-WORD REPETITION TASK

<u>2-syllable words</u>
habba
kitton
hamma
thaljun
sa’ra
a’da
ma’la
abi
ummun
wardun

<u>3-syllable words</u>
wasala
alima
wathaba
kala’mun
zahratun
hari’kun
kataba
russumun
amta’run
buyu’tun

<u>4-syllable words</u>
tahadatta
yuma’riso
tarannama
yuja’milo
sayyaratun
darrajatun

a’amidatun
tassa’ala
awssimatun
yurassilu

5-syllable words

mutafa’ilun	
mutasha’imum	
yataallamo	muhandissu’na
yutassa’kato	mutaja’nissun
yuma’rissuna	yataa’tafo
assa’tidatun	yahtassibu’n

APPENDIX F

DIRECTIONS FOR NARRATIVE MONOLOGUE TEST

1. I will show you a children's picture book and ask you to narrate it in Spanish. The book is about a big dog named Carl who plays with a group of children. There are words on the first page, but there are no words on the rest of the pages. You will try to tell a story about the pictures in your own words, in Spanish.

2. Try to say something about every picture in the book. You don't have to go into detail about what you see. If you can't think of a word that you're looking for, try another word or try describing something else in the picture.

3. Speak as quickly but also as accurately as you can. Telling a story in a foreign language is very hard. Do your best!

4. You can browse through the book for a minute before we start, just to familiarize yourself with the contents.

APPENDIX G

ENGLISH WORDS USED IN THE WORD TRANSLATION TEST

<u>Trial items</u>

1. Man
2. Fork
3. Teacher

<u>Test items</u>

1. Door
2. Book
3. Dog
4. Hand
5. Pants
6. Eye glasses
7. Pencil
8. Bed
9. Clock
10. Spoon

11. Boat
12. Sun
13. Window
14. Apple
15. Table
16. Coat
17. Socks
18. Belt
19. Bread
20. Fish

21. Tree
22. Hat
23. Finger *
24. Foot *
25. Glove
26. Pen
27. Ball
28. Airplane
29. Umbrella
30. Drinking glass

31. Suitcase
32. Leg *
33. Chair
34. House
35. Nose *
36. Shirt
37. Eye *
38. Desk
39. Shoe
40. Ear *

* Participants' answers for these items were not considered in the data analysis.

APPENDIX H

IMITATION/GRAMMATICALITY TEST

(Ungrammatical forms are marked with an asterisk.)

Practice sentences:
P1. ¿Quién es Marta?
Es la hermana de Pedro.
Who is Marta?
She's Pedro's sister.

* P2. ¿Cuántas clases tomas?
Tomas cuatro clases.
How many classes are you taking?
You're taking four classes. (second-person verb form)

P3. ¿A qué hora es la clase?
Es a las ocho y media.
What time is the class?
It's at eight-thirty.

* P4. ¿Dónde está el profesor?
Es en la clase. (*ser* instead of *estar*)
Where's the professor?
He's in the class.

*P5. ¿De qué color es la computadora?
La computadora es negro. (masc. adj. w/ fem. noun)
What color is the computer?
The computer is black.

P6. ¿Sabes tocar la guitarra?
Sí, sé tocar la guitarra.
Do you know how to play the guitar?

Yes, I know how to play the guitar.

Test items

1. ¿Dónde está mi libro nuevo?
Está sobre el escritorio.
Where is my new book?
It's on the desk.

* 2. Hola, Verónica. ¿Adónde vas?
<u>Vas</u> a la biblioteca. (second-person verb)
Hi, Veronica. Where are you going?
I'm going to the library.

3. En general, ¿eres optimista o pesimista?
Soy muy optimista.
In general, are you an optimist or a pessimist?
I'm very optimistic.

4. ¿Cómo es la madre de Jorge?
Ella es una mujer elegante y generosa.
What is Jorge's mother like?
She's an elegant and generous woman.

* 5. Allí hay un restaurante. ¿Quieres cenar?
Sí, <u>soy</u> hambre. ("be" instead of "have")
There's a restaurant. Do you want to have dinner?
Yes, I'm hungry.

* 6. Tu hermana es bonita. ¿Es estudiante?
No, <u>trabajamos</u> en un restaurante. (first-person plural verb form)
Your sister is pretty. Is she a student?
No, she works in a restaurant.

7. ¿A qué hora comienza la clase?
La clase comienza a la una y media.
At what time does the class begin?
The class begins at 1:30.

* 8. Jennifer es muy inteligente, ¿no?
Sí, es una mujer <u>listo</u>. (masc. verb form for fem. noun)
Jennifer is very intelligent, isn't she?
Yes, she is a smart woman.

9. ¿Cuánto cuesta esta camisa?
Cuesta veinticinco dólares.

How much does this shirt cost?
It costs 25 dollars.

10. Tu amiga estudia español también, ¿verdad?
No, ella no estudia español, ella estudia italiano.
Your friend is studying Spanish too, right?
No, she isn't studying Spanish, she's studying Italian.

*11. ¿Dónde está Madrid?
Madrid <u>es</u> en el centro de España. (*ser* instead of *estar)*
Where is Madrid?
Madrid is in the center of Spain.

* 12. ¿A qué hora sale usted?
Hoy, voy a <u>salgo</u> a las tres y media. (requires infinitive form)
What time do you leave?
Today I'm going to leave at 3:30.

* 13. La camisa es bonita. ¿Tu lo compraste?
No, no compré<u>la</u>. (D.O in wrong position)
Did you buy it?
No, I didn't buy it.

14. ¿Cómo está tu padre?
Está muy bien, gracias.
How's your father?
He's very good, thanks.

* 15. ¿Qué estás haciendo?
<u>Soy</u> estudiando español. (*ser* instead of *estar*)
What are you doing?
I am studying Spanish.

16. ¿Comes en la cafeteria o en tu casa?
Generalmente, como en mi casa.
Do you eat in the cafeteria or in your house?
Generally, I eat in my house.

*17. ¿Por qué compraste un diccionario nuevo? Ya tienes un diccionario excelente.
Yo lo compré <u>por</u> mi amigo. (*por* instead of *para*)
Why did you buy a new dictionary?
You already have an excellent dictionary.
I bought it for my friend.

18. ¿Qué te gusta hacer los sábados?

Me gusta bailar y visitar a mis amigos.
What do you like to do on Saturday?
I like to dance and visit my friends.

* 19. ¿Cuántos hermanos tiene usted?
Yo <u>tiene</u> tres hermanos. (third-person verb form)
How many brothers do you have?
I have three brothers.

20. ¿Jorge tiene el pelo corto o largo?
Tiene el pelo muy largo.
Does Jorge have short or long hair?
He has very long hair.

* 21. Este libro es interesante. ¿Vas a comprarlo?
No, yo <u>tengo no</u> dinero. (incorrect negative form)
This book is interesting. Are you going to buy it?
No, I don't have any money.

22. ¿Es casado tu hermano?
No, él es soltero.
Is your brother married?
No, he is a bachelor.

* 23. Esta camisa es muy cara.
Sí, pero los zapatos son <u>barato</u>. (lack of number agreement)
This shirt is very expensive.
Yes, but the shoes are cheap.

24. ¿De qué color es el autobús?
El autobús es blanco y negro.
What color is the bus?
The bus is black and white.

* 25. ¿A qué hora te despiertas?
Yo __ despierto a las siete. (reflexive *me* absent)
What time do you wake up?
I wake up at 7.

*26. ¿Quiénes son Madonna y Britney Spears?
Son cantantes muy <u>famosa</u>. (lack of number agreement)
Who are Madonna and Britney Spears?
They are very famous singers.

27. ¿Quién es más alto, Jorge o Tomás?
Jorge es más alto que Tomás.

Who's taller, Jorge or Tomas?
Jorge is taller than Tomas.

28. ¿De dónde es Isabel?
Ella es de Perú.
Where is Isabel from?
She is from Peru.

* 29. ¿Vive usted con sus padres?
No, yo vivo en Pittsburgh y mis padres <u>vivo</u> en Florida. (first-person verb)
Do you live with your parents?
No, I live in Pittsburgh, and my parents live in Florida.

30. ¿Cómo es la casa de Jorge?
Su casa es grande y muy bonita.
What is Jorge's house like?
His/ house is large and very pretty.

31. ¿Cuántos años tiene Marta?
Ella tiene veintidos años.
How old is Marta?
She's 22 years old.

*32. ¿Te gustan las manzanas?
Sí, me <u>gusta</u> las manzanas. (lack of number agreement)
Do you like apples?
Yes, I like apples

33. ¿Qué vas a beber?
Voy a beber vino blanco.
What are you going to drink?
I'm going to drink white wine.

34. ¿Cómo se llama tu amiga?
Se llama Ana.
What's your friend's name?
Her name is Ana.

* 35. ¿Cuándo hablaste con ella?
<u>Hablo</u> con ella ayer por la noche. (present tense verb form)
When did you talk with her?
I talk with her last night.

36. ¿A qué hora desayunas?
Desayuno a las ocho.
What time do you eat breakfast?

I eat breakfast at 8.

* 37. ¿Cuándo termina la clase de español?
La clase termina <u>al</u> cuatro y media. (*al* instead of *a las*)
When does Spanish class end?
The class ends at 4:30.

* 38. ¿Qué hay sobre la mesa?
No hay <u>algo</u> sobre la mesa. (*algo* instead of n*ada*)
What's on the table?
There's nothing on the table.

39. ¿Qué tienes que hacer hoy?
Tengo que estudiar para el examen de español.
What do you have to do today?
I have to study for my Spanish exam.

* 40. Hay muchas niñas aquí. ¿Cuál es Elizabeth?
Es la muchacha <u>pequeño</u> a la derecha. (masc. adj. with fem. noun)
There are a lot of girls here. Which one is Elizabeth?
She is the little girl on the right.

APPENDIX I

SPANISH PROFICIENCY TEST

Part 1. Listening. You will hear 10 short recordings in Spanish. Listen carefully, and after each recording answer the question. Write the letter of the best answer in the blank space. The first 6 questions are in English, the last 4 are in Spanish.

1. What is the purpose of this annoucement?
(a) to prevent childhood diseases
(b) to promote safe playgrounds for children
(c) to explain school entrance requirements
(d) to recall a historical event — Your answer: _____

2. What is a requirement to apply for this job?
(a) The applicant must own a car and be able to drive.
(b) The applicant must be attending high school.
(c) The applicant must be currently employed.
(d) The applicant must speak both languages. — Your answer: _____

3. For what occasion are the cyclists competing?
(a) a sightseeing tour
(b) a presidential election
(c) a national holiday
(d) a bicycle sale — Your answer: _____

4. To whom is this advertisement directed?
(a) people who want to purchase discount travel

packages.
(b) people who want to work for an airline.
(c) people who want to start their own business
(d) people who want to submit an article Your answer: _____

5. What is unusual about this Olympic competition?
(a) the location of the events
(b) the nationality of the participants
(c) the number of competitors
(d) the age of the athletes Your answer: _____

6. What did the two young people from Madrid win?
(a) a trip to a Disney theme park
(b) a chance to go hiking in Spain
(c) free meals at a restaurant
(d) a study abroad program Your answer: _____

7. ¿Qué puede hacer con este producto?
(a) limpiar su cocina
(b) cocinar todo tipo de comida
(c) adquirir muchas vitaminas
(d) bajar de peso Your answer: _____

8. ¿Qué sugiere este dentista?
(a) necesita cerrar la boca
(b) necesita visitar al médico
(c) necesita usar una nueva pasta de dientes
(d) necesita hacer otra cita Your answer: _____

9. Según el anuncio, ¿por qué es especial este tipo de pizza?
(a) viene con un regalo gratuito
(b) está en venta por un tiempo limitado
(c) viene sin ingredientes encima
(d) tiene una variedad de tamaños Your answer: _____

10. ¿Qué van a iniciar estos dos actores?
(a) un canal de televisión nuevo en Miami

(b) dos nuevas telenovelas
(c) un viaje para niños y padres
(d) unas clases y cursos Your answer: _____

Part 2. Reading. Read the following passages and answer the questions that follow them. Choose the best answer and write the letter of your answer in the space provided.

Nuevo lavavajillas "Miele Oceanis":
Una decisión que no te sorprenderá.

El nuevo lavavajillas Miele Oceanis, incorpora la tecnología más avanzada para poder dar a cada tipo de plato el tratamiento que necesita: seis programas de lavado, dos temperaturas, mayor capacidad en el cesto superior...y por supuesto es muy silencioso y cuida al máximo el medio ambiente. Una decisión que no te sorprenderá cuando lo conozcas. Y que celebrarás todos los días...a la hora de lavar los platos.

1. What is emphasized in this advertisment?
(a) a free set of dishes with the purchase of the product
(b) the product's low price
(c) the features of this new product
(d) a decision to change the name of the product Your answer: _____

Lo del Momento
Quince años de graduadas

El próximo sábado 16 de julio, las integrantes de la Promoción de Bachilleres egresadas del Colegio San José de Tarbes de La Florida en el año de 1988, festejarán el décimo quinto aniversario de su graduación.

La celebración consiste en un almuerzo en el restaurante Vecchio Molino y las integrantes del grupo pueden ponerse en contacto con la señora Leonor Sarmiento, (telefono 77.03.91), en el colegio, en La Florida.

2. What is being celebrated?
(a) the reopening of a restaurant
(b) a class reunion
(c) a wedding anniversary
(d) a job promotion Your answer: _____

La mujer en la historia

En el número de mayo/junio leí el excelente
artículo titulado "El noble rival del
Libertador," escrito por Martha Gil-Montero.
En el artículo se menciona que la autora acaba
de terminar una novela basada en la vida de
Manuela Sáenz. Me interesa mucho las
historias de las mujeres de la conquista y la
independencia, y especialmente la de doña
Manuela. Quisiera saber si la novela ya está
publicada y dónde se puede comprar.

3. What is the purpose of this letter to the editor?
(a) to inquire about the price of a subscription
(b) to contribute an article to the magazine
(c) to get information about a new book
(d) to complain about a recent article Your answer: ______

GRANDES ALMACENES "EMY"
Venga a "Emy" y seleccione de nuestra
gran variedad de ropa.
Somos mayoristas y distribuidores de las
marcas de más prestigio.

_ Venga y pregunte sobre nuestros programas de al contado y de crédito-mayoreo, nosotros le asesoramos.

_ Conozca nuestros 4 pisos con el mayor surtido en ropa y accesorios para Damas, Caballeros y Niños, con las mejores marcas como Vanity, Vitos, Nina Ricci, Marsel, Christian Dior, Yves Saint Laurent y muchas más.

_ Aceptamos las principales tarjetas de crédito.

_ Estacionamiento anexo al edificio.

_ En su primera compra de mayoreo le pagamos su boleto de transporte terrestre y en el D.F. mandamos por Usted a su domicilio.

4. This announcement would appeal most to people who want to...
(a) start a new business
(b) work in a department store
(c) learn about international finance
(d) purchase clothing Your answer: _____

UBICACION DE LA UDLA

La Universidad de las Américas-Puebla fue construida en lo que fue la hacienda de Santa Catarina Mártir, en el Municipio de San Andrés Cholula, Estado de Puebla.

El campus, junto a Cholula, la ciudad más antigua del continente, queda a 15 minutos de Puebla, la capital del estado del mismo nombre y está a 120 kilómetros de la cuidad de México.

Toda la zona es un importantísimo sitio arqueológico debido a su relevancia prehispánica como centro religioso. Por lo mismo, también es significativo lo colonial. Es común que cerca de una construcción prehispánica exista arquitectura colonial. Por ejemplo, la pirámide de Tlachihualtépetl sostiene la iglesia de Nuestra Señora de los Remedios.

5. Which statement best describes the campus?
(a) It is in danger of closing.
(b) It is an example of modern architecture.
(c) It is located in the capital of Mexico.
(d) It is situated in an area rich in historical sites. Your answer: _____

Rita Moreno: Leyenda Latina

"Soy un dinosaurio, una especie que está a punto de desaparecer," dice Rita Moreno, una de las pocas artistas que han ganado los cuatro premios más grandes del mundo del espectáculo- un Oscar por su papel en *West Side Story*, un Tony por *The Ritz*, dos Emmys por su actuación en *The Muppet Show* y *The Rockford Files*, y un Grammy por contribuir en un disco con *The Electric Company*. Esta artista puertorriqueña que canta, baila y actúa debe su celebridad a su persistencia y perseverancia personal. Por eso, sirve como ejemplo positivo para los latinos de todas las generaciones.

Desde el momento en que dejó su ciudad natal, Humacao, Puerto Rico, Rita aprendió a luchar para sobrevivir. Cuando tenía cuatro años, Rita llegó a Nueva York con su madre. Se encontró en un ambiente totalmente distinto al que había dejado en Puerto Rico. Cuando fue a la escuela por primera vez, su mamá la dejó en el salón sin saber la minima palabra de inglés. En este mundo extraño, Rita tuvo que trabajar con mucha energía para adaptarse a un nuevo idioma y cultura. Con su talento y su fuerte labor, no le llevó mucho tiempo para dominar lo difícil. Una década más tarde a los trece años, debutó en Broadway con su participación en *Skydrift*, un drama de 1945. Pero aquella primera aventura teatral solo duró siete presentaciones y para poder mantener a su familia, Rita tuvo que trabajar en bares cantando y bailando flamenco.

Rita trató de demostrar su versatilidad con varios proyectos. Durante los años '70 y '80 trabajó sin cesar para sobresalir en los campos del teatro, del cine y de la televsión. Ella nunca abandonó su batalla de romper los estereotipos ante los latinos en los Estados Unidos. Además de su carrera artística, Rita Moreno es una activista política, a favor de las causas humanitarias, especialmente en cuanto a los niños latinos que tienen desventajas. En el futuro parece que no hay extinción de este "dinosaurio" talentoso.

6. ¿Cuál es la cosa más notable sobre Rita Moreno?
(a) nació en un lugar muy cerca de Hollywood
(b) solamente quiere cantar en español
(c) ha recibido una variedad de premios
(d) tiene interés en la conservación Your answer: _____

7. ¿A qué se debe la fama de esta artista?
(a) a su pura determinación
(b) a su buena suerte
(c) a su familia rica
(d) a su educación en la universidad Your answer: _____

8. ¿Qué problema enfrentó ella a llegar a Nueva York?
(a) no podía alquilar un apartamento
(b) no tenía buena salud
(c) no ganaba el concurso
(d) no entendía bien la lengua Your answer: _____

9. ¿Cuándo empezó su carrera profesional?
(a) cuando era una jovencita
(b) después de hacerse abuela
(c) en los años setenta
(d) cuando vivía en Puerto Rico Your answer: _____

10. ¿Cuál es un deseo actual de Rita?
(a) ganar más dinero
(b) ayudar a la gente latina
(c) volver a su país de origen
(d) escribir su autobiografía Your answer: _____

Part 3. Writing. An exchange student from Spain named Fernando is going to visit your campus next month. Write him a letter in Spanish telling him about yourself and your campus. You can write as much as you want, but try to write at least 60 words. Be sure that your letter has a beginning, a middle and an end, and that the sentences are logically connected. Demonstrate your range of vocabulary knowledge, and try to avoid repeating yourself. You may wish to include some of the following information:

_ what your classes are like
_ what your language experience is
_ what your favorite pasttimes are
_ a description of the area where you live
_ suggestions for possible activities
_ a description of your campus
_ suggestions for meal arrangements

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. The Quarterly Journal of Experimental Psychology 49-A (1), 216-233.

Adams, A-M. & Willis, C. (2001). Language processing and working memory: A developmental perspective. In Working Memory in Perspective, J. Andrade (ed.), pp. 79-100. Hove, U.K.: Psychology Press.

Anderson, J.R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Ardila, A. (2003). Language representation and working memory with bilinguals. Journal of Communication Disorders 36, 233-240.

Atkinson, R.C. & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In L.W. Spence (Ed.), The Psychology of Learning and Motivation: Advances in Research and Theory, (pp. 85-195). New York: Academic Press.

Babcock, R.L. & Salthouse, T.A. (1990). Effects of increased processing demands on age differences in working memory. Psychology and Aging 5, 421-428.

Baddeley, A.D. (1996). Exploring the central executive. The Quarterly Journal of Experimental Psychology 49A-1, 5-28.

Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? Trends in Cognitive Science 4-11, 417-423.

Baddeley, A.D. (2003). Working memory: Humans. In: J.H. Byrne (Ed. in chief), Learning and Memory (2nd ed.) (pp. 672-676). New York: Thomson-Gale.

Baddeley, A.D., Gathercole, S. & Papagno, C. (1998). The phonological loop as a language learning device. Psychological Review 105, 158-173.

Baddeley, A.D. & Hitch, G.J. (1974). Working memory. In G. Bower (ed.), The Psychology of Learning and Motivation, vol. 8 (pp. 47-89). New York: Academic Press.

Baddeley, A.D. & Hitch, G.J. (2001). Foreward. In Working Memory in perspective, J. Andrade (ed.), xv-xix. Hove, U.K.; Psychology Press.

Baddeley, A.D. & Logie, R.H. (1999). Working memory: The multiple-component model. In: Models of Working Memory: Mechanisms of Active Maintenance and Control, A. Miyake & P. Shah (eds.), 28-61. Cambridge, U.K.; Cambridge University Press.

Baddeley, A.D., Papagno, C. & Vallar, G. (1988). When long-term memory depends on short-term storage. Journal of Memory and Language 27, 586-595.

Beardsmore, H.B. (1974). Testing oral fluency. IRAL 12-4, 317-26.

deBot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. Applied Linguistics 13, 1-24.

Bower, G.H. (2000). A brief history of memory research. In: The Oxford Handbook of Memory, E. Tulving & F.L.M. Craik (eds.), 3-32. New York: Oxford University Press.

Broadbent, D.E. (1958). Perception and communication. New York: Pergamon Press.

Brown, H.D. (1994). Principles of language learning and teaching (3rd ed.). Englewood Cliffs, N.J.; Prentice Hall Regents.

Brumfit, C. (1984/2000). Accuracy and fluency: The basic polarity. In: H. Riggenbach (ed.) (2000), Perspectives on fluency, 61-73. Ann Arbor: University of Michigan Press.

Burgess & Hitch (1996). A connectionist model of STM for serial order. In Models of short-term memory, S.E. Gathercole (ed..), 51-72. Hove, U.K.: Psychology Press.

Cantor, J. & Engle, R. (1993). Working memory capacity as long-term activation: An individual differences approach. Journal of Experimental Psychology: Learning, Memory, and Cognition 19, 1101-1114.

Carpenter, P.A., Miyake, A., & Just, M.A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In: M.A. Gernsbacher (ed.), Handbook of Psycholinguistics, (pp. 1075-1122). San Diego; Academic Press.

Carroll, J.B. & Sapon, S.M. (1959). Modern Language Aptitude Test. New York: The Psychological Corporation/Harcourt Brace Jovanovich.

de Castells, M.O., Guzman, E., Lapuerta, P., & Garcia, C. (2002). Mosaicos Spanish as a world language (3rd ed.). Upper Saddle River, N.J.: Prentice Hall.

Chase, W.G. & Simon, H.A. (1973). Perception in chess. Cognitive Psychology 1, 33-81.

Chincotta, D. & Underwood, G. (1997). Bilingual memory span advantage for Arabic numerals over digit words. British Journal of Psychology 88, 295-310.

Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge: MIT Press.

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

Committee on Developments in the Science of Learning (2000). How people learn Brain, Mind, Experience, and School. Washington, D.C.: National Academy Press.

Conway, A.R.A. & Engle, R.W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. Memory 4, 577-590.

Corder, P. (1983). Strategies of communication. In: Strategies in interlanguage communication, C. Faerch & G. Kasper (eds.), 15-19. London: Longman.

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior 11, 671-684.

Crowder, R.G. (1976). Principles of learning and memory. Hillsdale, N.J.; Erlbaum.

Crystal, D. (1997). The Cambridge encyclopedia of langauge (2nd ed.). Cambridge: Cambridge University Press.

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 Learning and Verbal Behavior 19, 450-466.

Daneman, M. & Case, R. (1981). Syntactic form, semantic complexity, and short-term memory: Influence on children's acquisition of new linguistic structures. Developmental Psychology 17, 367-378.

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

Dechert, H.W. (1983). How a story is done in a second language. In C. Faerch & G. Kasper (eds.), Strategies for Interlanguage Communication (pp. 175-195). London: Longman.

DeGroot, A.D. (1965). Thought and choice in chess. The Hague: Mouton.

Deese, J. (1980). Pauses, prosody, and the demands of production in language. In: Temporal variables in speech: Studies in honor of Frieda Goldman-Eisler, 69-84. The Hague: Mouton.

DeKeyser, R.M. (1988). Communicative processes and strategies. Annual Review of Applied Linguistics 9, pp. 108-121.

DeKeyser, R.M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. Studies in Second Language Acquisition 19, 195-221.

Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In: H. Riggenbach (ed.), Perspectives on Fluency (pp. 287-313). Ann Arbor: University of Michigan Press.

Ellis, N. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. Studies in Second Language Acquisition 18, 91-126.

Ellis, N. & Schmidt, R. (1997). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. Studies in Second Language Acquisition 19, 145-171.

Ellis, N. & Sinclair, S.G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. The Quarterly Journal of Experimental Psychology 49-A, 234-250.

Engle, R.W. (2001). What is working memory capacity? In: H.L. Roediger III, J.S. Nairne, I. Heath & A.M. Suprenant (eds.), The nature of remembering: Essays in honor of Robert G. Crowder (pp. 297-314). Washington, D.C.: American Psychological Association.

Engle, R.W., Kane, M.J., & Tuholski, S.W. (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, 102-134. Cambridge, U.K.; Cambridge University Press.

Ericsson, K.A. (1996). The acquisition of expert performance: An introduction to some of the issues. In: K.A. Ericsson (ed.), The road to excellence: The acquisition of expert performance n the arts and sciences, sports, and games, 1-50. Mahwah, N.J.: Erlbaum.

Ericsson, K.A. & Charness, N. (1994). Expert performance: Its structure and acquisition. American Psychologist 49, 725-747.

Ericsson, K.A. & Delaney, P.F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In: Models of working memory Mechanisms of active maintenance and control, A. Miyake & P. Shah (eds.), 257-297. Cambridge: Cambridge University Press.

Ericcson, K.A. & Kintsch, W. (1995). Long-term working memory. Psychological Review 102, 211-245.

Faerch, C. & Kasper, G. (1983). Plans and strategies in foreign language communication. In C. Faerch & G. Kasper (eds.), Strategies in Interlanguage Communication (pp. 20-60). London: Longman.

Fillmore, C.J. (1979/2000). On fluency. In: H. Riggenbach (ed.) (2000), Perspectives on fluency, 43-60. Ann Arbor: University of Michigan Press.

Flynn, S. (1986). Production vs. comprehension: Differences in underlying competencies. Studies in Second Language Acquisition 8, 135-164.

Fortkamp, M.B. (1999). Working memory capacity and aspects of L2 speech production. Communication & Cognition 32-3/4, 259-296.

Freed, B. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In H. Riggenbach (ed.), Perspectives on Fluency (pp. 243-265). Ann Arbor: University of Michigan Press.

French, L.M. (2003). Phonological working memory and L2 acquisition: A developmental study of Quebec Francophone children learning English. Unpublished Ph.D. dissertation. Laval, Canada: L'Universite Laval, Quebec.

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

Gathercole, S.E. & Baddeley, A.H. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. British Journal of Psychology 81, 439-454.

Gathercole, S.E. & Baddeley, A.H. (1993). Working memory and language. Hove, U.K.: Erlbaum.

Goldman-Eisler, F. (1968). Psycholinguistic experiments in spontaneous speech. New York: Academic Press.

Goldman-Rakic, P.S. (1999). Working memory and the mind. In: The Scientific American book of the brain, 91-104. New York: The Lyons Press.

Groeger, J. (1997). Memory and remembering: Everyday memory in context. New York: Longman.

Guillot, M-N. (1999). Fluency and its teaching. Clevedon, U.K.: Multilingual Matters.

Hammerly, H. (1991). Fluency and accuracy. Clevedon, U.K.: Multilingual Matters.

Harris, D.P. (1969). Testing English as a second language. New York: McGraw-Hill.

Hieke, A.E. & Dunbar, Ronald W. (1985). Building fluency in English: Authentic Speech. Englewood Cliffs, N.J.: Prentice-Hall.

Henning, G.H (1973). Remembering foreign language vocabulary: Acoustic and semantic parameters. Language Learning 23, 207-222.

Hughes, A. (1989). Testing for language teachers. Cambridge, U.K.: Cambridge University Press.

Hunt, K.W. (1970). Syntactic maturity in school children and adults. Monographs of the Society For Research in Child Development, 53 (134), Chicago: University of Chicago Press.

James, W. (1890). The principles of psychology. New York: Henry Holt & Co.

Juffs, A. (2003). Working memory: Understanding current issues in mainstream psychology. Paper presented at Second Language Research Forum. Tucson, AZ: October, 2003.

Juffs, A. (2004). Representation, Processing and Working Memory. Transactions of the Philological Society 102-2, 195-225.

Just, M.A. & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. Psychological Review 99, 122-149.

Just, M.A., Carpenter, P.A., & Keller, T.A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. Psychological Review 103-4, 773-780.

Kane, M.J., Conway, A.R.A., & Engle, R.W. (1999). What do working-memory tasks really measure? Behavioral and Brain Sciences 22, 101-102.

Kemper, S. (1986). Imitation of complex syntactic constructions by elderly adults. Applied Psycholinguistics 7, 277-287.

Kenyon, D.M. (1995). An investigation of the validiy of task demands on performance-based tests of oral fluency. In: A.J. Kunnan (ed.), Validation in language testing, 19-40. Mahway, N.J.: Erlbaum.

King, J. & Just, M.A. (1991). Individual differences in syntactic processing: The role of working memory. Journal of Memory and Language 30, 580-602.

Kinkade, J. (1995). Predictors of ESL fluency ratings by native and non-native raters. Unpublished M.A. thesis. University of Pittsburgh: Pittsburgh, PA.

Kozyrev, Joann Rishel (2002). Talk it over! Listening, speaking, and pronunciation. Boston: Houghton-Mifflin.

Krashen, S.D. (1981). Second language acquisition and second language learning. Oxford: Pergamon.

Krashen, S.D. (1982). Principles and practices in second language acquisition. London: Pergamon.

Krashen, S.D. & Terrell, T.D. (1983). The natural approach Language acquisition in the classroom. Hayward, CA: Alemany Press.

Kroll, J.F. & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In: M.M. Gruneberg, P.E. Morris & R.N. Sykes (eds.), Practical Aspects of Memory: Current Research and Issues, Vol. 2. London: John Wiley & Sons.

Kroll, J.F., Michael, E., & Sankaranarayanan, A. (1998). A model of bilingual representation and its implications for second language acquisition. In: A.F. Healy & L.E.Bourne Jr. (eds.), Foreign language learning: Psycholinguistic studies on training and retention (pp. 365-411). Mahwah, N.J.: Erlbaum.

Leeson, R. (1975). Fluency and language teaching. London: Longman.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. Language Learning 40-3, 387-417.

Lennon, P. (2000). The lexical element in spoken second language fluency. In: H. Riggenbach (ed.), Perspectives on Fluency (pp. 25-42). Ann Arbor: University of Michigan Press.

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

Logan, G.D. (1988). Toward an instance theory of automatization. Psychological Review 95, 492-527.

MacDonald, M.C. & Christiansen, M.H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). Psychological Review 109, 35-54.

Mackay, A., Philp, J., Egi, T., Fujii, A. Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P.Robinson (ed.), Individual differences and instructed language learning (pp. 181-209). Amsterdam: John Benjamins.

Mayer, M. (1969). Frog, where are you? New York: Dial Press.

Meara, P. (1984). The study of lexis in interlanguage. In A. Davies, C. Criper & A.P.R. Howatt (eds.), Interlanguage (pp. 225-235). Edinburgh: Edinburgh University Press.

Michael, E.B. & MacWhinney, B. (2003). Span, Stroop and second language acquisition. Poster presented at the Psychonomic Society, Vancouver, B.C.

Miller, G.A. (1956). The magical number seven plus-or-minus two: Some limits of our capacity for processing information. Psychological Review 63, 81-97.

Miyake, A. & Friedman, N. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A.F. Healy & L.E.Bourne Jr. (eds.), Foreign language learning: Psycholinguistic studies on training and retention (pp. 339-364). Mahwah, N.J.: Erlbaum.

Miyake, A., Just, M.A., & Carpenter, P.A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral context. Journal of Memory and Language 33, 175-202.

Miyake, A. & Shah, P. (1999a). Models of working memory: An introduction. In: A. Miyake & P. Shah (eds.), Models of Working Memory: Mechanisms of Active Maintenance and Executive Control, 1-27. Cambridge, U.K.: Cambridge University Press

Miyake, A. & Shah, P. (1999b). Toward unified theories of working memory: Emerging general consensus, unresolved theoretical issues, and future research. In: A. Miyake & P. Shah (eds.), Models of Working Memory: Mechanisms of Active Maintenance and Executive Control, pp. 442-481. Cambridge, U.K.: Cambridge University Press.

Neath, I. (1998). Human memory: An introduction to research, data, and theory. Pacific Grove, CA; Brooks/Cole Publishing Co.

New York State Library. (2004). Archive of past Regents Exams in Spanish. Available at: <nysl.nysed.gov/edocs/education/exams/spanish.htm> Accessed March, 2004.

Norman, D.A. & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R.J. Davidson, G.E. Schwartz, & D. Shapiro (eds.), Consciousness and self-regulation, vol. 4. New York: Plenum Press.

Omaki, A. & Ariji, K. (2003). Working memory capacity for sentence processing: Is it language-independent? Paper presented at Second Language Research Forum. Tucson, AZ: October, 2003.

Oppenheim, N. (2000). The importance of recurring sequences for nonnative speaker fluency and cognition. In Perspectives on Fluency, H. Riggenbach (ed.), 220-240. Ann Arbor: University of Michigan Press.

O'Reilly, R.C., Braver, T.S., & Cohen, J.D. (1999). A biologically based computational model of working memory. In: A. Miyake & P. Shah (eds.), Models of working memory Mechanisms of active maintenance and control, 375-411. Cambridge: University of Cambridge Press.

Osaka, M. & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. Bulletin of the Psychonomic Society 30-4, 287-289.

Page, M. & Henson, R. (2001). Computational models of short-term memory: Modelling serial recall of verbal material. In J. Andrade (ed.), Working memory in perspective (pp. 177-198). Hove, U.K.: Psychology Press.

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

Pashler, H. (2003). Attention and memory. In J.H. Byrne (ed. in chief), Learning and Memory (2nd ed.) (pp. 48-51). New York: Thomson-Gale.

Pawley, A. & Syder, F.H. (2000). The one-clause-at-a-time hypothesis. In: H. Riggenbach (ed.), Perspectives on Fluency, 163-199. Ann Arbor: University of Michigan Press.

Peterson, L.R. & Peterson, M.J. (1959). Short-term retention of individual items. Journal of Experimental Psychology 58, 193-198.

Pienemann, M. (1998). Language processing and second language development: Processability theory. Amsterdam: John Benjamins.

Raupach, M. (1983). Analysis and evaluations of communication strategies. In C. Faerch & G. Kasper (eds.), Strategies in Interlanguage Communication (pp. 199-209). London: Longman.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. Discourse Processes 14, 423-441.

Riggenbach, H. & Lazaraton, A. (1991). Promoting oral communication skills. In: M. Celce-Murcia (ed.), Teaching English as a second or foreign language (2nd ed.), 125-136. Boston: Heinle & Heinle.

Ringbom, H. (1983). Borrowing and lexical transfer. Applied Linguistics 4, 207-212.

Roberts, R. & Gibson, E. (2002). Individual differences in sentence memory. Journal of Psycholinguistic Research 31-6, 573-598.

Robinson, P. (1997). Generalizability and automacity of second language learning under implicit, incidental, enhanced and instructed conditions. Studies in Second Language Acquisition 19, 223-247.

Rost, Michael (1998). Strategies in speaking. Hong Kong: Longman.

Sajavaara, K. (1987). Second language speech production: Factors influencing fluency. In H.W. Dechert & M. Raupach (eds.), Psycholinguistic models of speech production (pp. 45-66). Norwood, N.J.: Ablex.

Salthouse, T.A. & Babcock, R.L. (1991). Decomposing adult age differences in working memory. Developmental Psychology 27-5, 763-776.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. Studies in Second Language Acquisition 14, 357-85.

Schneider, P. (1999). Working memory in a multilevel hybrid connectionist control architecture (CAP2). In: A. Miyake & P. Shah (eds.), Models of working memory Mechanisms of active maintenance and control, 340-374. Cambridge: Cambridge University Press.

Schneider, W. & Schiffrin, R.M. (1977). Controlled and automatic information processing: Part 1 detection, search and attention. Psychological Review 84, 1-66.

Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In: H. Riggenbach (ed.), Perspectives on Fluency, 200-219. Ann Arbor: University of Michigan Press.

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

Singer, M. & Ritchot, K.F.M. (1996). The role of working memory capacity and knowledge access in text inference processing. Memory & Cognition 24, 733-743.

Skehan, P. (2002). Theorising and updating aptitude. In Individual Differences and Instructed Learning, P. Robinson (ed.), 69-93.

Snodgrass, J.G. & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psycholgoy: Human Learning and Memory 6, 174-215.

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

Stevick, E.W. (1976). Memory, meaning & method Some psychological perspectives on language learning. Rowley, MA: Newbury House.

Tarone, E. Cohen, A.D. & Dumas, G. (1983). A closer look at some interlangauge terminology: A framework for communicative strategies. In: C. Faerch & G. Kasper (eds.), Strategies in interlanguage communication, 4-14. London: Longman.

Temple, L. (1997). Memory and processing modes in language learner speech production. Communication & Cognition 30-1/2, 75-90.

Towse, J. N. & Houston-Price, C.M.T. (2001). Reflections of the concept of the central executive. In Working Memory in perspective, J. Andrade (ed.), pp. 240-260. Hove, U.K.; Psychology Press.

Vinther, T. (2002). Elicited imitation: A brief overview. International Journal of Applied Linguistics 12-1, 54-73.

Ward, G. (2001). A critique of the working memory model. In J. Andrade (ed.), Working memory in perspective (pp. 219-239). Hove, U.K..: Psychology Press.

Waters, G.S. & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). Psychological Review 103-4, 761-772.

Waters, G.S. & Caplan, D. (1999). Verbal working memory and sentence comprehension. Behavioral and Brain Sciences 22, 77-126.

Weir, C. (1993). Understanding and developing language tests. New York: Prentice Hall.