

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Working memory's meager involvement in sentence repetition tests

Permalink

<https://escholarship.org/uc/item/5p40b7nd>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 34(34)

ISSN

1069-7977

Authors

Okura, Eve
Lonsdale, Deryle

Publication Date

2012

Peer reviewed

Working memory's meager involvement in sentence repetition tests

Eve Okura (eveokura@hawaii.edu)

Department of Linguistics, 558 Moore Hall,
Honolulu, Hawaii 96822 USA

Deryle Lonsdale (lonz@byu.edu)

Department of Linguistics & English Language, 4039 JFSB
Provo, UT 84602 USA

Abstract

Elicited imitation (EI) is a testing method for learners' oral language proficiency. One common criticism aimed at EI is that performance might not require linguistic knowledge, but mere rote memorization. This study explores the issue by administering two tests to the same group of students studying English as a second language: (1) a working memory test, and (2) an English EI test. Participants came from a range of English language proficiency levels. Our goal was to test whether scores from these two treatments (English EI scores and working memory scores) would correlate significantly. If not, this would suggest that there is some difference in what they measure. The results did fail to show a significant correlation between working memory and English EI scores. On the other hand, there was a significantly positive correlation between students' English EI scores and their placement level.

Keywords: working memory; elicited imitation; oral language testing; explicit/implicit linguistic knowledge

Introduction

Numerous methods exist for assessing language competence, ranging from written grammar examinations to live interviews with a native speaker. Some discussion revolves around the different types of tests: what they actually measure, and whether they directly assess linguistic knowledge and ability versus some other capability. One such testing method is commonly referred to as *elicited imitation* (EI). EI is a testing method in which researchers present a series of sentences to a person who then imitates each sentence by repeating it as accurately as possible. These responses are recorded and subsequently analyzed. Since 1967 EI has been used in linguistic research, particularly for studying language acquisition (Slobin & Welsh, 1973). Many have further proposed that EI tests can assist in measuring language ability (Connell & Myles-Zitzer, 1982; Vinther, 2002; Erlam, 2006). Others have argued (Hamayan, Saegert, & Larudee, 1977) that EI cannot measure language ability, and that listeners can simply repeat what they hear, without involving any linguistic processing. Such arguments claim that a person could repeat the utterance without understanding what it means, and without having any ability in that language. In this paper we report on a study meant to contribute more human performance data to this debate.

Addressing the issue of whether or not EI test-takers rely solely on rote repetition in responses requires some consideration of memory. We assume here an admittedly simple but traditional tripartite view of memory: long-term, short-term, and working memory (Cowan, 1996). Furthermore,

we will focus on only the third component, working memory (WM). WM contains the information held momentarily, as it is needed to analyze, solve a problem, or perform a task.

Working memory is the part of memory that would determine whether high performance levels on EI tests are a result of mere parroting or whether linguistic knowledge and competency are also needed. An EI test-taker who did not know English would have to hear the stimulus utterance, retain it in WM, and then repeat the utterance exactly to get a perfect score on the item. If the item were beyond their WM capacity, they would be expected to get only part of it correct at best, all of it wrong at worst.

Clearly WM and second-language (L2) ability are not completely independent. Some WM capacity is necessary to even be able to respond in an EI test (Doughty & Long, 2003; Robinson, 2005); some WM capacity is also necessary in all analytical and linguistic tasks. In this paper we probe the degree of overlap between them with two EI tests, one targeting English ability and the other WM capacity.

Background: WM, language, and EI

Early research in working memory includes Miller's famous discussion of *immediate memory span* positing a "magic number" of 7 ± 2 (Miller, 1956), meaning that a person could generally hold up to seven completely unrelated items in mind simultaneously. Several researchers have subsequently advocated the number 4 ± 1 instead as the relevant metric defining the scope of average WM capacity (Cowan, 2001). Miller also introduced the process of *recoding* now more commonly referred to as *chunking*. Chunking takes multiple separate items and agglomerates them into patterns, reducing the number of items to remember and speeding up processing.

Another thread of research has sought to differentiate WM, which is highly evanescent, from other types of memory that are longer-lasting. For example, Baddeley and Hitch (1986) experimented to see if storing several numbers in working memory affected the ability to carry out comprehension tasks. It did not, suggesting that memory is stored in a different "place" from where processing and problem-solving occur, as one did not impede the other. A distinction is typically drawn between WM and short-term memory (STM): the latter is a form of memory that lasts for up to 10-30 seconds after one receives a stimulus (Cowan, 1996). Discussions of how language interacts with WM and STM also involve the articulatory loop, a time-constrained buffer that temporarily

stores information when comprehending and producing language. Issues of activation, decay, and interference all complicate and enrich investigation of these areas; to the extent possible we will abstract away from or simplify these factors in the work reported in this paper.

STM involvement in conversational language includes managing spontaneity, multiple interlocutors, conversational turn-taking, contextual factors, meta-awareness, and explicit language knowledge. Evaluating L2 conversational speech abilities typically involves oral proficiency interviews (OPI's) (Lehman & Tompkins, 1998) which are costly and challenging to grade since it involves skilled human interlocutors.

Our study differs from natural speech in that spontaneity does not occur, as it would in a traditional OPI. In an EI test, as opposed to natural conversation, there is a response, but it is imitative rather than a spontaneous utterance invented by the individual. STM is what holds the stimulus information when the response is being planned (Cowan, 1996).

Language acquisition researchers distinguish between *acquisition* and *learning*. Acquisition is done intuitively, similar to a child developing her native language abilities (Krashen, 1982). Competence, the result of acquisition, is also subconscious: we are generally not consciously aware of the grammatical rules of the languages we have acquired. Instead, we have a “feel” for correctness, and errors do not “sound” or “feel” right, even if we do not consciously know what rule was violated. Krashen associates language acquisition with *implicit learning*.

On the other hand, Krashen defines *language learning* as a conscious awareness of grammar, vocabulary, syntax, etc., and refers to this as *explicit learning* (Krashen, 1982). Ellis echoes Krashen's use of the terms implicit and explicit and extends them beyond the learning process to linguistic knowledge itself (N. C. Ellis, 2008). Thus explicit and implicit knowledge of language are distinct and dissociated, they involve different types of representation, they are substantiated in separate parts of the brain, and yet they can come into mutual influence in processing. Explicit knowledge of a language is form-focused rather than meaning-focused, and involves conscious thought. Presumably much of the implicit and explicit knowledge of a language are stored away in long-term memory (LTM) and marshaled as necessary by STM when required.

EI testing is interesting for a variety of reasons. First, it is time-constrained; the responses must be immediate, so extensive deliberation is not possible. The time constraint also prevents use of the articulatory loop to enhance WM, and thus scores on the EI test. Second, the test involves repetition instead of full-fledged conversational interaction so that deliberation about of grammatical form is less salient. These two elements—time limit and avoiding meta-awareness of form—are thought to shift EI tests towards the assessment of implicit language knowledge instead of explicit knowledge (R. Ellis, 2005).

Some combination of WM, STM, and LTM is pressed into

service for language-based interactions including simultaneous listening and comprehension. STM, like WM, is limited in its capacity and reactivity, so processing language must involve optimally combining layers of representation of meaning (morphemes, lexemes, lexical items, phrases, sentences), that allow STM to group items together, thereby expanding its ability to deal with the situation at hand.

Meaning is another important consideration in memory, language, and learning (Doughty & Long, 2003). Robinson (2005) noted that phonological WM capacity in particular is associated with L2 speaking abilities. DeKeyser notes the difference between *meaningful* versus *nonmeaningful* associations, each in its relation to memory capacity: meaningful form-function mappings associate constituents in a sentence according to lexical and grammatical principles of the language. Establishing meaningful relationships between abstract entities draws more on insight, whereas associating nonmeaningful co-occurrence of concrete elements logically draws more on memory (DeKeyser, 2005).

Utterances are constructed as intonation units, and substantive units are fairly strongly constrained to have a typical length of about four words in English, indicative of the cognitive limits on how much information can be fully active in the mind at any one time (N. C. Ellis, 2002). At the syllable level most information has already been chunked up, reflecting the phonemic, morphological, and lexical properties of the language: speakers rarely deliberate at the syllable level in languages they are proficient in. On this basis, one would expect a nonce syllable test would measure memory, and a “meaningful” (i.e. language-based) test would draw less on memory than the nonce test, and more on linguistic ability.

Previous research has also shown that knowledge of semantics, grammar, syntax, etc., affects verbal STM performance. This suggests that such linguistic working memory chunking also occurs during the process of verbal repetition. The articulatory loop is also implicated in both vocal and subvocal repetition of a phrase, which helps to maintain it in short-term memory. The connection is complicated by factors including articulation rate, semantic properties, grammar class, word frequency or familiarity, and word sequencing. Since chunking can happen at these linguistic levels as well, an interesting question is the interaction of linguistic versus articulatory loop chunking; current thinking is that the former can supersede the latter (Morra, 2000).

It follows that, due to WM limitations, correctly repeating longer utterances in an EI test is only possible via linguistic recoding. Hence WM tasks involving processing of nonce syllables will be different from WM tasks enhanced by language ability, and the difference indicates to what degree the EI tests are measuring language.

Elicited imitation (EI) and sentence repetition testing (SRT) are the most common terms used to refer to the same testing method. Though a thorough review of the extensive history in EI testing is not possible here, we mention some of the strands of work most relevant to our current experiment.

EI is the “...repetition of a model sentence presented in a context calling for imitation, as opposed to...spontaneous imitation of...utterances” (Slobin & Welsh, 1973). Bley-Vroman and Chaudron (1994) suggest that EI performance requires both linguistic processing and short-term memory. They sketch a 4-step process of EI performance; the test-taker: (1) hears the utterance; (2) forms their version of it in their mind (a *representation*); (3) stores this representation in short-term memory; and then (4) verbalizes their representation of the initial utterance.

EI has been instrumental in measuring first language acquisition. By studying the EI of adult speech in children, Slobin and Welsh (1973) conclude that the process requires linguistic processing, particularly since children cannot repeat beyond what they could produce spontaneously, unless the utterances were short enough to leverage immediate memory. He argues that the same conclusions hold for adult L2 learners.

A well known basic asymmetry in language capabilities exists between comprehension and production; this also underlies EI testing. Vinther points out whereas some test subjects may be able to understand the stimulus but not reproduce it accurately, on the other hand “the opposite situation, that subjects should be able to produce well-formed imitations of sentences they have not understood and are not able to remember, is highly improbable” (Vinther, 2002).

Prior literature also provides guidelines for EI testing. Jessop et al. outlined the advantages and challenges of using EI as an L2 acquisition assessment, providing suggestions for using EI effectively. This paper addresses the central problem they delineate: “In the 1970s, EI’s validity was challenged: the major criticism being the possibility of rote repetition in response to stimuli (i.e., participants may be simply parroting what they hear)” (Jessop, Suzuki, & Tomita, 2007). They call for more study to validate the use of EI in measuring L2 performance.

EI tests thus measure verbal STM performance in the sense that they measure knowledge of the various linguistic elements discussed above, rather than rote memory ability. It is also apparent that WM is essential to perform well on a language EI test. The question is how much do language EI tests overlap with working memory tests in what they measure? If the two do not correlate perfectly, then whatever does not overlap implies that the remaining portion, the part of the EI test that does not overlap with WM, is measuring something else. As suggested by previous research, that “something else” is knowledge of grammar, vocabulary, syntax, and other linguistic features. To quote one researcher, “Elicited imitation goes beyond rote memory and repetition; rather, sentences are assumed to be ‘filtered’ through one’s grammatical system” (Gass & Mackey, 2007). This would predict that highly proficient L2 learners will do better at EI tests since their knowledge and expertise will help overcome memory capacity and time constraints.

Conversely, an L2 EI test would function more similarly to a WM test for individuals who do not yet have any ability

in the language. This is because without necessary language competence, the individual will not be able to “chunk” and use other opportunistic linguistic strategies; syllables would instead sound like nonce syllables to the student. For longer sentences mere rote repetition will not be possible since WM capacity will be exceeded.

If there were a significant WM/EI correlation across the board, one could conclude that memory is dominating the measurement of language ability. This would support arguments that EI tests are mere “parroting”, and hence offer no contribution to measuring language ability. If, however, there is minimal correlation between WM scores and EI language test scores, the tests cannot be primarily memory tests. We also predict that some students who have approximately the same WM capacity but differing language ability will score differently on the EI language test. In summary, we claim that EI language tests, when constructed properly, primarily measure linguistic ability, and demonstrably do not measure an individual’s working memory to a significant degree.

Testing EI and WM

We administered two tests to a pool of students: an English EI test and a WM test. For the EI test, we used our own pre-existing English instrument: EI Form D. It contains 60 questions of varying lengths and complexity, carefully designed and engineered to target different levels of proficiency. Half of the items were also paired with a comprehension task.

The comprehension task involved two illustrations appearing side-by-side on the screen after an EI stimulus sentence was given, but before the test-taker recorded their response. One illustration depicted the meaning of the stimulus sentence, and the other picture was irrelevant to the stimulus. Test-takers were to click on the relevant picture. Only then were they allowed to record their audio response. These tasks were intended to divert attention away from form, and shifting the focus to meaning.

We chose to create our own working memory test—as opposed to using a previously existing one—in order to assure that its content, format, and delivery method integrated seamlessly with the English EI test. Our WM test uses nonce syllables (Cowan, 2005) drawn from a pre-existing repository of nonce words (McGhee, 2010). Meaningless monosyllabic words such as “kish” were used directly, and longer ones were split into nonce syllables. Caution was exercised to ensure that the individual syllables were not homophones of actual English words, either in isolation or in sequence; this would have enabled chunking and hence complicated WM measure interpretation. The 55 nonce syllables were arranged into ten items of increasing length: item 1 has one syllable, and item 10 has 10 syllables.¹

Half (i.e. five) randomly selected WM items were recorded

¹A reviewer has observed—and we concur—that a more commensurate WM test would have had sentences with (nonce) syllable totals equalling those in the EI test sentences. Unfortunately we no longer have access to the subject population for another within-subjects test.

by a male voice, and the other by a female voice, presented in randomized order. Syllables within a single item were separated by pauses of approximately one second in length. Care was also taken to assure flat prosody when pronouncing the syllables, avoiding any intonational contours that might make them easier to remember in sequence.

We administered the English EI test in January 2011 to incoming adult students at the BYU English Language Center (ELC) as a part of an initial placement assessment. The ELC in essence has 6 grade levels, Levels 0 through 5. Students are placed at these levels based on multiple assessments of their abilities, including grammar, writing, and listening comprehension tasks. Our test was given at the same time as other diagnostic language tests. Our EI test results were not used in placing students in their ELC levels.

The 94 students were randomly divided into two equally sized groups (A and B). Group A took the 60-item English EI test with 30 items paired with distractor comprehension tasks. Group B was given the same 60-item English test, but with the other half of the 60 items paired with comprehension tasks.

The presentation order of English EI items was randomized within each set (questions with tasks, and questions without tasks, respectively) to eliminate any possible item-order effects. In the test both groups were presented the section without the tasks first, and afterwards the section with the tasks. Time allowed for responses for individual items was proportional to their length. Due to technical difficulties (microphone malfunctions, web server problems, etc.) about half of the students' tests had to be discarded.

A few months later, in March and April, as many of the original students as could be located were invited to take the 10-item working memory test. This resulted in 67 students; some students had left the program, and others were unavailable due to scheduling problems. Again, various technical difficulties resulted in tests that were lost or unusable. The final pool of data consisted of 40 students for whom we had a complete set of English EI and WM scores.

Both tests were scored by a percentage method, i.e. the overall score on the test is a percent of the total number of syllables the student uttered correctly over the total number of syllables in the stimuli of the test. As with our prior work, we had trained human graders (all native English speakers for this project) score the test recordings via a web-deployed user-friendly interface. The rater listens to the audio recording, and marks each syllable that is present with a "1," and each syllable in the utterance that the test-taker missed with a "0." Six individuals graded results for the 5,430 English test items (89 test-takers times 60 items per test), and the 890 working memory test items. Individual EI item scores were aggregated across each test.

Results and Discussion

To ensure that the WM test was actually testing memory, we analyzed the scores of all participants for each item. All WM

scores ranged from 27.27% to 52.73% (mean and mode at about 40%), with one outlier at 20%. As indicated in Figure 1, average accuracy across students on 1, 2, and 3-syllable items were almost 100% as expected. At 4 syllables, the average score dropped to just below 80%. At 5 syllables the score fell below 50% (49.55%), reminiscent of the 4 ± 1 WM "magic number". These findings suggest that our WM test was measuring WM capacity as intended. In addition, the boxplot shows that skewness on both tests was minimal.

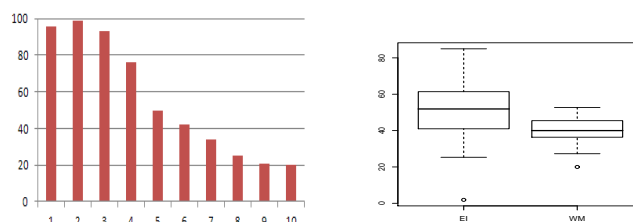


Figure 1: WM scores histogram (left) and boxplots for both tests' scores (right).

We then ran three Pearson correlations across individuals to quantify relationships of interest; see Table 1.

Test	EI	ELC Level
WM	$r=.249, n=40, p=.121$	$r=.130, n=40, p=.430$
EI		$r=.786, n=40, p=.000$

Table 1: Pearson correlations between WM and EI scores and placement level.

The correlations between WM and EI scores and between WM scores and ELC level do not reach significance, though the correlation between EI scores and ELC level does.

Consider the scatterplot graphs in Figure 2: each plot-point represents an individual student who took both tests, modulo overlap. Graph (a) shows that most students have average WM scores (around 40% overall), but have very divergent EI test scores, ranging vertically throughout almost the entire spectrum of English language ability, from 1.9% to 84.97%. We also see students with below-average WM scores but average English EI scores; on the other hand, some students who did not do as well on the EI test have above-average WM scores.

Scatterplot (b) shows that the student at the highest ELC level (5) had a WM score of 41.82%, but was outperformed on the WM test by a student at the lowest ELC grade (0), who had a WM score of 47.27%.

Scatterplot (c) reveals a much clearer correlation, the only significant positive correlation we found: EI versus ELC placement level. Most of the points align much more closely with the diagonal.

Of the three students in Level 0 (the ELC's lowest beginning level), one of them scored the very lowest on the English EI test, correctly repeating only 1.9% of the syllables.

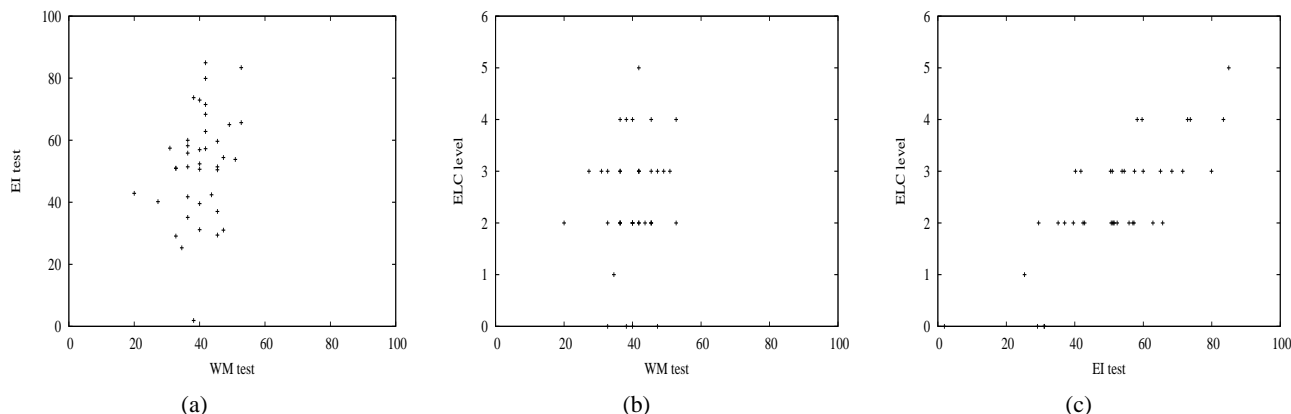


Figure 2: Scatterplots: (a) WM score vs. EI score; (b) WM score vs. ELC level; and (c) EI score vs. ELC level

This student’s working memory score was perfectly fine, at 38.18%. Further, note that the lowest beginning level students could not break past the 30-40% barrier on the English EI test.

With their little-to-no knowledge of English, the lowest-scoring students were unable to repeat anything beyond their WM capacity. It seems reasonable to assume that below scores of roughly 30%, working memory tends to be the primary contributor to the test-takers’ English EI results. Above 30%, knowledge of the language seems to have a greater influence, and working memory proportionately less so. This particular result follows our prediction—that the English EI test would primarily be an assessment of an individual’s linguistic ability, and not an assessment of their working memory capacity—not as a rote memory test.

As mentioned above, most types of second language capacity tests assess explicit knowledge. Our test responds to Ellis’ two criteria for measuring implicit knowledge: (1) the test must be time-constrained, and (2) the test must get at meaning, and, as much as possible, avoid focusing on form (R. Ellis, 2005). Our test was time-constrained: test-takers were only allowed a few seconds to respond. This did not give students a chance to analyze sentence structure or grammar, or indeed rehearse the item before repeating it.

Although we were not measuring actual language chunking in this test, if chunking by meaning enabled higher-level ELC students to perform better on the English EI test, then it may be that this test (or at least some of its items) allowed them to focus on meaning rather than form.

Shorter EI stimuli that fall within the constraints of working memory function more like a nonce syllable working memory test. This is especially true of students at Level 0, who have little knowledge of the language. But lengthier sentences are too long for WM capacity, and perhaps form-focused chunking does not sufficiently recode so many syllables into few enough chunks to retain in working memory. The longest sentences on the test can be remembered and uttered correctly more easily by those who can recode it into its

meaning, which accesses their implicit knowledge of the language. Otherwise these lengthier items become impossible for low-proficient learners to reproduce based on WM alone.

For example, one of the items on our English EI test, “Joe writes poetry,” is short enough that working memory may have a greater correlation to scores on this particular item than the test overall. While due to its length it could be considered a working memory test item, its carefully designed grammatical features make it a test of at least some explicit linguistic knowledge.

Complex, lengthy items such as “When Jim entered the office he was immediately afraid of the uncommunicative boss.” contain too many separate features for a person wholly dependent on explicit knowledge to maintain all the separate elements in STM and/or WM; a round-trip to implicit knowledge representations would be required. If such is the case, then specific items could be used to distinguish between working memory, explicit knowledge, and implicit knowledge of the language.

In summary, the lack of significant correlations between working memory and English EI scores and between working memory and ELC levels, and the significant correlation between English EI scores and ELC levels suggest that there is more to performance on EI tests than working memory capacity. Ellis proposed that EI tests could be developed to test for different types of knowledge, namely, explicit linguistic knowledge and implicit linguistic knowledge. He stated that time-limitations and a focus on meaning were necessary to construct an EI test that measured implicit linguistic knowledge. The test-takers’ errors indicate that using items with specific grammatical features can distinguish between levels of ability. Beginning level test-takers had difficulty with even very short items, if they contained specific grammatical structures beyond their linguistic ability. Long sentences were also difficult or impossible for the beginning-level students, but doable for advanced speakers. These findings produce further research questions. Do utterances that are too long to repeat using average working memory capacity require chunking by

meaning to repeat correctly? If EI performance on longer sentences does access a translation into meaning, then according to Ellis' framework, these types of long utterances also test implicit knowledge. A further question still is, do test-takers even recall hearing those morphemes they were unable to accurately repeat, such as the word-final 3rd person singular "s" (e.g. "runs")?

Conclusions and future work

The findings above indicate a lack of significant correlation between EI and WM scores for the same population. Though a direct causal relationship cannot be made to language use from this result, our measures do lend circumstantial support to the idea that WM testing does not primarily target linguistic ability, and hence directly or even significantly influence EI scores. This is in agreement with the findings of DeKeyser (2005) that elements without meaning (in this case, nonce syllables) draw on memory more than elements of meaning. Though some working memory is involved in language learning and production (Robinson, 2005), it was not shown here to be positively correlated with EI testing. This lends credence to the suggestion that the English EI test is testing linguistic knowledge.

There were a few limitations in the scope of this study, all of which can be overcome with further testing or analysis: (1) The sample size of forty test-takers is relatively small, though the largest we have seen yet for the task at hand. (2) We cannot guarantee that the nonce syllable sequences are meaningless in all of the non-English languages known by test-takers. (3) It would be informative to correlate other metrics (e.g. the participants' OPI scores) to working memory. (4) We have not yet analyzed the effect of the distractor tasks. (5) More detailed analyses of our results could be undertaken including such techniques as structured equation models or ANOVA analyses.

Acknowledgments

We express our appreciation for support from the BYU Pedagogical Software and Speech Technology Research Group, the BYU Center for Language Studies, and the BYU English Language Center.

References

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

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second language competence. *Research methodology in second-language acquisition*, 245-261.

Connell, P. J., & Myles-Zitzer, C. (1982). An analysis of elicited imitation as a language evaluation procedure. *Journal of Speech & Hearing Disorders*(47), 390-396.

Cowan, N. (1996). Short-term memory, working memory, and their importance in language processing. *Topics in Language Disorders*(17), 1-18.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*(24), 87-185.

Cowan, N. (2005). *Working memory capacity*. Hove, East Sussex, UK: Psychology Press.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? a review of issues. *Language Learning*(55), 1-25.

Doughty, C. J., & Long, M. H. (2003). Optimal psycholinguistic environments for distance foreign language learning. *Language Learning & Technology*(7), 50-80.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*(24), 143-188.

Ellis, N. C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The Modern Language Journal*(92), 232-249.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*(27), 141-172.

Erlam, R. (2006). Elicited imitation as a measure of 12 implicit knowledge: An empirical validation study. *Applied Linguistics*(27), 464-491.

Gass, S. M., & Mackey, A. (2007). *Data elicitation for second and foreign language research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hamayan, E., Saegert, J., & Larudee, P. (1977). Elicited imitation in second language learners. *Language and Speech*(20), 86-97.

Jessop, L., Suzuki, W., & Tomita, Y. (2007). Elicited imitation in second language acquisition research. *The Canadian Modern Language Review*(64), 215-238.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press Inc.

Lehman, M. T., & Tompkins, C. A. (1998). Reliability and validity of an auditory working memory measure: data from elderly and right-hemisphere damaged adults. *Aphasiology*(12), 771-785.

McGhee, J. (2010). *Lexical decision reaction times: The effects of ambiguity and bilingualism*. (Unpublished)

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*(63), 343-355.

Morra, S. (2000). A new model of verbal short-term memory. *Journal of Experimental Child Psychology*(75), 191-227.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*(25), 46-73.

Slobin, D., & Welsh, C. A. (1973). Elicited imitation as a research tool in developmental psycholinguistics. In C. Ferguson & D. Slobin (Eds.), *Studies of child language development* (p. 485-489). New York: Holt, Rinehart and Winston Inc.

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