Dear Author,

Here are the proofs of your article.

- You can submit your corrections **online**, via **e-mail** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- For fax submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the **journal title**, **article number**, and **your name** when sending your response via e-mail or fax.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal’s style. Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.
- Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**
- The **printed version** will follow in a forthcoming issue.

**Please note**

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL: http://dx.doi.org/[DOI]. If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information go to: http://www.springerlink.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us if you would like to have these documents returned.
This study measures whether number and type of morphemes in an elicited imitation string results in a greater number of modifications with L2 experience. Rationale is drawn from L2 working memory processing limitations at distinct levels of proficiency. 38 subjects (L2 Spanish university students) comprise three proficiency groups: beginning, undergraduate majors and graduate students. Number of morphemes was varied within each syllable count; and responses were either correct or modified (lexemically/inflectionally as deletions or substitutions). Two way ANOVAs determined significance between mean proportions for each group. Findings indicate that increases in number of morphemes yielded significant differences; and that while the lowest proficiency group produced higher proportions of lexical deletions, the more advanced groups’ modifications were inflectional substitutions.
**Author Query Form**

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the ‘Author’s response’ area provided below.

<table>
<thead>
<tr>
<th>Query</th>
<th>Details required</th>
<th>Author’s response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Please check and confirm the edits made in article title.</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Citation of reference Lust et al. (1986) has been changed as Lust et al. (1987) inorder to match the list. Please check and confirm.</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>References Baddeley (1996), Baddeley (2007), Baddeley et al. (1998), Baddeley and Hitch (2001), Baddeley et al. (1988), Lust (1986), Stockwell and Bowen (1965) are given in list but not cited in text. Please cite in text or delete from list.</td>
<td></td>
</tr>
</tbody>
</table>
Assessing L2 Lexical Versus Inflectional Accuracy Across Skill Levels

Donna E. West

Abstract This study measures whether number and type of morphemes in an elicited imitation string results in a greater number of modifications with L2 experience. Rationale is drawn from L2 working memory processing limitations at distinct levels of proficiency. 38 subjects (L2 Spanish university students) comprise three proficiency groups: beginning, undergraduate majors and graduate students. Number of morphemes was varied within each syllable count; and responses were either correct or modified (lexemically/inflectionally as deletions or substitutions). Two way ANOVAs determined significance between mean proportions for each group. Findings indicate that increases in number of morphemes yielded significant differences; and that while the lowest proficiency group produced higher proportions of lexical deletions, the more advanced groups’ modifications were inflectional substitutions.

Keywords Elicited imitation · Working memory · Spanish · Adult L2 assessment · Morphemes

Introduction

Particular to this inquiry is the measurement of implicit as opposed to explicit L2 knowledge across distinctive proficiency levels by means of oral elicited imitation (EI) testing. Chunking information, at the morphemic level and beyond, becomes more likely with L2 developmental increases given the additional wealth of input and output, together with changes in WM and the interlanguage (IL). According to Bley-Vroman and Chaudron (1994, p. 245), “…If the subjects’ grammar corresponds to the grammar used in producing the string, the imitation is more likely to be correct. Specific inaccuracies may point to specific differences between the subject’s grammar and the grammar of the target string.” Lust et al. (1987, 1996) likewise support this claim. Increases in vocabulary coupled with morphosyntactic advances account
for a greater likelihood to chunk beyond the phonological level and responses should reflect reconstructions indicating the level at which chunking is operationalized for different L2 proficiency groups. Moreover, “the limitations of working memory drive learners to construct higher order chunks out of lower order chunks” (Ellis 2008, p. 469). Ellis (2001) contends that individual differences appear to exist between L2 learners of similar developmental levels/experiential levels; nonetheless, he argues that such differences are small in comparison to differences which exist across distinct developmental levels, especially if such levels are sufficiently different to warrant separation into skill groups. Mackey et al. (2002) are in accord—learners’ developmental stage is a factor in noticing phonological strings in WM. Mackey et al., however, note some exceptions, namely, that learners with less L2 experience, with greater than typical WM capacity, performed comparably or better than did some of their more advanced counterparts. This issue demands further investigation—establishing whether and to what degree proficiency level and increases in morphemic chunking in WM capacity are correlated such that an increase in one is associated with an increase in the other. Hameyer’s (1980, p. 21) claim underscores the need for further research into the types of morphemic modifications across proficiency groups: “…Incorrect inflections and incorrect substitutions are the surest single indicators by which to distinguish proficiency groups.”

In the present study the same EI prompts were used for three L2 Spanish proficiency groups (beginning, advanced undergraduate, and advanced graduate students). Increases in number of morphemes in the prompts were made by adding plural and person inflectional morphemes while keeping lexemes and number of syllables constant. Number of unaltered morphemes in imitations rather than number of unaltered syllables may more adequately measure WM differences across distinct proficiency levels determined by the linguistic level at which chunking is operationalized.

**Background**

Elicited imitation has a lengthy history. Its superiority to other methods is grounded in three rationale: (1) it does not leave production of a particular grammatical structure/lexical item to chance; (2) it can measure implicit knowledge by means of structured responses to stimulus sentences; and (3) it establishes a uniform base of linguistic input for all learner participants. “Elicited imitation may be advantageous over spontaneous speech sampling because it allows greater control over contextual variables than is possible in most naturalistic settings” (Corrigan 1982, p. 223). EI’s initial purpose was to determine the limits of short term memory (STM). STM refers to the ability to store and process a limited amount of information online for a limited amount of time, typically 2 or 3 s (Eysenck 2001, p. 160; Baddeley 2003, p. 675). Miller (1956) determined that the limit of STM is seven units, plus or minus two, although Simon (1974) determined that STM does not go beyond five chunks. Simon’s lower limit could be a consequence of the fact that the composition of the elicited imitation prompts may not have been meaningful in STM. Miller’s (1956) determination that STM consists of a greater number of chunks may represent the number of digits and not the number of chunks. Chunks within STM can be larger than a single digit/phoneme/syllable/morpheme, perhaps depending on the proficiency level.¹ The issue of types and composition of “chunks” has, to date, been given little attention. Lewandowsky and Murdock (1989) found that digits

¹ According to Gathercole and Baddeley (1993, p. 25) “The most convenient and widely used index of this developmental increase is provided by auditory digit span, which is the maximum number of digits that someone can immediately remember and then repeat back in the same order,” which is measured by a prompt provided to the listener, and then having the listener within 3 s articulate what he or she remembers.
are more easily remembered in STM than are linguistic chunks. Erlam’s (2006) findings further suggest that linguistic units that are meaningful are more likely to be remembered in STM/working memory. Since the 1970s, STM has been replaced by working memory, extending its limitations in terms of functionality and capacity. Because working memory is responsible for additional functionality and capacity when compared to STM, its means to hold additional chunks likewise surpasses that of STM (Skehan 1998, p. 50). Establishing the limits of STM and/or working memory lays the groundwork for the use of elicited imitation as a measure of memory units (See Skehan 1998, p. 78).

According to Skehan (1998, p. 44),

short term memory has been replaced by the concept of working memory, a system which still contains rehearsal loops (both phonological and visual) and also a central executive component [the most primary of the working memory (WM) systems] which is concerned with the allocation of a limited amount of attention. In addition working memory contains those records from long-term memory (LTM) which are currently in a high state of activation.

Skehan’s account of working memory appears to be based primarily on Baddeley and Hitch’s (1974) and Baddeley’s (1986) model which has been slightly modified in Baddeley and Logie (1999) and in Baddeley (2000) to include greater influence of LTM on WM and an episodic buffer. In emphasizing the contribution of WM in learning, Baddeley’s model enumerates several primary functions of the central executive system: (1) coordinating information flow between the visio-spatial and phonological loop with LTM; (2) focusing and refocusing attention; (3) manipulating information from the input (suppressing irrelevant data and sequencing relevant data); and (4) directing retrieval of LTM information. Other models of WM include: the connectionist model (Schneider 1999; Just et al. 1996; Engle et al. 1999; Waters and Caplan 1996, 1999); and models which emphasize LTM’s import minimizing that of WM (O’Reilly et al. 1999; Ericsson and Charness 1994; Ericsson and Kintsch 1995; Ericsson and Delaney 1999). For a more extended account of working memory models and their potential effects on learning, especially on L2 learning, see Mizera (2006, pp. 10–22).

Levett applies much of Baddeley’s WM model to issues of L2 learning and proposes that WM plays a primary role in L2 acquisition, especially at beginning and intermediate proficiency levels. Levett proposes several specific L2 related functions for WM: accessing lexical/semantic chunks from LTM; altering LTM chunks to accommodate to the input/message; and assembling LTM chunks and message chunks either simultaneously or in quick succession. Levett’s model makes prominent three factors in WM: recall, modification, and assemblage. According to Lennon (2000, pg), Oppenheinm (2000, pg) and Segalowitz (2000, pg), lexical access is slower and more deliberate in L2 at lower and intermediate proficiency levels since chunks are often smaller (fewer multi-word chunks). As a consequence, greater attentional resources must be expended on organizing the many chunks than on integrating LTM chunks into the message or assemblign/sequencing linguistic chunks once in WM. The latter function, namely, sequencing morpho-syntactic items in WM is likely to be less utilized since attentional resources are expended on LTM retrieval and message coordination, creating an increased likelihood of L2 errors or producing slower serialized articulation (non-automatic speech) (Temple 1997, p. 87).

As mentioned above, Levett (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. (Mizera 2006, p. 87)
Lust et al. (1987, 1996) used EI to determine children’s syntactic abilities. Modifications in children’s responses to EI prompts established the children’s competence at that point in language development. Valian and Aubry (1996, 2005) found that children made a higher proportion of inflectional errors rather than lexical errors beginning at 2;0 to 4;0. Even though Valian et al. did not code some omission of inflections as errors, they still found that inflectional errors were more prominent than were lexical errors. Gathercole and Baddeley (1993, p. 25) have reported that working memory capacity at 4;0 extends to three units and increases at 10;0 to the adult limit of seven. The extent of adult working memory for native speakers of a language extends to sixteen syllables according to Wanner (1974). Whether each syllable consists of a meaningful component (a word, or a free or bound morpheme) was not the objective of Wanner’s study. These theorists attempted to use EI to establish a benchmark for native language competency and to match degree of working memory capacity to age/linguistic experience while acknowledging some existence of individual differences in L1 WM capacity. Their findings indicate that the ceiling for STM/WM increases with age/MLU and with developmental advances.

Elicited imitation as a measure of linguistic competence was used in L1 acquisition to measure linguistic competence at different developmental levels. Smith (1973), Corrigan and Di Paul (1982), Hood and Lightbown (1978) and Keller-Cohen (1981) employed EI to investigate how children construct semantic hierarchies. Potts et al. (1979, pp. 24–26) cast doubt on the reconstructive nature of EI responses by demonstrating that sizable discrepancies exist between linguistic structures under productive control within natural speech and those structures present in imitations. Still other researchers (Slobin and Welsh 1973; Lust et al. 1987, 1996) present strong evidence to the contrary: that reconstructive processes are inherent in L1 EI responses.

Lust et al. (1987) use EI tests to evaluate the state of L1 grammar at early stages. Their findings unequivocally indicate that EI is a reconstructive/restructuring process and not merely verbatim mimicry of the prompt: “EIs were not passive copies or parroting, but active restructurings of the stimulus sentence. Not only were Echo’s EI’s often not identical to the stimulus sentence but the distortions were not random” (1987, p. 285). Lust et al. (1987) provide the most convincing evidence of restructurings given their frequent citations to subjects’ systematic modifications to stimulus sentences. The imitation modifications which Lust et al. (1987) cite (L1s whose MLU ranges from 3.75 to 4.76) can be classified into two types of reconstructions: (1) modifying subordinate clauses to coordinates; (2) reducing redundant lexemes which follow the antecedent by pronominalization of a previously mentioned noun (NP) in the sentence or by deleting the second (redundant) NP/VP altogether. To illustrate the former, Lust et al. (1987) refer to children’s imitations of adverbial subordinate clauses, e.g., using “while” as a systematic reconstruction since their imitations always converted such sentences into coordinate/compound sentences inserting the conjunction “and” between the two independent clauses. Lust et al. (1987) use “restructuring” to mean syntactic reconstruction and even indicate that semantic content is “lost” in the event of syntactic reconstruction. Semantic and syntactic features are assumed to be independent; but reconstruction need not be forced into this syntactic paradigm. Reconstructions can be a consequence of semantic or pragmatic reformulations as opposed to syntactic reconstructions.

Semantic reconstructions can take the form of more amplified meanings, which do not preclude the meaning expressed in the prompt, but which allow for additional temporal/spatial connections between the events of the independent and dependent clauses. A stimulus sentence from Lust et al.’s (1987) test instrument is: Prompt: “Mary was sad while Mary was
playing ball”; Imitation: “Mary was sad and (Mary/she/no subject) was playing ball”. “And” may have replaced “while” not to conform to a structure (coordinate), but to allow for the possibility of simultaneity or sequentiality of the two events—Mary’s sadness and Mary’s ball playing. In the event that the meaning of “while” was unclear to the learner, “and” which is more frequent in the input, may have been substituted to allow for a more general relationship between the two events while maintaining semantic content. Still further amplification includes the possibility that the two events may not be causally integrated, i.e., the state of Mary’s sadness may not be an outgrowth of her ball playing independent of whether the two are simultaneously experienced from the child’s perspective. The L1 learners were obviously reconstructing, but perhaps on a lexical level as Bley-Vroman and Chaudron (1994) suggest, or at a semantic level, which would account for meaning connections between the two independent clauses at a more sentential/ across-sentential level. This semantic restructuring may be useful especially given the absence of spatio-temporal context when elicited imitations are administered. In the absence of contextual cues, L1 learners reconstruct semantically and pragmatically by providing a host of possibilities for interpretation.

To illustrate the second type of reconstruction, deletions of redundant NPs/VPs, Lust et al. (1987, p. 292) cite to reductions of redundant NPs/VPs. “Push the truck and pull the truck” received more accurate responses than did “Push and pull the truck.” Moreover, when VP or NP redundancies were deleted, they were the second mention and never the first mention in the prompt, e.g., “Push the truck and push the car” was reduced to “Push the truck and the car”; but “Push the truck and pull the truck” was never reduced to “Push and pull the truck,” a backward reduction. Lust et al. (1987, p. 294) contend that both forward and backward reductions are available in the adult grammar, and that these forward reductions/modifications are a consequence of a preference for a conjunction structure. Structural templates may not account for increased errors; instead, L1 learners may be operating on the semantic level within working memory to produce a response which is free of contradiction. The latter EI prompt, “Push and pull the truck,” can be perceived as contradictory if interpreted as a single simultaneous event as opposed to sequential events, which appears to be the intended meaning of the prompt.

The above examples taken from Lust et al. (1987) provide the strongest evidence that EI responses are not verbatim reproductions of the stimulus sentence but reflect systematic linguistic operations likely to be unconscious. Such systematic changes demonstrate the state of the learners’ grammar at a given point in development, and not mere parroting or a release of phones from STM in an invariant sequence. These systematic restructurings indicate meaning based repetitions since the learner often does not imitate as an exact replica, but changes the stimulus sentence while maintaining the meaning. Lust et al. unswervingly attribute all reconstructions to be a consequence of syntactic restructuring. A more likely explanation is that in view of the wealth of linguistic input that learners experience, they modify the input semantically and pragmatically to compensate for the lack of context inherent to EI productions. In fact, any failure to match the surface structure of the prompt does not indicate that meaning is lost, or compromised as Lust et al. contend, but that learners merely package the meaning to allow for several possibilities of interpretation when linguistic and extra-linguistic cues are not available. This semantic/pragmatic reconstruction does not illustrate a loss of semantic meaning, but is an indication that the learner, even at this early stage, realizes the effect of the lack of contextual cues on sentence interpretation.

Valian et al. (1996) and Valian and Aubry (2005) likewise used EI to measure L1 competency across developmental levels but from a cognitive performance-based point of departure (WM limitations are responsible for alterations in imitations). In Valian and Aubry (2005), 28 subjects were divided into two groups of developmentally different (based on MLU) L1s.
Valian et al. used two tests, spontaneous speech and EI; and within the latter each stimulus sentence was presented twice, but not successively. “Extensive pilot testing (N = 44) had demonstrated the need for two intervening sentences before asking a child for a second repetition. Otherwise, the child tended to repeat her earlier attempt verbatim” (Valian, et al. 2005, p. 626). Valian et al.’s rationale is founded on the fact that after two intervening stimulus sentences, memory of the first pass of the prompt would have decayed. MLUs were determined from the spontaneous speech samples for developmental categorization purposes.

Valian et al. note a trend that EI MLUs exceeded spontaneous speech MLUs (Valian et al. 2005, p. 629), but percentage of words imitated correctly, as opposed to imitative MLU, was their ultimate competency measure, especially given increased MLUs from pass one to pass two of the stimulus sentences. The issue whether L1s had productive control of structures from either linguistic measure (EI vs. spontaneous data) is unaddressed; hence the meaningfulness of prompts remains in question and responses and imitations which are not direct replicas of their stimulus sentences are not provided as examples of reconstructions. The EI testing instrument consists of 30 prompts testing five distinctive grammatical structures (topic, expletive, infinitive, modal, and past tense), six tokens for each grammatical structure. Although prompt instrument design in Valian et al.’s study is well-formed in that an equal number of tokens are included for each grammatical structure, care was not taken to vary prompt morpheme length for the same grammatical structure, e.g., prompts containing past tense markers ranging from 3 to 7+ morphemes. According to Valian et al.’s coding practice, words, not morphemes, were consistently counted as a single unit but omission of certain words/morphemes (determiners, adjectives, and/or plural markers on nouns) in the imitations were not coded as errors. These practices cast doubt on coding consistency and obscure the function of morphemic length and complexity as a factor in EI production (though morphemic count and complexity may have greater influence on EI performance in more inflected languages than is the case in English).

Elicited imitation is experiencing resurgence in the L2 literature because it is deemed to be a good measure of second language competence, not merely for younger learners but for adults as well. In fact, EI as a measure of WM capacity may serve a more critical purpose in L2 than in L1 development especially for lower and intermediate proficiency learners since lexical retrieval is typically slower and more conscious than for native speakers (Lennon 2000; Oppenheim 2000; Segalowitz 2000) and since chunks are smaller and may consist of lower order linguistic units, e.g., the phonological/morphological as opposed to the lexical/semantic.

Naiman (1974) was one of the first to use EI to measure L2 competence. First and second grade children learning L2 French produced identical structures in the EI experiment as in natural speech (tense markers and object nouns and pronouns) (Naiman 1974, p. 70). Naiman only vaguely alludes to the types of modifications/reconstructions that his subjects made to the EI prompts, and he does not focus on developmental issues–how competency increases with experience. Moreover, since all of Naiman’s EI prompts were 15 syllables in length, the effect of syllable length or, more importantly, chunks (phonemic, morphemic, syntactic, and the like) on WM was not addressed. Although Naiman indicates that 15 syllables “overloaded

2 Actually, many of the subjects would have experienced three passes of the stimulus sentence were their imitations of the first pass accurate.

3 This decay in WM may be a consequence of two factors: (1) a time delay of more than 3 s between first presentation of the prompt and imitation of the second presentation; and, (2) the influence of the 2+ intervening prompts as additional chunks in the input, discarding previously presented chunks or shunting them into other memory systems.

STM” (p. 72) for his subjects, he does not establish its limits for these younger L2s before and after some immersion.

Hamayan et al. (1977), tested slightly older children’s (8, 11) L2 competence of syntactic items, namely, compound and complex structures, and wh-questions. Although mean scores were not provided to compare the effect of L2 experience between the two groups, Hamayan et al. revealed that the older group had just slightly more L2 exposure. The findings indicate no between-group differences on any of the syntactic items, word order and the like. The groups’ experience and exposure were not significantly distinct to find differences, and hence were not developmentally distinguishable. Further, these investigators do not address the effect of increased morphemic length on EI performance or prompt meaningfulness.

Although Perkins et al. (1986) claim that one of their research questions entails measuring whether the difficulty level relates to number of syllables/bound morphemes/number of words/transformations/propositions, it appears that only the latter two were measured, nor does the test instrument measure what it purports to measure. Eight of the 26 grammatical prompts, together with responses, were omitted from the analysis because none of the subjects repeated them entirely correctly. The eight omitted stimulus sentences contain remote vocabulary items which are relatively unfamiliar, even to native speakers, together with the fact that these stimulus sentences extend from 16 to 25 syllables, most of which extend beyond the ceiling of L1 working memory which would overload the WM of even most native speakers. A case in point is: “Each culture developed in a manner dictated by its own needs, sources, and ingenuity.” An additional design defect merits mention—within the 26 reported prompts, the authors fail to hold constant a single variable which could increase complexity while increasing or decreasing the incidence of another. Neither syllable length nor number of lexical items were kept constant while varying number of bound morphemes, or number of transformations, or number of propositions. To illustrate: while the number of prompts in a string can vary from 3 to 25, an equal number of prompts for each syllable length with grammatical complexity variations within prompts of the same syllable length would vastly improve the effectiveness of the measurement instrument by providing additional tokens. An even more serious defect is noted in the fact that reconstructions/meaningful responses are obscured by the coding technique. Binary coding of either wholly correct (an exact replica) or incorrect without coding the type of variation from the prompt (syntactic/semantic/morphemic variations in the form of either substitutions or deletions) does little to isolate which grammatical items are suppressed in WM/deleted and which structures/lexemes are reconstructed while maintaining meaning. If responses reflect a more significant proportion of deletions/substitutions of bound morphemes as opposed to lexical items or the reverse, a determination of at which linguistic level learners at particular L2 proficiency level (phonemic/morphemic/semantic/syntactic) are operating can be demonstrated.

To demonstrate grammatical competence, specifically that of anaphora, from a UG based perspective Munnich et al. (1994) constructed an EI test instrument consisting of grammatical and ungrammatical prompts in four categories– 48 items, 12 in each. Merely twelve subjects from an advanced adult ESL level only were tested. Although implicit WM constraints were considered, in that all prompts consist of 15 syllables, the structure of both the grammatical and ungrammatical prompts is obscure and ambiguous even for native speakers; e.g., “The foreman informed the architect who the engineer called.” The dependent clause, including the anaphoric reference, even in this grammatical prompt, is ambiguous and it is puzzling that these structures would be used as a benchmark for ESL learners. Piloting the test items on native speakers could preclude ambiguities and/or semantically confounding components in the test items. Insurance of the meaningfulness of test items/prompts is paramount since without such insurance subjects’ reconstructions may not be measured—only verbatim recall.
Munnich et al. assume that significantly greater deletions of subject pronouns than object pronouns from ungrammatical prompts was a consequence of subjects’ intent to make grammatical sentences. The upshot of their conclusion is that their participants more often corrected ungrammatical subject pronouns than ungrammatical object pronouns. In line with Bley-Vroman and Chaudron’s (1994, p. 256) observation, Munnich et al. unwittingly ignore the pivotal role of serial effect in the stimulus sentence. A close perusal of their test instrument reveals that subject pronouns are always in the middle of the prompt (at least four syllables back from the concluding syllable) while all object pronouns appear as the last syllable. This design defect—in not varying placement—confounds the factors that contribute to pronoun deletion. Hence their explanation that participants were more likely to correct/reconstruct ungrammatical subject pronouns is called into question—the subject pronouns were corrected not necessarily because of their subject constituency but perhaps as a consequence of their invariant placement. An additional design feature which Munnich et al. failed to include, likewise for purposes of precluding verbatim recall and enhancing reconstructive responses, is incorporating a time delay between stimulus administration and participants’ responses.

The issue of adult L2 reconstructions in Erlam (2006) was carefully integrated as a design feature in her EI test in that following oral administration of the stimulus sentences (grammatical/ungrammatical) a delay was implemented prior to the imitations. Subjects were instructed to judge whether each stimulus sentence was true or false prior to the respective imitation. While the delay in imitations is effectual in precluding parroting and increasing the likelihood of restructuring, the incorporation of truth-value judgments prior to imitations may have served as too great a distractor. Rather than encouraging a focus on meaning and hence increasing reconstructive processes, considerations necessary to judgments incumbent to determining truth-value may have interrupted focus on meaning. To illustrate: truth-value judgments of a sentence containing more than a single clause can produce conflicting considerations in terms of which clause (if not both) is true. In “It is a silly question to ask do a woman need to marry.” (Erlam 2006, p. 475), truth-value judgments may consider either whether a woman needs to marry or whether the question of needing to marry is silly. Moreover, truth-value judgments of ungrammatical stimulus sentences may confound the participants as in the case above. Ungrammatical stimuli may appear less true than grammatical stimuli especially when the ungrammatical structure affects meaning as in the case of English passives, e.g. “The number of Africans with AIDS was increased last year.” (Erlam 2006, p. 476). The passive here indicates that an agent perpetuated the increase rather than indicating that AIDS increased in a particular population. Furthermore, since truth-value judgments are binary there is a 50% likelihood of accuracy; and when the processing load is overloaded or when fatigue interferes with performance after many repetitions, truth-value judgments may fail to facilitate a focus on stimulus sentence meaning. While Erlam implements the time delay factor to create an EI task which goes beyond verbatim recall to measure reconstructions, her use of truth-value judgments may not have produced the desired effect. Erlam assumes that a reconstruction is tantamount to an error since she codes an alternative correct possessive response to be incorrect, e.g. Stimulus sentence: “Princess Diana’s death shocked the world.” Incorrect response: “The death of Princess Diana shocked the world.”

The Study

Thirty-eight undergraduate and graduate L2 Spanish students voluntarily participated in this study. The L1 for all students is English, even for those who are L1 Spanish speakers. Fifteen students belong to the beginning group—they experienced a mean of 1.4 semesters of univer-
sity foreign language coursework; and a mean of 2.3 courses in high school. This beginning
group did not experience any other non-classroom exposure to Spanish. Seventeen students
belong to the advanced undergraduate group—they experienced a mean of 3.7 university
courses (2 of which constitute Spanish major courses); and a mean of 2.8 courses in high
school. Approximately one fourth of these advanced undergraduate students studied in the
target culture for one semester. Six students belong to the advanced graduate group; they
experienced a mean of 2.5 graduate level Spanish courses, 10 Spanish major undergraduate
courses, 2.6 beginning/intermediate undergraduate target language courses and 4.0 Spanish
courses at the high school level.

A set of 24 oral elicited imitation prompts in Spanish were developed from vocabulary
and structures already presented from students’ text (“Vistas”) and were evaluated by two
native Spanish speakers. As a consequence of their review, four were modified lexemically to
preserve a more natural linguistic stream. These 24 prompts were then piloted on 4 students
from each level. All students heard the same recording (24 prompts) produced by a female
native Spanish speaker who articulated slowly and clearly. The order of the prompts with
respect to syllable length and morphemic complexity was randomly sequenced; but, the
order remained constant for all subjects. At minimum two prompts separated prompts with
identical lexemes according to V alian and Aubry’s (2005) design, e.g., “Entiende el titulo
extranjero” (He/she understands the foreign title) was presented at least two prompts before
or after “Entiendes los titulos extranjeros” (You fam. understand the foreign titles).

All of the prompts consist of four words: a determiner, noun, adjective, and present tense
verb with differing syntactic sequences altering the placement of the verb. The adjective
always followed the noun, which is the typical sequence for descriptive adjective placement
in Spanish. Subject pronouns were excluded, which conforms to typical Spanish patterns, the
affix on the verb is the only person marker in those instances. Object pronouns were likewise
excluded. Number of syllables and number of morphemes varies; but, the number of words
does not. Syllable length ranges from 7 to 11 with 8 prompts at each syllable count.

Prompts of additional syllable length (13, 15, 17) were likewise used for the two more
advanced groups but are not included given the lack of data for the beginning group. This
author chose 11 syllables as the ceiling since one-third of the students’ L2 level is below
intermediate. Although Erlam extended syllable length to 18, her subjects’ L2 level was at
minimum intermediate. Failing to successfully imitate a large proportion of the prompts, were
they to consist of more than 11 syllables, may have resulted in a plethora of deletions for the
less advanced groups. Lexemes were selected from the beginning students’ L2 textbook to
ensure some familiarity for the L2s at the lowest level, while ensuring meaningfulness to the
more advanced learners.

Syntactic features are in keeping with the accustomed scheme of the least experienced
group, as not to depress performance and to further ensure the prompts’ meaningfulness.
Nonetheless, the syntax was varied, allowing for the rather free sequence of SVO. Although
Spanish is an SVO language, syntactic variations, such as subject following verb, or object
preceding verb, are not uncommon, e.g., “La carne deliciosa recomienda.” (He/she rec-
ommends the delicious meat). All sentences are declarative and half include direct objects.
Determiners and nouns constitute subjects within ten of the prompts; the subject is marked as
a verb inflection in the remaining fourteen. Placement of the verb within the string was varied
(12 verbs at the conclusion, and 12 at the outset), not uncommon in Spanish given its flexible
syntactic variation. The latter was implemented to ensure that the same constituent type
did not always appear as the first/last component of the string to guard against confounding
constituency type with serial effect. Bloom’s (1990) account indicates that a recency effect
exists in the STM loop such that memory of the latter unit is more accurate, militating in favor
of sequence variation of SVO. Absent this variation, substitution or deletion of a particular
constituent is confounded with effects of recency/primacy.
Singular forms contain zero number morphemes since the absence of the plural morpheme
means singular. The number of morphemes in the prompts ranges from 6 to 11; the differential
in each pair is typically four morphemes (see “Appendix” for the twenty-four prompts used
in this study). Half the prompts contain either a second person “-s” or the “-n” plural person
marker on the verb, 5 and 7 tokens, respectively, while the remaining 12 prompts are singular
third person forms. All verbs with singular subject NP’s accordingly were third person
singular present tense; all verbs in prompts for which subject NP’s were plural were present
tense third person plural, or present tense second person singular to increase the number
of morphemes without increasing syllable length. The present tense\textsuperscript{5} was used exclusively
since any other tense was unfamiliar to a majority of the least proficient group. The person
morphemes on the verb (“-n”, “-s”) were chosen to increase the morpheme length since
each consists of one phoneme only and since their meanings are less ambiguous than the
first person affixes “-o” and “-mos”. The first person singular “-o” is used in the past tense
perspective aspect as a stressed vowel to refer to third person; the first person plural “-mos”
contains an additional syllable and it has varied uses. Consequently, they were excluded.
Great care was taken to have sufficient tokens of particular types of inflectional morphemes
to compare lexical to inflectional performance. The types of inflectional morphemes were
scrupulously controlled within each prompt to compare across prompt performance for types
of inflectional accuracy. Only one person inflection was included in each prompt since only
one verb appeared within each, independent of syllable length or number of morphemes.
Half of the nouns are masculine, and half feminine, and half of each of those end in “-e”
which does not distinguish gender. Nouns which end in “-e” were deliberately chosen
since gender is not apparent from the noun inflection and hence could not be carried to
the adjective and/or determiner. Likewise, the number of gender morphemes was controlled
across prompts—3 per prompt; while number morphemes amount to 4 per prompt, including
determiners, nouns, adjectives, and verb affixes. This design was developed to discern whether
participants’ performance was a consequence of phonemic STM or deductions made from
constituent heads and imputed to other members of the constituent.

To measure the possible effect of morphemic complexity on working memory across L2
developmental levels, morphemic complexity was varied within the same syllable length.
Four prompts (half) within each of the syllable length groups represent singular nouns and
third person singular verbs, while the other four include plural nouns and additional person
morphemes on the verb—“-s” or “-n” (second person singular, third person plural, respec-
tively). In fact, the lexemes of the singular strings appear twice, but with plural markings on
determiners, nouns, and verbs (except for the second person “-s”). While including the same
lexemes in two prompts (one plural, one singular) may have had minimal practice effect, such
is outweighed by the advantage of maintaining the same lexemes while increasing morphemic
load. Such instrument design can measure the effect of increased numbers of morphemes in
the memory stream without confounding increases in the number of morphemes with changes

\textsuperscript{5} The present tense in Spanish as in English is realized as a zero morpheme; hence, it is devoid of an
individualized phonetic/phonic representation. For example, the “-s” on the second person singular present
tense verb has three meanings: second person, singular number, present tense; but present tense is not realized
as a distinctive morphemic form. (For further elaboration, cf. West 2010b) Both the singular and the plural
prompts contain gender morphemes on determiners, nouns and adjectives. Gender on the noun for four of
the nouns (likewise reflected in four of the plural prompts) ends in “-e” and their gender was not apparent
(see “Appendix”). Consequently, there is less redundancy for gender morphemes in those eight strings. Such
provided one fewer morpheme in those strings, which permits greater variation of number of morphemes.

\copyright Springer
in lexical use. Using different nouns and adjectives, and at the same time increasing numbers of inflectional morphemes, fails to isolate morpheme length as the only factor contributing to accuracy in reproducing the string—one lexeme may be more common, more meaningful to the learner than is another. A second advantage of “a second pass” of the lexical items within the stimulus sentences rests in part on Valian et al.’s (2005) double presentation design and in part on the rationale that limited memory resources together with previously stored articulatory information from recent LTM can enhance meaningful recall. The possibility of a practice effect is minimized by the separation of the two like stimulus sentences from one another in the prompt presentation sequence—a minimum of three prompts were presented between the two lexically identical prompts in keeping with Valian et al.’s (2005) design.

Each student heard the same instructions in English; students were informed of the number of prompts and were instructed to repeat exactly what they heard and as much as they heard. They were likewise instructed that they would be asked to answer a short question pertaining to the meaning of each prompt before repeating the prompt. The questions ranged from 2 to 3 short words; the subjects were told that their answers needed to consist of only one to two words (e.g., Prompt: La niña buena llega. (The good child arrives.) Question: ¿Quien llega? (Who arrives?) Typical Response: La niña. (The child.)). As a consequence of the intervening question and response to that question, the interval between the presentation of the EI prompt and the response was a minimum of 3 s. If a subject failed to repeat more than two entire prompts, the data were excluded from analysis. The intent in having students imitate between 2 and 3 s after presentation of the respective prompt was to ensure for meaningfulness by attempting to preclude the use simply of the phonological loop in the imitation in light of Eysenck’s (2001) and Baddeley’s (2000) findings that the upper limit for storage of working memory is 2–3 s.

A coding scheme accounting for correctness/incorrectness of morphemes and syllables was employed. The coding scheme likewise accounted for whether a morpheme/syllable was deleted or substituted and whether the substitution was lexical or inflectional. Data were coded within 2 months of collection. A second individual was trained extensively in the use of the coding scheme. In fact, the primary investigator and the independent coder simultaneously coded 40 % of the data; and agreed on 95 % of their coding decisions.

Results
The procedures described in this study were successful in eliciting imitations from L2s of differing proficiency levels. A repeated measures 2-way ANOVA was conducted to test for differences in error types (Inflectional Substitutions, Lexical Deletions, and Lexical Substitutions) across the three proficiency groups. Proportions represent means for each proficiency level; and means were derived from the total number of errors within the three categories: Inflectional Substitutions, Lexical Substitutions, and Lexical Deletions. Inflectional Deletions were not included as a category, since it is infrequent that suffixes in the form of inflections are deleted, leaving the root bare. The main effect for proficiency was not significant $F(2, 35) = .24, p = 0.80, \eta^2 = .01$. However, the main effect for error type was both significant and large $F(2, 70) = 25.83, p = 0.001, \eta^2 = .43$. As reflected in Fig. 1, all three proficiency groups produced a greater proportion of lexical deletions, as compared to the proportions of inflectional errors and lexical substitutions (see Fig. 1 for the standard deviations). However, the significant and moderately large interaction between the proficiency and error type, $F(4, 70) = 2.94, p = 0.026, \eta^2 = .14$, suggests that the pattern of errors differed across proficiency group.
The hypothesis that lexical errors in the form of substitutions and deletions were more likely to be produced by the beginning proficiency group is not supported since $p > 0.05$ (Fig. 1). It is interesting to note however, that proportions across groups show a downward trend in degree of lexical errors indicating some greater likelihood of lexical errors at the beginning proficiency level, although no significant differences were found between the groups ($p > 0.05$).

The hypothesis that proportions of inflectional errors including substitutions and deletions increase with proficiency level is partially supported since a significant difference exists between the beginning proficiency groups and that of the undergraduate and advanced groups ($p > 0.05$) (Fig. 1). No significance was found between the advanced undergraduate and the more advanced graduate proficiency groups, or between the beginning and the graduate advanced proficiency groups, although the proportion of the latter group indicates some additional inflectional errors with respect to the beginning proficiency level.

The hypothesis that morphemic complexity depresses performance for the less advanced proficiency groups is strongly supported (Fig. 1). A significant difference was found between the beginning proficiency group and the advanced undergraduate group as well as between the beginning proficiency group and the advanced graduate proficiency group ($p > 0.05$). Although the proportion of morphemic substitutions and deletions for the undergraduate

---

Fig. 1  Proportion of lexical, inflectional, morphemic, and syllabic errors (7, 9, 11 syllables collapsed)

---

6 The significant $p$ values remain so even when Bonferroni corrections were made. Bonferroni corrections indicate the following: lexical errors (deletions, substitutions)—beginners to intermediate & advanced, $p > .004$ (uncorrected), $p = 0.008$ (corrected); inflectional substitutions—$p = 0.019$ (uncorrected), $p = 0.038$ (corrected). Linear effect with main effect for proficiency in 2 way ANOVA was $p = .001$ (uncorrected) and $p = 0.003$ (corrected).
advanced group is higher than that of the advanced graduate group, the difference is not significant.

The hypothesis that the lower proficiency groups will produce a higher proportion of syllabic errors than the more advanced proficiency group is unsupported—$p > 0.05$, while the hypothesis that within groups syllable length would have no effect is confirmed—$p > 0.05$ (Fig. 1). Nonetheless, a rising trend exists for the beginning proficiency group with increased syllable length, although the differences are not significant.

Discussion

The sentential position of certain parts of the NP constituent (adjective determiner) may have been confounded with issues of serial order since although the adjective position was varied somewhat from middle to end of string and determiner order from beginning to middle of string, the adjective never appears at the outset, nor the determiner at the conclusion. Meaningful variation could not be orchestrated in Spanish, especially with respect to determiners, since their placement is relatively fixed. The less varied placement of determiners and adjectives may have influenced the lowest proficiency group moreso as a consequence of greater WM constraints. As discussed, access from LTM into WM is slower and more deliberate (Lennon 2000; Oppenheim 2000; Segalowitz 2000) given that learners of lower proficiency levels likewise have experienced less L2 practice. Hence, they would be less likely to develop “higher order chunks out of lower order chunks” (Ellis 2008, p. 469) as do L2s with greater experience in the target language. Further investigation should disentangle lexical deletions from lexical substitutions and which substitutions are meaningful to show which lexical errors are more prevalent for lower as opposed to more advanced L2s. Additionally, coding for phonemic errors (substitutions/deletions) as opposed to morphemic errors with respect to particular lexemes likewise should be undertaken to demonstrate more definitively whether phonological as opposed to morphemic errors are more likely for lower proficiency levels than for higher levels. Such analysis could isolate the level at which L2s are chunking at distinct developmental levels. This analysis focused on morphemic chunking across proficiency groups and indicates significantly greater morphemic errors on the part of the lowest proficiency group. Many morphemic errors include phonemic ones since Spanish is an inflectional language; and an alteration/deletion in a phoneme likewise represents a morpheme alteration/deletion. As discussed below, the lack of significance in inflectional errors between the lowest and highest proficiency groups may be a consequence of two factors: (1) a ceiling effect/fewer differences between the undergraduate advanced and the graduate advanced groups, and (2) the possibility that the lower group’s proportion of inflectional errors may not reflect phonemic errors made within certain morphemes. The latter, failure to consider a morphemic error in the event of a phonemic error, may have deflated the proportion of morphemic errors made by this lowest developmental group. The more advanced groups appear to make fewer phonemic errors; hence deflation of inflectional errors would not be an issue. The issue of degree of attention to the prompt and to which morphemes in the string is perhaps another factor which explains the significance in inflectional errors between the lowest and the advanced undergraduate, but lack of significance between the two more advanced groups, and between the lowest and highest proficiency groups. The attention of more advanced learners may not have been as riveted on the sentence stimuli in view of their short length, and simple meanings and structures (however, necessary for the lowest proficiency group). Longitudinal data might provide some indication of whether attentional issues account for inflectional errors in that each participant’s degree of produc-
tive control over particular inflections is more likely to be measurable. Furthermore, this cross-sectional study could have included additional implicit language testing methods (not merely EI) to determine the productive control of each subject over particular morphemes such as spontaneous speech.

An alternative explanation for the significantly fewer inflectional errors of the beginning learners is difficulty noticing and storing inflections in WM. As a consequence, the beginning group may have deleted inflections more often than substituting them or producing them accurately. In sum, the significantly lower proportion of inflectional errors on the part of subjects of the lowest proficiency level appears to be a consequence of two factors: (1) greater dependence on verbatim recall rather than on reconstructive processes and (2) the lack of WM resources at early developmental stages to notice/call up from LTM morphemic units within a string, especially affixes. This latter rationale is in line with Ellis’ (2006) claim that bound morphemes, especially suffixes, are less noticeable as a consequence of their placement and their abbreviated phonetic duration, e.g., “-s” referring to number within the NP or person within the VP consists of but one phonemic unit and often appears in the middle of a string. Ellis’ argument fails to explain the developmental sequence of increased inflectional errors with proficiency level, since presumably according to Ellis, L2 noticing skills would increase with WM maturation, which would produce findings contrary to the present findings. In other words, difficulties noticing affixes independent of proficiency level cannot explain how it is that more advanced L2s would make significantly greater inflectional errors, in view of increased L2 experience, IL advances and increases in WM capacity. Were Ellis’ contention amplified to incorporate WM maturation and developmental advances in linguistic control, it could account for how failure to notice inflections at more advanced proficiency levels leads to increased errors and reconstructions, while less proficient L2s lack the developmental resources to notice inflection, hence their significantly lower proportion of inflectional errors.

Likewise, Jarvis’ (2000, p. 246–7) conjecture that negative interlanguage transfer is more operational at lower levels than at higher levels of proficiency fails to account for the significantly lower proportions of inflectional errors on the part of the lowest proficiency group and the increase in proportion of inflectional errors for the more advanced proficiency groups. Were transfer a primary factor in error production, the reverse findings would have materialized. The influence of negative interlanguage transfer, especially when the L1 and the L2 are radically distinctive (Kellerman 1995, p. 126; Stockwell et al. 1965) is likewise implausible if applied equally to all linguistic levels of control (the phonological, morphological, semantic and syntactic), but its plausibility increases on the grammatical level. For example, while person inflections in English (the L1) are lexicalized, they typically appear as an affix in Spanish (the L2); and gender inflections, although grammaticalized as inflections in Spanish (the L2) have no grammatical status in English (the L1), illustrating on the grammatical level how the two linguistic systems greatly differ. Kellerman’s argument has less weight on the phonological level than on the semantic/syntactic level since the former is more finite, more abbreviated and is experienced fully in the input at early developmental stages. The development of morphemic/semantic/syntactic skills, however, requires the retrieval and application of meaning-based regularities from the IL, which become operational at more advanced levels of proficiency. Noticing phonological regularities may be operational at early developmental stages with little reference to morphemic/syntactic structure necessary to noticing inflections.

A more plausible explanation of increased inflectional errors at more advanced levels of proficiency is drawn from the particular level of linguistic (phonemic/morphemic/semantic/syntactic) control that L2s of different proficiency levels use in WM and impose on the
input, in this case the stimulus sentences. Given the paucity of L2 experience for beginning learners in terms of input and productive output together with limitations in WM functionality, the lower proficiency L2s appear likely to have drawn primarily upon the phonological attributes in the stimulus sentence as opposed to accessing semantic/syntactic regularities from their LTM/interlanguage. Moreover, the less developed interlanguage of the lowest proficiency L2s may account for a reliance on the phonetic distribution of the stimulus sentence rather than on reconstructing its morphological/semantic/syntactic components. It is this factor (reconstruction) which appears to be responsible for the modifications made to inflections within the stimulus sentences by the more advanced L2s. The more advanced L2s made increased inflectional errors (person/gender/number) as a direct consequence of reconstructing stimulus sentences on the morphological level and perhaps beyond; otherwise, alterations, especially coherent inflectional substitutions, would not have increased with proficiency level. The factor of which level of linguistic control (phonological, morphological, et cetera) is operational in WM appears to account for significant performance differences between the beginning and undergraduate advanced groups but not between the two advanced groups. Increased inflectional errors/reconstructions on the part of the more advanced proficiency groups are a consequence of two factors: (1) regularly operating beyond the phonological level of linguistic control in WM such that inflectional morphemes are noticed and (2) increased negative interlanguage transfer at the morphological/syntactic level particularly where the language systems differ. These rationale appear to account for the significant difference in mean proportions between the lowest proficiency group and the advanced undergraduate group, and may likewise account for the fact that significance in mean proportion of inflectional errors was not found between the two more advanced proficiency groups. In view of more efficient WM processing and reconstructive processes at higher linguistic levels, a ceiling effect appears to be reached at the undergraduate advanced level, accounting for the lack of significance between the two more advanced proficiency groups.

Although proportion of inflectional errors/reconstructions increased with proficiency level, overall morphemic errors decreased with proficiency level. While this appears at first glance to be contradictory, findings from Hypothesis 1 demonstrate how it is that the lowest proficiency group could make significantly greater morphemic errors while at the same time producing significantly lower proportions of inflectional errors. Findings from Hypothesis 1 indicate somewhat higher (but not significantly) proportions of lexical errors on the part of the lowest proficiency group compared to proportions of lexical errors for the more advanced proficiency groups. What appears to account for the beginning group’s significantly greater morphemic errors is their increased lexical errors. While the more advanced L2s produced significantly higher proportions of inflectional morphemic errors than did the beginning L2s, their production of overall morphemic errors was significantly lower. The significant difference between the beginning and undergraduate advanced groups validates the primary claim of this study that it is the number of morphemes within the string which operates to distinguish proficiency groups, especially obvious between the lowest and more moderately proficient groups. The fact that significance was not found between the undergraduate and the graduate advanced groups is likely to be a consequence of two factors (independently and/or combinatorially): (1) Greater proficiency similarities between the two groups, and (2) The lack of differentiation between deletions and substitutions as morpheme errors. It may

7 For a more detailed explanation and rationale for the validity of this argument and its particular application to L2 Spanish, cf. West (2010a).
be the case that with increased proficiency level, L2 errors consist more of substitutions than deletions.

Since morphemic errors include inflectional (gender, number, person, theme vowel), derivational and lexical substitutions or deletions and for purposes of this hypothesis a differentiation is not made between them, the more advanced groups may be more likely to make substitutions than deletions. Their familiarity with the language permits them to make meaningful substitutions/reconstructions. In fact, subjects of the two more advanced groups substituted more sensically than did subjects of the beginning group, e.g. substituting a noun having the same syllable structure, gender, and number for one which they could not remember, “entiende pájaro extranjero” (He/she understands the foreign bird) for “En-tien-des-los-tí-tu-los-ex-tran-je-ros,” (You fam. understand the foreign titles) or “pide el pescado delgado” (He/she orders the thin fish.) for “pi-de-el-pe-sca-do-sa-la-do.” (He/she orders the salty fish). The beginning group was more likely to delete a morpheme, typically a free morpheme, as in “entiende título,” “entiende extranjero” or simply “extran-jero.” It appears that the higher the linguistic level of control in WM, the more likely the learner is to substitute meaningfully where gaps in memory exist. When more advanced learners operate on the morphemic level and beyond, they are able to recover/hold meanings in WM while retrieving reasonable substitutions from the IL into WM for those morphemes temporarily suppressed/not readily retrievable. This operation on the part of more advanced L2s materializes simultaneously with maintenance of the phonological and syllabic structure likewise within WM. In contrast, beginning L2s are less likely to maintain syllable length since they appear to delete morphemes/lexemes and may not simultaneously hold phonological information in WM while accessing a suitable substitution from the IL.

Although no significance was found in proportion of error production for any of the groups with an increase in number of syllables (7, 9, 11) within the stimulus string, the proportion of syllabic errors of the lowest proficiency group indicates an upward trend—errors increasing with syllable length. This trend shows that although increases in the number of syllables may have been a factor in increased error production early on in the learning process, afterward, at higher levels of proficiency, the number of syllables had no effect. No trend is apparent for either of the two more advanced proficiency groups. While findings addressing Hypothesis 3 demonstrate that generally morphemic errors increase with number of morphemes, findings addressing Hypothesis 4 show no such positive relationship—errors do not increase with additional syllables in the stimulus string. It is the number of morphemes within the stimulus sentence, and not the number of syllables which is associated with increased morphemic errors, suggesting that the number of meaning-based units affects error production. Nonetheless, this instrument design practice may not be as easily applicable to languages that are not inflectional in nature, such as English, since an increase in morphemes is more likely to likewise increase the number of syllables. Number of morphemes within the stimulus string must be isolated as a factor which impinges on WM and performance. In the Romance languages, where gender and number (or person and number) markers are often within the same syllable, the effect of number of morphemes on performance can be adequately measured without increasing number of syllables.

Conclusion

This study incorporates viable methods to resolve three outstanding issues in ELI test development. The first entails a suitable instrument design to compare distinctive L2 proficiency
levels; the second incorporates a delay between prompt presentation and response, while the third provides an alternative framework to determine prompt meaningfulness. In light of the design features of this study, construction of EI tests to evaluate L2 performance across proficiency levels should incorporate the following: (1) using short, syntactically uncomplicated stimulus sentences; (2) varying number of morphemes within stimulus strings while keeping lexemic and syllabic factors constant; (3) rotating the position of distinctive grammatical classes within prompts to guard against serial effect in WM, and (4) ensuring that a similar opportunity count is supplied within the stimulus string for repetition of inflectional morphemes, e.g., person/number/gender. EI design should likewise incorporate a non-artificial time interval between the presentation of each stimulus sentence and its corresponding repetition (which does not detract from prompt meaningfulness) to preclude verbatim repetition which may be devoid of meaning. The issue of meaningfulness of prompts is paramount; and to further ensure their meaningfulness (which likewise non-artificially fills the 3-s interval between prompt administration and response) a short question requiring meaningful prompt processing rather than truth value judgments is preferable.

Findings from this study suggest the need to use short and/or simple syntactic/morphemic structures within the stimulus sentences, such that performance of the beginning and intermediate proficiency groups can be adequately compared to that of the more advanced proficiency groups. Although some motivational/attention-based fallout can surface for the more advanced groups with this approach, without it performance comparisons between beginning L2s and more advanced L2s are unlikely to be measurable. Hence, including more than 12 morphemes in a single stimulus string is likely to result in numerous/frequent deletions especially for the lower proficiency group/groups, despite the greater challenge/greater interest which additional morphemes within such strings might engender for more advanced L2s. Increasing the number of morphemes while holding the number of syllables in the string constant is a practice in need of further implementation.

The findings from this study further suggest that the number of morphemes in the stimulus string is a primary factor in distinguishing the proficiency groups, not the number of syllables. More specifically, findings from this study indicate that an increase in number of morphemes had the most dramatic effect on the lowest proficiency group; but inflectional errors were higher for the more advanced L2s. What requires further investigation is types of morphemic “errors”, i.e., whether derivational or inflectional modification is more likely to contribute to production alterations, as well as which kind of inflectional modification contributes more heavily (person, number, gender). L2 maturational factors may influence the type of “error” produced such that beginning L2s produce greater proportions of within constituent alterations (gender, number), and more advanced L2s produce a higher proportion of across constituent alterations (person). Such analysis could uncover more dramatic performance differences between lower and higher proficiency L2s consequent to morphemic factors than was evidenced in this study.

Appendix

Stimulus Sentences
<table>
<thead>
<tr>
<th>Spanish Expression</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>El viaje extensor encanta</td>
<td>The lengthy trip is enjoyed</td>
</tr>
<tr>
<td>Comen los postres ricos</td>
<td>They/you (pl.) eat the delicious desserts</td>
</tr>
<tr>
<td>El vinagre morado alimenta</td>
<td>The wine vinegar is nourishing/nourishes</td>
</tr>
<tr>
<td>Los vinos tintos caen</td>
<td>The red wines fall</td>
</tr>
<tr>
<td>Toma la leche fría</td>
<td>He/she drinks the cold milk</td>
</tr>
<tr>
<td>Limpias las ventanas antiguas</td>
<td>You (fam.) clean the old windows</td>
</tr>
<tr>
<td>Pide el pescado salado</td>
<td>He/she orders the salty fish</td>
</tr>
<tr>
<td>Entiendes los títulos extranjeros</td>
<td>You (fam.) understand the foreign titles</td>
</tr>
<tr>
<td>La noche oscura regresa</td>
<td>The dark night returns</td>
</tr>
<tr>
<td>Los vinagres morados alimentan</td>
<td>The wine vinegars nourish</td>
</tr>
<tr>
<td>La carne deliciosa recomienda</td>
<td>He/she recommends the delicious meat</td>
</tr>
<tr>
<td>Come el postre rico</td>
<td>He/she eats the delicious dessert</td>
</tr>
<tr>
<td>Toman las leches frías</td>
<td>They/You (pl.) drink the cold milk</td>
</tr>
<tr>
<td>La niña buena llega</td>
<td>The good girl arrives</td>
</tr>
<tr>
<td>Los viajes extensos encantan</td>
<td>The lengthy trips are enjoyed</td>
</tr>
<tr>
<td>El vino tinto cae</td>
<td>The red wine falls</td>
</tr>
<tr>
<td>Pides los pescados salados</td>
<td>You (fam.) order the salty fish</td>
</tr>
<tr>
<td>Limpia la ventana antigua</td>
<td>He/she cleans the old window</td>
</tr>
<tr>
<td>Repite las palabras españolas</td>
<td>You (fam.) repeat the Spanish words</td>
</tr>
<tr>
<td>Entiende el título extranjero</td>
<td>He/she understands the foreign title</td>
</tr>
<tr>
<td>Las carnes deliciosas recomiendas</td>
<td>They/you (pl.) recommend the delicious meat(s)</td>
</tr>
<tr>
<td>Las niñas buenas llegan</td>
<td>The good girls arrive</td>
</tr>
<tr>
<td>Repite la palabra española</td>
<td>He/she repeats the Spanish word</td>
</tr>
<tr>
<td>Las noches oscuras regresan</td>
<td>The dark nights return</td>
</tr>
</tbody>
</table>

References


