An Item/Order Tradeoff Explanation of Word Length and Generation Effects.

Liam Hendry

Gerald Tehan

University of Southern Queensland, Toowoomba, Australia

Hendry, Liam and Tehan, Gerald (2005) *An item/order tradeoff explanation of word length and generation effects.* Memory, 13 (3 and 4). pp. 364-371. ISSN 0965-8211

Authors' final manuscript corrected version

Mailing Address:
Liam Hendry
Department of Psychology
University of Southern Queensland
Toowoomba, 4350
AUSTRALIA
Phone number: + 61 7 4631 2376
e-mail: hendry@usq.edu.au

Abstract

The item-order hypothesis suggests that under certain conditions increased item processing can lead to deficits in order processing, and that this produces a dissociation in performance between item and order tasks. The generation effect is one such example. The word length effect is seen as another instance where this tradeoff might be observed. The following experiments compare word length and generation effects under serial recall and single item recognition conditions. Short words are better recalled than long words on the serial recall task but long words were better recognised than short words. The results are consistent with the item-order approach and support a novel explanation for the word length effect.

An Item/Order Tradeoff Explanation of Word Length and Generation Effects.

The word length effect is the finding that lists containing short words are better recalled in immediate serial recall than lists containing long words (Baddeley, Thompson & Buchanan, 1975). This effect is a central feature of what have come to be known as trace decay plus rehearsal (TDR) models of immediate recall (Brown & Hulme, 1995). The major assumptions underlying these models have recently been sufficiently challenged (for a review see Nairne, 2002) to warrant an alternative theoretical account of the word length effect. The current study proposes such an account based on a tradeoff between item and order processing; an explanation that is currently popular in explaining such long-term memory phenomena as the generation effect (Nairne, Reigler & Serra, 1991), the perceptual interference effect (Mulligan, 1999), and the word frequency effect (DeLosh & McDaniel, 1996) among others.

The current study therefore has two main objectives: To introduce an effect in which such an item-order tradeoff has been observed (the generation effect) into the short-term domain, and additionally, to employ the methodology used to demonstrate the item-order tradeoff to show that the word length effect will show a similar dissociation in recall between tasks that utilise order information and tasks requiring the use of item information.

In a typical generation experiment, participants are presented with two types of words. Some of the words are intact (control items), and are read aloud by the participant. For the experimental items, a cue is typically provided, and participants must generate an appropriate response. Memory is then compared for the control and generated items.

The direction of generation effects seems to be dependent upon the type of recall test used. For example, Nairne et al. (1991) presented participants with 24 trials each consisting of eight unique items. On half of the trials each of the eight items on a list were presented with one letter missing and participants were required to generate a word from the fragment. The remaining trials were presented with intact words that were read aloud. Each list was followed by a 30 second distractor activity before participants were asked to respond. A recall cue specified one of two recall options. If a line of asterisks was presented, participants were instructed to not respond and to prepare for the next trial. However, on the non-asterisk trials, the items from the list were re-presented in a random order and participants were requested to put these words into the original presentation order. On this order reconstruction task the read words were better recalled in order than the generated words.

Following this phase of the experiment, a surprise single-item recognition test was given to participants to test for memory of the items that had been presented earlier on the asterisk trials. The recognition test comprised both target and distractor items and on this test, the generated items were better recognised than the read items.

These results clearly demonstrated that the generate/read manipulation dissociated on order reconstruction and item recognition tasks. Nairne et al. argued that the order reconstruction task primarily measured order information and that the recognition task measured item information, and explained the dissociation by arguing that the generated items received more item-specific processing at the expense of order processing, which became evident on an item task. The read items

received more order processing at the expense of item processing, which was apparent on an order task.

It is possible that the word length effect may be another example of the item/order tradeoff. Our assumption is that is that because of rapid presentation rates long words take more time to identify than short words. As a consequence more time is available for processing order information for short words. However the additional time required by long words means that they receive additional item processing. If this is indeed the case, then the typical short word advantage that is observed on immediate serial recall (an order memory task) should reverse when an item recognition task is utilised. Moreover, if the word length and generation effects have the same foundations, then similar patterns of effects should be apparent across variations in the serial recall task.

Experiment 1

Method

Participants. 119 introductory psychology students from the University of Southern Queensland volunteered to participate in the experiment, in return for which they were given course credit, or a ticket in a raffle for cash prizes ranging from A$20 to A$200. Of these, 20 students participated in each of five groups, and 19 participated in the immediate/word length group. Ages ranged from 18 to 49, with equivalent balance in each group.

Materials. Experiment 1 comprised six independent groups, three of which tested the word length effect under conditions of immediate serial recall, delayed recall and irrelevant speech. For the three groups involving the word length effect, two word-pools were created from the MRC Psycholinguistic Database (Quinlan, 1992) comprising 120 short and 120 long words. Short words were all monosyllabic, and contained three phonemes. Long words were either two or three syllables, and contained seven phonemes. Short and long words were matched for word frequency, imagery and concreteness.

The other three groups tested the generation effect under the same three conditions as word length, and a second word pool was constructed containing six phoneme words from the MRC database (Quinlan, 1992). The 240 words used were chosen such that if a single letter was replaced with a blank space, only one letter could be substituted to produce a legitimate English word. The 240 words and their corresponding fragments served as the word pool. Half the word pool items were randomly assigned to the generation condition for each participant, with the other half remaining as read control items.

For the serial recall component of the experiment, 90 words were randomly selected from each pool, to form 15 six-word lists of long words, and 15 six-word lists of short words in the groups involving the word length effect. For the generation effect groups, another 90 words were randomly selected from each pool to form equivalent generation or read lists. The 15 lists of each type were divided into 10 non-recall lists, and 5 serial recall lists for both long and short words, and generated and read words. The order of the 30 lists was then randomised. This procedure was carried out for each participant and ensured that a unique set of trials was created for each participant. The same 20 sets of lists were repeated for the three serial recall conditions – immediate, delayed and irrelevant speech.

The recognition component of the experiment was created using the 10 non-recall lists of each type from the initial serial order study trials, as well as distractors which comprised the unused 30 words of each type from the initial pools. A unique

list of 180 words resulted for each participant. These items were randomly ordered and arranged in six columns of 30 words on a single A4 sheet.

Procedure. All participants were tested individually, in sessions of approximately 30 minutes' duration. The study component of the experiment was administered on a computer with a monochrome screen. Words from each six-item trial were presented singly, in the centre of the screen, at the rate of one per second. Participants in the three groups involving word length were instructed to read each word aloud as it appeared on the screen. Participants in the generation groups were instructed to read the words aloud if they were undegraded, but to generate a word and say it aloud if the presented word was a fragment (had a letter missing). In all groups, however, the instructions stated that the six words on each trial were to be remembered in order.

Participants were instructed that on some trials they would be asked to recall the six items in order, but on other trials they would not be tested. They were told that they would not know in advance what type of trial it would be and that they would only find out at the end of each trial. The serial recall lists were signified by a row of question marks after the final word, at which point participants attempted to recall the six words aloud in order. The trials which were going to be later tested for recognition ended with a row of asterisks. The instructions stressed that on these trials recall was not required and participants were to simply wait for the next trial to begin. No rationale for this request was provided.

In the two groups involving immediate serial recall, participants were required to repeat the items aloud while the word lists were presented on the screen. On lists with the question mark cue for recall, they were directed to recall the list aloud in correct serial order, substituting the word "something" for forgotten words to preserve the serial order of remembered items.

In both delayed recall groups, following the last word in each list, a series of four, four-digit numbers appeared on the computer screen at a rate of one per second. As the digits appeared on the screen participants were required to read each four digit number aloud. After the last number appeared and was read, the recall cue appeared (whether it was question marks or asterisks) and participants responded appropriately. In both groups involving irrelevant speech, recordings of Russian news broadcasts were played during presentation of the six words only, and participants were instructed to do their best to ignore the irrelevant sounds. At the end of each list, they were instructed to respond to the recall cues in the same manner as participants in the other groups.

For the recognition component of the experiment, participants were given the recognition sheet after a three minute delay during which administrative details were completed. They were simply asked to circle any words they remembered as having been presented in the experiment they had just completed, and were given as much time to do so as they required.

All participants completed practice lists before commencing the experiment. The experimenter recorded their responses to the serial recall lists on a separate sheet as either correct in position, omitted, transposed or extralist intrusions. The primary recall measure was correct in position, and scores in the generation effect groups were conditionalised upon correct generation of target words at study. Total items correct in position for the serial recall lists were then converted to proportions (probability of correct recall) for analysis. For the recognition task, proportions were then computed, maintaining original serial position, for words correctly recognised from the experiment. Distractor words incorrectly recognised were recorded as false alarms.

Results

Means (and 95% confidence intervals) of the correct in position serial recall and item recognition data from the experiment are shown in Figure 1. In the generation groups, participants were able to generate the target item from the fragment on 89.3% of occasions. On the recognition tests, false alarms varied from 6% to 11% across groups, however the majority were found to be from a small proportion of participants.

Partly in the interests of space we have chosen to report the results relying upon 95% confidence intervals as suggested by Loftus and Masson (1994). For those who prefer traditional analysis of variance procedures, our verbal descriptions are supported by significant main effects and interactions from a 3 (experimental task) by 2 (effect) by 2 (recall condition) by 2 (item difficulty) by 6 (serial position) mixed design ANOVA.

With regards to the primary aims of the study, the results were clearly as expected for word length. Under serial recall conditions, there was a short word advantage, but this reversed under recognition. For the generation effect, the generate advantage was observed consistently in the recognition task, but the expected read advantage under serial recall did not emerge consistently.

Discussion

The first thing to note in the current data is that we replicate all the previously established effects. Word length effects were present in immediate serial recall, under delayed recall and under irrelevant speech. Moreover the generation effect was present under recognition procedures.

With regard to the central issues of the research, the suggested tradeoff between item and order information was clearly evident for word length. Short words enjoyed a typical advantage in serial recall, which reversed at recognition to become a marked advantage for long items. These results are very robust and provide compelling support for the item/order tradeoff explanation.

However, the picture was not as clear with the generation effect. There was a strong advantage for the generated items in the item test, however the expected advantage for the read items in the order test did not emerge. One explanation for the absence of the effect in serial recall is that our manipulation of generation lacked sufficient sensitivity to show the effect at the serial order stage of the experiment. Thus, with relatively long words to read, the time available to read the words may have put item processing under stress in the same way and for the same reasons that we argue in respect of the long words in the word length groups. In fact the read and long words only differed in length by one phoneme. Furthermore, omitting one letter from a multisyllabic word made many of the words extremely easy to generate (generation failures, where they occurred, were generally limited to the same few items from the word pool as they appeared in each group). Thus, there may not have been a large difference in the degree of item processing between the generated and read items. We will return to this issue after Experiment 2. This explanation receives further support when one considers the strength of the tradeoff effects. For word length, where there is a strong processing difference in serial recall there is a correspondingly strong reverse effect in recognition. With the generation groups, the weak processing difference in serial recall is matched by a correspondingly weak, though reliable, effect in the recognition component.

One potentially problematic finding for the item/order tradeoff approach deals with the fact that word length effects are attenuated or eliminated under articulatory

suppression (Baddeley et al., 1975; Tehan, Hendry, & Kocinski, 2001). This is readily accounted for in TDR models by assuming that suppression prevents rehearsal. From the item/order perspective it is possible that item processing becomes more difficult under suppression conditions, and that the easy items are more prone to be affected by the changes. That is, the difficult items already require substantial item processing, so suppression can have little additional effect with these items. In contrast, the short words now require extra item processing which decreases the amount of order processing conducted. Consequently, the differential effects of item and order processing are reduced under suppression and the consequential order and items effects are likewise reduced. The key unknown here is what will happen to item recognition. The results of Experiment 1 indicate that it is possible for item effects to still be present when order effects are eliminated. If generation and word length are based upon the same processes then the expectation would be that the word length effect would attenuated on the order task, but that the reverse effect would be present on the item task.

Experiment 2

Method

Participants. 20 introductory psychology students from the University of Southern Queensland volunteered to participate in the experiment, in return for which they were given course credit, or a ticket in a raffle for cash prizes ranging from A$20 to A$200.

Materials. The same materials from Experiment 1 (WLE) were used.

Procedure. The experiment was conducted in exactly the same manner as the immediate serial recall word length components of Experiment 1, with the only exception being the use of articulatory suppression. In addition to the instructions provided to participants in the previous experiment, they were also required to articulate “the the the” aloud as rapidly as possible during presentation of the lists. Serial recall, recognition and scoring procedures were identical to those used in Experiment 1.

Results

Means and 95% confidence intervals from the experimental conditions are shown at Figure 2. False alarms from the recognition component of the experiment were 10.0%, and again came mostly from a small minority of the 20 participants. Under articulatory suppression, there was no difference between recall of long and short words in the immediate serial recall task, however a reliable advantage for long words was evident at recognition.

Discussion

As predicted, the reduction in sensitivity of the word length effect by the use of articulatory suppression has produced an attenuated word length effect under suppression. Reverse word length effects were present in item recognition. These results are similar to those found with our manipulation of the generation effect in serial recall. In fact, across the board, the suppression results are very similar to the generation effects in Experiment 1. In both instances the absence of an effect on the order task does not preclude an item advantage under recognition. Thus, the current results suggest that there is no fundamental difference between word length and generation effects under these conditions.

The current results also provide an explanation for articulatory suppression within an item-order tradeoff perspective. It is clear that compared with conditions where participants are not required to suppress articulation, it is the short words which

are affected more by suppression than the long words. This pattern is compatible with the notion that item processing becomes more difficult under suppression conditions, and that the easy items are more prone to be affected by the changes.

General Discussion

In the current experiments, we tested an item/order tradeoff explanation of the word length effect. While this account makes the same prediction as most other models, that short words will be better recalled than long words in serial order, it makes the unique prediction that reverse word length effects would emerge on a test of item information. We provided a further test of the explanation by arguing that one of the markers of the item/order tradeoff perspective, the generation effect, should show an equivalent pattern of results to that found with word length.

The word length effects emerged exactly as expected. Strong word length effects in serial recall were matched by strong reverse effects in item recognition. Moreover, when serial recall differences for short and long words were reduced in Experiment 2, a similar reduction in the strength of the reverse word length effects was found at recognition. Given that item/order dissociations have been observed in many other areas, parsimony would suggest that the item/order account is a viable alternative explanation for the word length effect.

The generation effects, while consistent, were not as robust as expected. The typical generation advantage was observed on the recognition component but the reverse effect was not present on serial recall. We have suggested that the absence of this effect is due solely to a sensitivity issue and does not represent a fundamental difference between word length and generation. Clearly, a stronger and more sensitive manipulation of the generation process is required to confirm this suggestion.

While the results appear to support an item/order tradeoff perspective, we have provided no evidence for our primary assumption that long words are more difficult to process at the item level than short words. In the visual word identification literature there is clear evidence that short words are identified more quickly than long words in lexical decision, word naming and perceptual identification tasks (Balota & Chumbley, 1985; Forster & Chambers, 1973; Samuels, LaBerge & Bremer, 1978). Furthermore, it is clear that item identification and lexical decision processes play an important part in immediate serial recall. The relationship between item identification time and span has been demonstrated on a number of occasions with both children (Case, Kurland & Goldberg, 1982; Hitch, Halliday & Littler, 1989) and adults (Tehan & Lalor, 2000) as participants. For example, Tehan and Lalor (2000) demonstrated that performance on lexical decision, word naming and other item identification tasks predicted individual differences in serial recall. In short, there is evidence in the literature supporting the notion that item processing speed is important in immediate serial recall and that short words are processed faster than long words.

The current results have a correlational flavour to them. Because word length produces similar patterns to the generation effect (and presumably other item/order effects as well) we are assuming that similar underlying processes are involved. Resolving this issue will necessitate further research. It is not at all certain whether the item/order tradeoff perspective will be able to explain all word length effects, and we suspect that this is unlikely. However, the item/order approach does predict that the standard word length effect should be apparent in any task that requires the use of serial order information. It thus readily predicts that word length effects should be found in complex span tasks (Tehan et al., 2001), backward recall (Cowan, Wood & Borne, 1994), serial recognition (Baddeley, Chincotta, Stafford & Turk, 2002) and

even in probed recall where it is reasonable to assume that participants are using serial rehearsal (Avons, Wright & Pammer, 1994). Even where word length effects have not been found in an order task (Cowan et al., 1994, backward recall in a continuous distractor task) the approach might still be found on an item recognition test. For present purposes, we believe that the data suggest that the item/order account of the word length effect is certainly a viable option to consider.

References

Avons, S. E., Wright, K. L. & Pammer, K. (1994). The word-length effect in probed and serial recall. *Quarterly Journal of Experimental Psychology, 47A,* 207-231.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behaviour, 14,* 575-589.

Baddeley, A. D., Chincotta, D., Stafford, L., & Turk, D. (2002). Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *Quarterly Journal of Experimental Psychology, 55A,* 353-369.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language, 24,* 89-106.

Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory & Language, 34,* 594-621.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33,* 386-404.

Cowan, N., Wood, N., & Borne, D. (1994). Reconfirmation of the short-term storage concept. *Psychological Sciences, 5,* 103-107.

DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22,* 1136-1146.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behaviour, 12,* 627-635.

Hitch, G. J., Halliday, M. S., & Littler, J. E. (1989). Item identification time and rehearsal rate as predictors of memory span in children. *Quarterly Journal of Experimental Psychology, 41A,* 321-337.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1 (4),* 476-490.

Mulligan, N. W. (1999). The effects of perceptual interference at encoding on organization and order: Investigating the roles of item-specific and relational information. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25,* 54-69.

Nairne, J. S. (2002). Remembering over the short term: The case against the standard model. *Annual Review of Psychology, 53,* 53-81.

Nairne, J. S., Riegler, G. L., & Serra, M. (1991). Dissociative effects of generation on item and order information. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17,* 702-709.

Quinlan, P. T. (1992). *The Oxford psycholinguistic database.* Oxford: Oxford University Press.

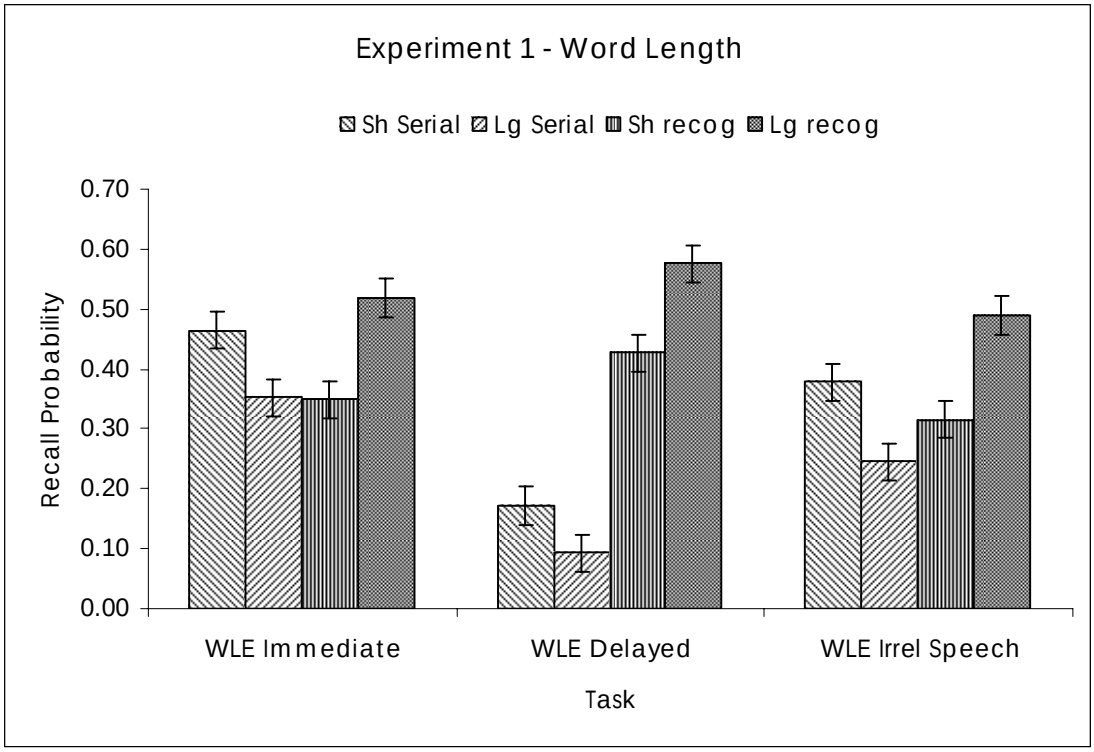
Samuels, S. J., LaBerge, D., & Bremer, C.D. (1978). Units of word recognition: Evidence for developmental changes. *Journal of Verbal Learning and Verbal Behaviour, 17,* 715-720.

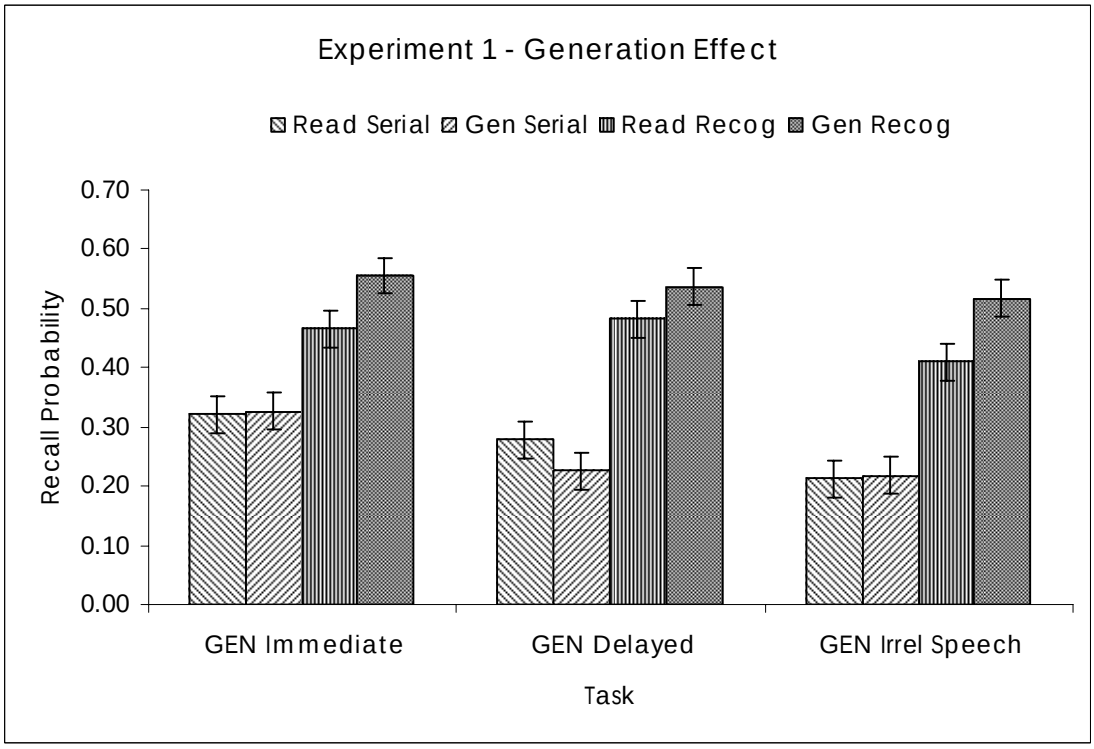
Tehan, G., Hendry, L., & Kocinski, D.G. (2001). Word length and phonological similarity effects in simple, complex, and delayed serial recall tasks: Implications for working memory. *Memory, 9,* 333-348.

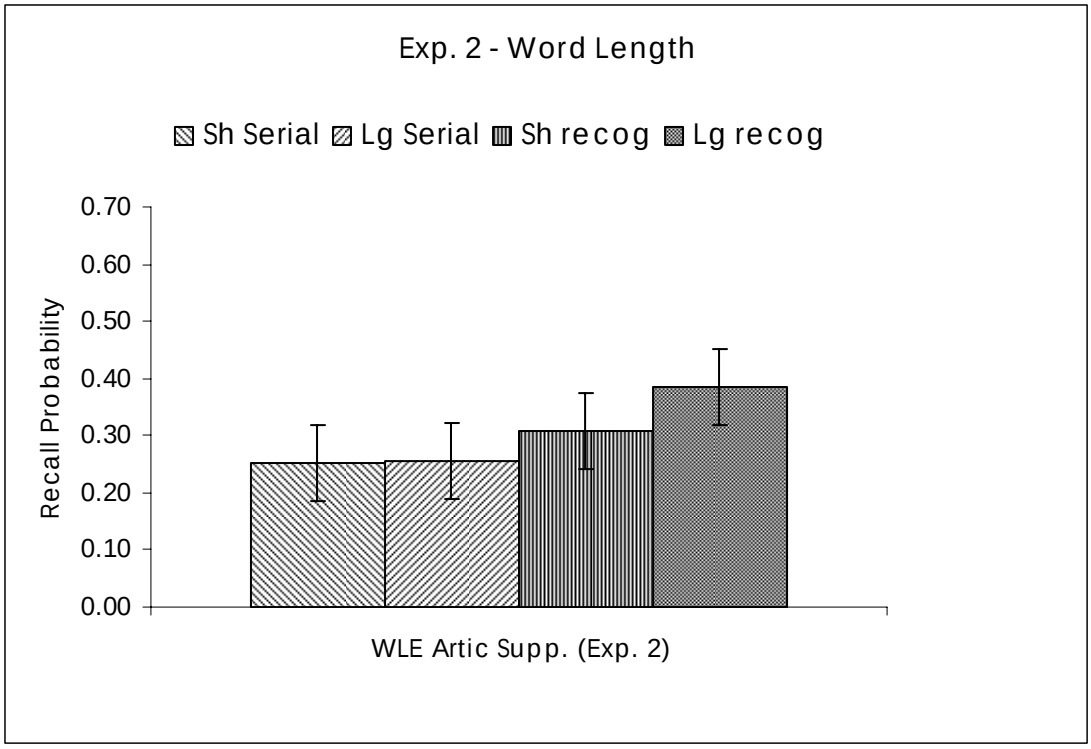
Tehan, G., & Lalor, D. M. (2000). Individual differences in memory span: The contribution of rehearsal, access to lexical memory and output speed. *Quarterly Journal of Experimental Psychology, 53A,* 1012-1038.

Figure Captions

Figure 1.
Word length and generation effects as a function of item and order tests.
Figure 2.
Word length effects for item and order tests under articulatory suppression conditions.


Experiment 1 - Word Length
Sh Serial
Lg Serial
Sh recog
Lg recog
Recall Probability
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
WLE Immediate
WLE Delayed
WLE Irrel Speech
Task


Experiment 1 - Generation Effect
Read Serial
Gen Serial
Read Recog
Gen Recog
Recall Probability
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
GEN Immediate
GEN Delayed
GEN Irrel Speech
Task


Exp. 2 - Word Length
Sh Serial
Lg Serial
Sh recog
Lg recog
Recall Probability
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
WLE Artic Supp. (Exp. 2)