AN ITEM AND ORDER PROCESSING ANALYSIS OF WORD LENGTH, GENERATION AND

PERCEPTUAL INTERFERENCE EFFECTS IN HUMAN MEMORY

Liam Hendry, B. Sc. (Psychology) (Hons.)

A Thesis Submitted in Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy at the

University of Southern Queensland

June, 2004.

STATEMENT OF ORIGINALITY

This thesis describes original research conducted by the author. I certify that the ideas, experimental work, results and analyses reported in this thesis are entirely my own effort, except where otherwise acknowledged. Data from Experiments 1 to 6 and Experiment 10 have been presented in poster form at the Quebec Short-term and working Memory Conference in July 2002, and published online in an article in the scientific journal Memory, January 2004 (Hendry & Tehan). No part of this dissertation has been previously submitted for an academic award at this or any other university.

Candidate: Liam Hendry

Principal Supervisor   Gerry Tehan

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Appendix 1   Words used in Experiments involving Word Length

Appendix 2   Words used in Experiments involving Generation

Appendix 3   Words used in Experiments involving Perceptual Interference

Appendix 4   Words used in Experiment 11 involving Generation revision.

Acknowledgements

This dissertation represents the end point of an 8-year odyssey through the levels of degrees of Psychology at the University of Southern Queensland. In that time, I consider I have been most fortunate to have had around me, right from the very first day, friends, colleagues and teachers of the highest calibre. Like an actor receiving an Oscar, I have a long list of people to whom I wish to express my deepest gratitude, for without their help and support I would not be where I am now.

On that very first day of undergraduate Psychology, I met the group of people who would remain great friends and colleagues throughout the programme to this very day, and although we have all followed different paths through the various levels of academic endeavour and are now scattered across the globe, we remain linked by our common experiences at USQ. For their special friendship and support, my thanks go out to David Kocinski, Craig Smith, and Nancey Hoare, who have always been there for me through the ups and downs of all those years of study.

Much can happen in 8 years, David is now a psychologist at the Gold Coast, Craig is working in San Francisco, and Nancey is still at USQ, also in the Ph.D. programme. We still have a terrific study group, although the faces have changed, and I feel privileged to have been working alongside my new colleagues in the Ph.D. room - Sharron, Geoff, Kerry, Sandra, Maxine, Coralie & Don to name but a few. The friendship and support of these wonderful people over the last few years has made a difficult task seem easy and fun. The social aspects of our time together have provided excellent relief from the rigours of research (special thanks to Maxine- the hostess with the mostest), and the intellectual benefit of having such a collection of keen minds close by has provided much discussion and food for thought over the years.

The University itself, and the Faculty of Sciences in particular, would have to be one of the best learning environments available in Australia. Here I have had the

xi

Abstract

When participants are presented with lists of items for immediate serial recall, tradition would suggest that a race begins – between the need to constantly refresh or recycle the memory trace of that list, and a tendency for the memory trace to decay. Standard models in the literature assumed a complex interaction of mental subsystems whereby a controlling attentional process strove to keep the memory of such a list alive for a sufficiently long period of time so it could be remembered and output in order, using a type of recirculating loop rehearsal and storage mechanism to offset the decay process. Evidence supporting such models stemmed from the observation that more short words could be remembered in order than long words (Baddeley, Thomson, & Buchanan, 1975). This word length effect, described in the second chapter, was a crucial piece of evidence for rehearsal and decay models, in the example given, the recirculating loop was seen as being time-based and extremely limited in capacity, such that memory was deemed equivalent to the amount of information which could be cycled through the rehearsal loop in about two seconds. A number of recent challenges to this model of remembering have cast doubt on the nature of the process as described in such models as that of Baddeley (1990; 1996).

Chapter 1 began by providing an overview of the development of such models from their earliest form, and also introduced some alternative ideas about the structure and function of human memory. A processing view was described, in which the probability of recalling a list of items depended not upon a race between decay and rehearsal, but on differential processing of items based on their nature. As remembering a list in its original order involves not only remembering the items themselves, but also information about how they relate together in the list, an alternate theory was advanced that in some cases the processing of information about the items, and information about their serial order could dissociate, producing a processing tradeoff. As individual items

were better remembered, information about their presentation order diminished. This observation (Nairne, Riegler, & Serra, 1991) was introduced as the item-order hypothesis.

The item-order hypothesis suggested that under certain conditions increased item processing could lead to deficits in order processing, and that this produced a dissociation in performance between item and order memory tasks. The generation effect (Slamecka & Graf, 1978) was one such example, as was the perceptual interference effect (Mulligan, 2000), and these were discussed in Chapter 3. The word length effect was seen as another instance where this tradeoff might be observed. A design incorporating elements of item and order tasks based upon Nairne et al. (1991) was detailed in the fourth chapter, leading on to empirical testing of the word length effect (Chapter 5), the generation effect (Chapter 6) and the perceptual interference effect (Chapter 7). This series of experiments compared word length and generation effects under serial recall and single item recognition tasks, using a range of test conditions designed to allow replication and extension of existing data from these separate streams of research.

Results did not appear as predicted for some aspects of generation and all aspects of perceptual interference, and further experiments in Chapter 8 attempted to address the current findings. For the experiments involving word length, short words were better recalled than long words on the serial recall task, but long words were better recognised in the recognition task. Following additional manipulations in Chapter 8, the generation effect began to produce a similar pattern, but the results for perceptual interference were inconclusive. Word length data, however, were consistent with the item-order approach and supported a novel explanation for the word length effect. Implications and conclusions were discussed in Chapter 9.

Chapter 1 Contents

CHAPTER 1

Introduction

1.1 Background and History

George Miller (1956) was the first of the modern researchers to estimate the capacity of human short-term memory (STM). He suggested the capacity was 'seven plus or minus two' chunks of information. Whether a 'chunk' of information was defined in terms of digits, letters or words, Miller's major finding has since been largely interpreted as the ability to retain about seven items in immediate memory, where items have been typically defined by experimenters as words (Baddeley, Thomson, & Buchanan, 1975).

Initial attempts to measure the characteristics of forgetting from STM by John Brown (1958) and Lloyd and Margaret Peterson (1959) spearheaded a wealth of research into the structure and function of STM. Rapid forgetting from STM was seen to occur when participants were distracted, for example by having to count backwards in threes while attempting to retain information. It was felt that Brown and the Petersons provided strong evidence for at least two different memory systems - in that short-term forgetting displayed different characteristics from forgetting in long-term memory (LTM). Forgetting over the short-term was seen as a function of time-based decay, whereas interference was the prime determinant of forgetting in LTM.

Waugh and Norman (1965) developed a model of human memory partitioned into primary memory and secondary memory, where primary memory referred to the theoretical system they assumed to be responsible for short-term storage. In addition, they used the descriptor "short term memory" to refer to an experimental situation in which typically a small amount of information is retained over a short period of time (Baddeley, 1990). Secondary memory was assumed to reflect a long-term store. Their major contribution to the field was the thought that both primary and secondary memory systems could be involved in STM, as well as LTM tasks.

At much the same time, Atkinson and Shiffrin (1968) introduced a multicomponent model, expanding on an earlier idea by Broadbent (1958) where information enters a sensory register, then

a short term store, then a long-term store. The short-term store was seen to have a very limited capacity. It was also described as a 'workspace' where information could be rehearsed and or processed. It could retain information for up to 30 seconds, this information being stored in a serial order manner. Information from the long-term store could be retrieved in parallel fashion i.e. many items at once. The short-term store was also believed to operate on acoustic information, whereas the long-term store operated primarily on semantic information. The main attraction of the Atkinson and Shiffrin (1968) model was that they specified a number of distinct stores - but more importantly, they described how information passed from store to store.

1.2 The Working Memory model

By the middle of the 1970's, most of the assumptions which underpinned the Atkinson and Shiffrin model had been disputed, and at the same time a new model of immediate memory was emerging. Existing ideas of multiple stores were expanded by Alan Baddeley in 1966, in an ongoing process culminating in the in the working memory model (Baddeley & Hitch, 1974; Baddeley et al., 1975). The working memory model was initially tripartite, comprising a Central Executive, which controlled and allocated resources, a Visuo-Spatial Scratchpad for the storage and manipulation of visual imagery, and a Phonological Loop for the recycling of verbal (acoustic or auditory) information.

The Phonological Loop component of the Baddeley model evolved from the previous idea of a short-term store. Based on performance on the immediate serial recall task, the Phonological Loop was assumed to hold a limited amount of verbal information in serial recall for a limited amount of time. As such, the measure of the Loop's capacity changed from being a (relatively fixed) number of items to a relatively fixed amount of time - specifically limiting short-term storage to the amount of information a person could rehearse in no more than two seconds (Baddeley et al., 1975; Schweickert & Boruff, 1986). The role of rehearsal became a dominant theme in the functioning of the Phonological Loop because forgetting was seen as being simply due to a rapid decay of memory traces. For traces to be maintained in a useful state, rehearsal was required.

That is, immediate recall involved counteracting the decay process by continuous rehearsal of the to-be-remembered items (Baddeley, 1996). Viewing the process of short-term memory performance in terms of a race between decay and rehearsal has become widespread in many current formal models of immediate memory (Nairne, 2002).

1.3 The Word Length Effect

One of the cornerstones of the trace decay plus rehearsal (TDR) models is the word length effect (WLE), the finding that immediate serial recall for short words is better than for long words. The basic assumption here is that rehearsal occurs in real time and that in any given period of time, more short words can be rehearsed than long words. Further support for this notion was that the word length effect was eliminated when rehearsal was prevented by the use of articulatory suppression (Baddeley et al., 1975). The word length effect is a robust phenomenon when length is measured in terms of the number of syllables and much is known of its boundary conditions.

The locus of the word length effect was initially identified as the spoken duration of words, however recent research has questioned this explanation, as well as the more general proposition that immediate serial recall can be sufficiently explained in terms of the decay-rehearsal tradeoff. In Chapter 2, the word length effect will be examined in detail, and for present purposes it is sufficient to note that it has no universally accepted explanation. The current thesis tests a new explanation of the word length effect.

1.4 Processing Accounts of Memory

All the above models are primarily storage models, involving multiple stores, each with its own unique characteristics. It is possible, however, that a single (unitary) system supports retention for both long-term and short-term tasks (Melton, 1963). At the same time as the working memory model was being developed, an alternative view of memory has also been proposed: one that concentrates on processing rather than storage (Searleman & Herrmann, 1994). The levels of processing model of Craik and Lockhart (1972) argued for less complicated models of memory, and instead of separating memory into different structural components such as primary and

secondary, or long-term and short-term, their model concentrated on how information was processed. The durability of memory was seen more as a function of the amount or quality of processing given to information. The deeper an item was processed, the easier it would be to remember.

According to this approach, the strength of a memory is related to the depth of the encoding process, which is of itself a continuum. An example of depth of processing would be where participants are shown a list of words, and asked to make different types of judgements about each word, relating to processing levels. A shallow processing approach would deal with the appearance of the word (for example, 'was the word in uppercase or lowercase?') An intermediate processing level question might ask 'does the word rhyme with fish?'; and a deep processing question might require the participant to see whether the word fits into a blank space to make a meaningful sentence, or perhaps to rate items on an emotional dimension such as pleasantness. Results typically were interpreted to mean that deeper processing produced better recall. As there was no direct measure of processing depth, it was somewhat roundly assumed that the better-recalled words were those which had received deeper processing. While the circularity of the assumption is not fatally problematic for levels of processing models (Searleman & Herrmann, 1994), it has led to wide criticism of this approach in the literature (e.g. Baddeley, 1978).

1.5 Item and Order Information

Craik and Lockhart's levels of processing approach to memory operates using an item by item approach, but says little about serial order. Order information, or the component(s) of a memory trace which allow the participant to relate items on a list to each other in memory tasks involving retention of serial order, featured in the approach used by Einstein and Hunt (1980) who introduced an hypothesis about unitary memory based on item-specific and relational processing. Item-specific processing related to attending to the features of the item itself, whereas relational processing involved attending to ways in which items in a list could be grouped together, and ways in which items related to the other items on a list. Individual-item processing led to better

overall performance on a recognition test, and relational processing had a beneficial effect in free recall in terms of the clustering of related items (Einstein & Hunt, 1980). However, recall was maximised when both item and relational information were successfully used in combination.

The distinction between item and relational processing has subsequently been applied to numerous effects in the study of human memory. There are two important extensions of this work that form the basis for the current series of experiments. The first is that other researchers have taken serial order information to be a form of relational processing. Secondly, with item and order processing as a research focus, more recent memory studies have provided examples of situations where they are seen to dissociate, depending upon the type of memory task employed.

1.6 Item and Order Tradeoffs: Generation and Perceptual Interference

One crucial development of the item/order literature has been the assumption that under some circumstances item processing and order processing can trade-off and this produces dissociations in memory performance on tasks assumed to be differentially sensitive to item or order processing. The generation effect and the perceptual interference effect, for example, have recently been explained in such terms (DeLosh & McDaniel, 1996; Engelkamp & Dehn, 2000; Mulligan, 1999, 2000; Nairne, Reigler & Serra, 1991).

The Nairne et al. (1991) study is particularly relevant to the hypotheses of this paper. In three experiments, they investigated the effects of generation on item and order retention, using different memory tasks within the same experiments. In their procedures participants either read items as they appeared on a screen or generated items from a word fragment, e.g. umbr_lla. Each list contained a small number of items that were processed in either of the above ways. At the end of each list participants attempted either serial recall after a brief retention interval or they were told to not respond and simply wait until the next trial began. After the lists had been presented in this way a surprise item recognition test was administered to test for memory of the items in the initial trials that had not been responded to. On the serial recall trials the read items were better remembered in order than the generated items. However on the recognition test the

items that had been generated were better recognised than the read items. In short, on a test of order there was a read advantage but on a test of item information there was a generate advantage: The different types of processing dissociated on the two types of test. The Nairne et al. (1991) explanation for the above results is that generation requires more item processing than does simply reading an item and that the facilitative effects of extra item processing will be evident on a test of item memory, such as recognition. However, if the task involves order processing, extra resources being devoted to item processing will necessarily lead to decreased time or resources being devoted to order processing, given that both item and order processing must be done in the same limited time frame. Thus, generated items will not receive much order processing because more time/resources are being spent on item processing and this deficit will become manifest in an order memory task, such as immediate serial recall. In short, the dissociation across memory tasks stems from a differential tradeoff in item and order processing.

As mentioned earlier this form of reasoning has been extended beyond the generation effect. For example, Mulligan (2000) found an equivalent pattern of results with a perceptual interference task. Again, items were either read or each item was presented briefly followed by a visual mask. Lists were tested for either serial recall or a later item recognition test. The read items were better recalled than the masked items on serial recall, but the masked items were better recognised than the read items. Mulligan argued that masking an item required additional processing resources to be expended at the item level, reducing the degree of order processing performed.

1.7 Chapter Summary

The current thesis explores the possibility that the word length effect could be another instance of this item-order tradeoff. The essential argument to be made is that long words are harder to process at the item level than short words. This means that the short words would receive better order processing than long words. Thus, short words would be better recalled than long

words in immediate serial recall, because they have received more order processing. The key prediction is that long words will be remembered more often in an item recognition test.

This thesis is organised around a series of experiments which explore word length, generation and perceptual interference effects in such a manner as to make the patterns of results directly comparable. The experimental paradigm used is very similar to that used by Nairne et al. (1991. In all cases order information is tested using immediate serial recall and item information is tested via a final surprise item recognition test. As the same basic method is used for all the experiments, a consistent pattern of results across the three effects would provide evidence of a generalised processing similarity. If this pattern emerges, it would no longer be necessary to consider the word length effect as a purely short-term phenomenon, and by extension it may no longer be necessary to view human memory in terms of a multicomponent model, if a simpler unitary system based on processing differences would suffice.

The item/order processing tradeoff approach potentially provides a very different account of standard serial recall effects. From this perspective, no special stores are required, and forgetting is not necessarily time-based, although the presentation rate of items is still important as it provides a limited opportunity for encoding/processing to occur. This thesis takes probably the most crucial effect supporting the decay and rehearsal approach – the word length effect – and subjects it to investigation from an item/order processing perspective.

Chapter 2 Contents

CHAPTER 2.

The Standard Model and the Word Length Effect

From the introductory notes in Chapter 1, the word length effect has been represented as a robust and stable short-term memory effect, about which much is known in the literature, yet it is now proposed to examine it from such a novel perspective as item and order processing tradeoff effects. If an existing theoretical framework was already providing an adequate explanation of the word length effect, it would seem unnecessary to advance an alternative theory. A review of the word length literature, however, reveals there are already several competing theories, with many studies reporting apparently contradictory data.

2.1 Decay Vs Rehearsal in the Original Account

The original account of the word length effect was provided by Baddeley, Thompson and Buchanan (1975). By observing reliable span differences between lists of monosyllabic and polysyllabic words, they were the first to empirically show that short-term memory capacity was not a function of a fixed number of items, rather a function of the spoken duration of those items. A model of working memory eventuated from these and additional data (Baddeley & Hitch, 1974), comprising a Central Executive to oversee the processing and storage of information, a Visuo-Spatial Scratchpad for the manipulation and storage of visual information, and an Articulatory Loop for both processing of auditory inputs, and the encoding from visual into auditory codes of visually presented verbal information. This model has evolved (Baddeley, 1996; Cowan, Baddeley, Elliott, & Norris, 2003; Gathercole, 1997)and withstood three decades of research, often with contradictory findings, to become established in textbooks and scientific journals. Memory models such as that of Baddeley, which are based on assumptions of trace decay and rehearsal, have been described under the general title of the Standard Model (Nairne, 2002).

The model assumes the above components, in various forms and combinations, are utilised for the storage and processing of activated memory traces, which are subject to rapid decay. In

the components reserved for storage, particularly that of phonological information, memory traces for items such as words were thought to rapidly decay unless they were actively and continuously rehearsed. The Phonological Loop is of most interest to short-term memory researchers, as it is seen as the process by which verbal material is rehearsed and recycled in order to be remembered. Since the Baddeley et al. (1975) article was published, the Phonological Loop in his model has been further divided into two separate components: the Phonological Store in which items are placed, where they decay rapidly unless refreshed by rehearsal, and the rehearsal process, responsible for simple rehearsal as well as the translation of visually presented material into an acoustic code (Gathercole, 1997).

2.2 The Baddeley, Thompson and Buchanan study

The word length effect is intimately related to this conception of memory. From the results of eight experiments, Baddeley et al. (1975) concluded the following:

(1) *That memory span was sensitive to word length across a range of verbal material.* In the first experiment, using small closed word pools, a recall advantage was found for short (monosyllabic) words across list lengths ranging from four to eight items. To control for linguistic differences between short (mainly Anglo-Saxon) and long (mainly Latin) words, the second experiment compared short and long country names, again using closed small word pools, and again a clear advantage was found for the short (monosyllabic and disyllabic) items when tested in lists of five items.

(2) *That the word length effect remained even when the number of syllables and the number of phonemes were held constant, and all that varied was the spoken duration of the words.* Their third experiment reduced the distinction between long and short items from an obvious difference in the number of syllables to the spoken duration of those items when syllabic length was the same. All items were disyllabic, the short word lists comprised words of a shorter spoken duration as measured on an oscillograph (e.g. *wicket, bishop, phallic*), while the long words had a longer measured spoken duration (e.g. *morphine, voodoo, Friday*). Word pools were again small (10

items) and closed. Five-item lists were tested at a 2 second presentation rate, with recall paced by use of a metronome. Again, shorter items were recalled more often, but only in the first three list positions. Baddeley et al.'s fourth experiment further refined the difference between long and short items, with the number of phonemes also held constant. This reduced the closed word pools to only five items, which were presented at the rate of one per second. A recall advantage for the shorter items was again evident, across all serial positions. This and the preceding experiments had utilised auditory presentation, however when the experiment was repeated using visual presentation at a 2 second rate, no effect of word length was found. The absence of an effect was explained in terms of participant strategies, with the 2 second per item presentation rate allowing some participants to use imagery rather than rehearsal. Their fifth experiment, therefore, replicated the preceding one with visual presentation and instructions to participants to rehearse. A recall advantage was again found for the shorter items, and additional measures were taken to assess reading rate and articulation rate.

(3) *There was a relationship between articulation rate and span, such that span was equivalent to the number of items which could be rehearsed in no more than 2 seconds.* Additional results from Experiment V (Baddeley et al., 1975) provided estimates suggesting a person's span for words was equivalent to what could be read in 1.6 seconds, or articulated in 1.3 seconds. Reading rate was assessed by requiring participants to read the entire set of lists four times as quickly as possible, with a reading rate calculated in words per second. Articulation rate was measured by requiring participants to read a set of three words as rapidly as possible, for ten repetitions, over four tests. Experiment VI tested reading rate and word length across five different lengths (one to five syllable words) and found a strong linear relationship between reading rate and memory span for the words, with a steady decline in the number of words correctly recalled in order (as well as reading rate) as the number of syllables per word increased. With a different pool of participants, the estimate of memory span increased to the number of items which could be read aloud in 1.8

seconds, suggesting that the capacity of short-term memory was constant when measured in units of time, rather than units of structure (Baddeley et al., 1975).

4) *That concurrent articulation (which prevented participants from rehearsing) removed the word length effect under visual presentation conditions.* Baddeley and his colleagues suggested that if memory traces were subject to decay, and that decay was offset by refreshing the memory traces through the process of rehearsal, then preventing rehearsal should remove the effect of word length, because the rehearsal process was seen as the locus of the effect. If short term memory depended upon a time-based rehearsal process, it followed that more short words could be rehearsed than long words in the estimated 1.8 seconds. Recall for one syllable and five syllable words was tested in the seventh experiment with rehearsal prevented for one group by the use of articulatory suppression. The group using suppression failed to show any effect of word length, while the control group produced a clear advantage in recall for the short words. This result, however, was premised on the assumption that articulatory suppression (requiring participants to repeatedly articulate an irrelevant or neutral item during presentation of lists of words) prevented rehearsal. The eighth experiment provided data suggesting that the effect of suppression depended upon list presentation modality, and could be observed with visual presentation but not with auditory presentation. This result in particular suggested a multicomponent model, with different systems involved in processing visual and verbal inputs. The articulatory process, vulnerable to suppression, was seen as the means of converting visual stimuli into acoustic or phonemic information, under the assumption that short-term memory traces are phonemic in nature. Auditory stimuli, however, required no conversion and directly accessed the store, and were thus relatively unaffected by suppression.

To summarise, the word length effect manifested as "a clear tendency for performance to decline as word length increases" (Baddeley & Hitch, 1974, p. 78). The correlation between span and rehearsal rate (memory span and reading rate, r = .685; Baddeley et al., 1975) became somewhat predictive, with a person's span being equivalent to the amount of information which

could be read aloud in approximately two seconds. The strong linear association between reading rate and span was described as "remarkably straightforward" (Baddeley et al., 1975, p.583.), but it is possible that the fast readers were good memorisers because of the influence of another unexplored variable.

The reading rate/span relationship has been used to explain materials (Schweickert & Boruff, 1986), developmental (Hulme, Thomson, Muir & Lawrence, 1984) and cross-linguistic (Ellis & Hennelly, 1980) differences in memory span. Because time, and not a fixed amount, was now seen as the underlying measure of immediate memory, it followed that more shorter items would be rehearsed in the same amount of time than longer items, as reflected in the word length effect.

To summarise, models of memory such as that of Baddeley (1996) view the word length effect as being indicative of a tradeoff between decay and rehearsal (Neath & Nairne, 1995) in an immediate memory system, such that if rehearsal does not occur quickly enough, the memory trace of an item 'will have decayed too far to be usable' (Neath & Nairne, 1995, p.429.).

2.3 Problems with Decay and Rehearsal

Challenges to rehearsal plus decay assumptions have recently been advanced, and data published which question the assumption that the spoken duration of words is responsible for the word length effect, made in conclusions 1 and 2 of Baddeley et al. (1975). Lovatt, Avons, and Masterson (2000; 2002) failed to replicate the finding that lists of short-duration words are better recalled than lists with longer spoken duration words, except in the case where the original Baddeley et al. (1975) materials were used. Other sets of long and short duration items, when tested, did not reliably show the effect found by Baddeley et al. (1975) and replicated by Cowan et al. (1992). Although Cowan and his colleagues found a word length effect for disyllabic items matched for all but spoken duration, the items they used were those of Baddeley et al. (1975). Alternate and similarly-constructed word pools created by Caplan, Rochon, and Waters (1992), as well as Lovatt et al. (2000) and a matched three-syllable word set (Lovatt et al., 2002) did not produce recall differences to mirror differences in their spoken duration, nor did a study of Finnish

words and nonwords (Service, 1998). These data seem to limit the generality of spoken duration as an adequate explanation of the word length effect, and thus cast doubt on an explanation of immediate memory based on a race between decay and rehearsal.

Other challenges to decay-rehearsal models include instances where rehearsal rate and span dissociate. Thus, it is possible to match items for spoken duration and still see differences in span (Hulme, Maughan & Brown, 1991; Schweickert, Guentert & Hersberger, 1990). Likewise it is possible to see differences in spoken duration without accompanying changes in span (Caplan et al., 1992; Service, 1998). It should be noted, however, that while spoken duration manipulations may or may not precipitate a reliable word length effect, the manipulation of number of syllables/phonemes remains a robust method of demonstrating it.

If the word length effect depended upon a rehearsal mechanism, it should be expected that the effect would disappear when rehearsal is prevented, as noted in the fourth conclusion of Baddeley et al. (1975), and supported by the results of their seventh and eighth experiments. The use of articulatory suppression as a device to eliminate rehearsal removed the effect of word length under visual presentation conditions. It is possible however, under some circumstances, to still observe the word length effect when rehearsal is prevented via articulatory suppression (LaPointe & Engle, 1990). When word pools were open, instead of being limited to only a handful of items, they found the word length effect remained under suppression, using visual presentation. Word length effects also remain present when other methods of preventing rehearsal are used. For example, Coltheart and Langdon (1998) presented items at an extremely rapid rate of 8-10 items per second. At these rates rehearsal was not possible, but the word length effect remained. In sum, there are sufficient instances in the literature to suggest rehearsal may not be a prime determinant of span (Brown & Hulme, 1995), and that the rehearsal/decay explanation of the word length effect is at best insufficient.

2.4 Alternative Explanations of the WLE

2.4.1. *Decay at output.*

An alternative explanation for the word length effect (the output time / decay model of Cowan and his colleagues) also involves decay, but it occurs during the recall (output) process rather than during the study of lists. Word length effects were examined in lists containing a mixture of short and long spoken duration words (Cowan et al., 1992). Spoken duration effects seemed to be stronger in the later serial positions than in the earlier serial positions. In lists with longer duration items in the earlier serial positions, recall was worse for items that were recalled towards the end of the list. The interpretation was that the act of recalling the early items affected retention of the later items. As items of longer duration were being recalled, the representations of the later items were undergoing more decay in active memory than when items of shorter duration were being recalled first. The longer it took to recall the initial items the harder it became to retrieve the later items. According to this account, the word length effect was occurring because of differential amounts of decay during the recall process itself resulting from the duration of the early items, rather than during rehearsal prior to recall. The locus of the effect was placed firmly at output, and if so, then it is difficult to assert that rehearsal and output could be occurring simultaneously (Nairne, 2002).

If decay during output is the causal variable underlying the word length, then the process of recall is affecting remaining memory traces in the manner of the distractor activity employed in a typical Brown-Peterson task. If so, the word length effect should become attenuated and possibly removed if recall is tested after a filled retention interval that is sufficiently long for both short and long items to have decayed. Tehan, Hendry and Kocinski (2001) presented participants with lists of four words that were tested immediately or after a filled retention interval involving 12 seconds of solving maths problems as the distractor activity. Recall probability was compared for three-phoneme words and seven-phoneme words and standard word length effects emerged on an immediate test. However, the effects were still present and were still strong on the delayed test.

Twelve seconds of doing mathematics problems should have provided sufficient time for trace decay to occur in the absence of rehearsal. This finding has recently been replicated and extended using free and serial recall with delays of up to 60 seconds, and open and closed word pools (Russo & Grammatopoulou, 2003).

2.4.2   Linguistic complexity.

Yet another explanation for the word length effect is the linguistic complexity argument put forward by Caplan, Rochon and Waters (1992). Monosyllabic and polysyllabic words differ on a wide range of possible measures, and in the linguistic complexity hypothesis the word length effect results from differences in complexity of output plans for the short and long words. To test this idea they used words that differed in spoken duration but were matched for number of syllables and phonemes. In their version of the task, participants were to respond by pointing to pictures of the words. Under conditions requiring no verbal output at recall, long spoken duration (more complex) words enjoyed a recall advantage in an apparent reversal of the word length effect. These and similar findings have produced both debate and further empirical evidence (Baddeley & Andrade, 1994; Caplan & Waters, 1994; Cowan, Nugent, Elliott & Geer, 2000; Service, 1998, 2000) and this issue remains unresolved.

2.4.3   *Compilation Errors.*

The Feature Model (Nairne, 1988) uses a different approach to memory, suggesting that, in addition to features of items in a to-be-remembered list, there are also segments, which must be reassembled in the correct sequence to produce accurate recall. The model suggests memory traces degrade over time, but as a result of the probability of becoming overwritten with new information, rather than simply decaying. Recall of list items is achieved when enough features of a degraded trace in a primary memory system correspond to features of an item in a secondary memory system. The secondary memory system is typically assumed to only contain features of the most recent list of items.

This model has been tested using simulations (Neath & Nairne, 1995) and has displayed many classic short-term memory effects. It models word length effects by seeing each list item as itself a list of segments. The numbers of segments do not correspond with the numbers of features of list items, and for the purposes of simulation the numbers of features were held constant between long and short items. Item length was defined as the number of segments in an item, and the reasoning was that longer items with more segments had therefore a greater probability of assembly error, due to the compounding of error associated with each segment. When segments were used as a measure of word length, simulation results showed a relationship between number of segments and correct recall, such that correct recall declined as the number of segments increased from 1 to 13.

Neath and Nairne (1995) also simulated the effect of articulatory suppression, and found it eliminated the word length effect. Their model, however, could not accommodate the findings of LaPointe and Engle (1990) above, who reported a word length effect to be resistant to suppression when a large pool of unique items was used.

The lack of any definition or idea as to what constitutes a segment is also problematic for this account, however its replacement of time-based decay with the probability of matching traces of segments provides a contrasting view of forgetting. If a memory trace can be likened to the striking of a tuning fork, decay theory would have the note gradually lose its intensity until it became inaudible, whereas the feature model would suggest the tone remained undiminished, but with an associated probability that the vibrations of the fork could be suddenly muffled out at any given time.

2.4.4   *Localist Vs Globalist Assumptions.*

A recent study (Cowan, Baddeley, et al., 2003) reviewed these same approaches and then devised a word length and suppression study to test comparisons of what they termed localist and globalist views of recall. A globalist view reasons that recall of items from lists is to some degree influenced by the other items in the list, whereas a localist view concentrates on characteristics of

individual items which affect their recall regardless of the other list items. Using six-item lists for

immediate serial recall in which the proportion of long and short items was varied (short items were

of one syllable, and long items were five syllables, both from small closed pools) from pure lists of

long items through to pure lists of short items, they reported a word length effect between the pure

lists which disappeared under articulatory suppression. Although this effect was statistically

significant (F (1,39) = 4.60, $p.$ < .04) the F-ratio is surprisingly small given the size of the data set and

the magnitude of the ratio for suppression in the same analysis (F (1,39) = 82.48, $p.$ < .001).

Serial recall performance in the mixed lists decreased as the proportion of long items in lists

increased, another effect which disappeared under suppression. Critically for the localist account

(e.g. Neath & Nairne, 1995), a prediction that the proportion correct for long and short items, when

examined separately, should not differ as a result of the proportion of long items in lists was

disconfirmed. In addition, anomalies in some of the results, including the presence of word length

effects under suppression and a clear isolation effect (von Restorff, 1933) in lists with one item of

different length, would not appear to support predictions derived from the Phonological Loop

model in its current form. The authors contended it was possible that participants were using visuo-

spatial storage to aid recall, or that the Phonological Loop was selectively being used for

rehearsing long items only. In general, although their data appeared to support a globalist view,

they could not be well described by the Baddeley model without some major renovations. In fact,

the data did not easily support any of the models described, and the authors concluded that a

different "flavour" of any one of the models would be required, rather than concluding the data

supported a particular model as it now stands.

2.4.5   *Proactive Interference.*

The final explanation included in this chapter is that proactive interference (PI) plays a role

in producing the word length effect. Nairne, Neath and Serra (1997) presented one group of

participants with four lists of five short words and another group with four lists of five long words (the

same stimuli that Cowan et al. (1992) had used in their studies). There was no significant difference

in recall between the groups. A different group received 24 lists of long and short words. The word length effect was not evident on the first four trials, as previously, but became apparent as more lists were presented.

These results supported the notion that the word length effect is related to proactive interference and that PI and word length effects build up over lists. Tehan and Turcotte (2002) have recently attempted to replicate the Nairne et al. (1997) study by using the Cowan et al. (1992) words, but also including a stronger manipulation of word length by varying the number of syllables in the short and long words. The study concluded that PI effects were not reliably observable but word length effects were. These data cast doubt on the viability of a PI explanation for word length effects, yet continued to support the notion of a syllable/phoneme manipulation producing a more reliable word length effect than a manipulation of spoken duration.

2.5 Summary and Conclusions

From the above studies, it can be suggested that traditional trace decay and rehearsal explanations are not sufficient to explain the word length effect under a range of methodological conditions, and alternative theories are at best problematic. In this environment, the possibility exists for another approach to be explored, using a combination of the recall techniques and conditions described above. A suitable perspective to investigate is the item and order processing tradeoff approach described in the first chapter. In many of the memory studies previously described, immediate serial recall is an order task where the word length effect has often been observed. The tradeoff approach to word length would suggest that item recognition for long words should be more successful than for short words, once a recognition task has been incorporated into the paradigm. In the short-term domain, much is known about the sensitivity of word length to a range of variations on the immediate serial recall task. As a further avenue for replication, data abound in the word length literature for such variations as delayed recall, articulatory suppression and irrelevant speech.

The irrelevant speech paradigm refers to an experimental situation where memory tasks, such as serial recall, are performed with an audible distractor, such as a foreign voice speaking in the background (Neath, 2000). In a summary of studies which used irrelevant speech, the robust findings have been that it eliminates, or at least diminishes, the word length effect in both visual and auditory presentation modalities, and affects word length in the same manner as articulatory suppression, though not as seriously (Neath, 2000).

In terms of the Phonological Loop model, the two conditions should affect memory for words in different ways. Irrelevant speech, by providing an additional source of auditory information, interferes with items in the phonological store, degrading their traces in some unspecified manner. Articulatory suppression affects the Phonological Loop by preventing rehearsal of items, which then decay. This crucial theoretical difference between them has been observed in the pattern of results from the above complementary experiments using the word length effect (Neath, 2000).

These different tasks provide known patterns of results for the word length effect, and form a basis for replication of these effects when word length effect is approached from an item/order processing tradeoff perspective. Word length in immediate serial recall, with the addition of a final surprise recognition test, is a suitable starting point, as patterns of recall performance are robust and stable in the literature.

Delayed recall is a further task where serial order recall is required, but is not produced immediately after presentation of the list. A delay period is inserted between the end of the list and the recall cue, and is referred to as a filled delay if that period contains distracting irrelevant information, such as the requirement to repeat random numbers aloud, or perform mathematical operations, or make judgements about stimuli. Tehan et al. (2001) found word length effects after a 12 second filled delay to be very much in evidence, although this pattern is not consistent with predictions from the standard model. While these results have since been confirmed (Russo &

Grammatopoulou, 2003), further replication will establish a pattern for comparison with the other effects (generation and perceptual interference) to be explored in this study.

Test parameters are therefore chosen and word length effects examined under a common framework where a short-term memory account would predict word length effects in immediate but not delayed serial recall. In addition, a further examination can be made of word length effects under more controversial test conditions, such as irrelevant speech and articulatory suppression. An overall pattern should then emerge within the item/order processing perspective which can be related back to the generation and perceptual interference effects to be discussed in the next chapter.

Chapter 3 Contents

CHAPTER 3

Generation and Perceptual Interference Effects

If a to-be-remembered word is not simply read aloud from a list, but has to be generated during study from a stimulus or cue, improved memory for the self-generated words usually results. In a typical generation experiment, participants are presented with two types of words. Some of the words are intact (control items), and are simply read aloud as they are presented to participants. The experimental items are presented as cues, and participants must generate an appropriate response. Memory is then compared for the control and generated words, and the items which had to be generated tend to be recalled more often.

3.1 Improved Recall from Generation

This effect was first explored by Slamecka and Graf (1978), who reported improved recall of items generated from semantic associates as well as rhyme cues. In their first experiment, stimuli in the experimental condition involved presenting a word, an associated rule, and the initial letter of the desired response. An example of an item from the experimental condition was given as "*synonym: rapid – f…*" In this case participants were required to generate a synonym for *rapid* which began with the letter f, i.e. *fast*. Rules used in the experiment were *synonym, associate, category, opposite* and *rhyme*.

Once participants had completed the study phase, they were given a recognition test, which required them to choose the correct response from three alternative answers for each stimulus. Participants were told before the test that the recognition task would follow the study phase. Recognition probabilities were significantly greater for generated items than for those simply read aloud across all rules, and regardless of whether generation was timed or self-paced.

The second experiment introduced changes to the design (from between to within-subjects) and introduced an intentional/incidental learning manipulation, with half the participants informed of the recognition test, while the other half were not. Neither of these manipulations

made a difference to the pattern found in the previous experiment. Slamecka and Graf (1978) reasoned that generated items required more attention and processing than the simply read items, and in further experiments demonstrated that the generation effect was occurring with responses and not with stimuli (at output but not at study), and in cued and multitrial free recall as well as in recognition tasks.

3.2 A Processing Account

Although the study reported above was largely exploratory and descriptive, suggestions were made by the authors as to the potential theoretical background of the effect. Among their speculations was the possibility that the generation effect could be explained from a processing perspective, with the self-generated items requiring a "more profound processing level" than simply read controls (Slamecka & Graf, 1978, p. 602). Other possible explanations they advanced, such as a paired-associates approach, where the generation task "forces a distinctive encoding of the relation between stimulus and response" (Slamecka & Graf, 1978, p. 603), would by their nature presumably require more intensive processing than controls.

Further studies revealed generation of items from antonyms, definitions, translations from other languages, and from abbreviations and word fragments all led to improved recall (Mulligan, 2001). The effect has been observed in such memory tasks as free recall, cued recall, and recognition, as well as in both intentional and incidental learning situations (Burns, Curti, & Lavin, 1993). To date, the generation effect has been described only as a long-term memory phenomenon (Burns et al., 1993; Nairne, Riegler, & Serra, 1991) although some of these studies have employed experimental situations more common to short-term memory research.

3.3 Reversed Generation Effects

In the majority of generation effect studies since Slamecka and Graf (1978) a robust recall advantage for generated items has been demonstrated under a wide range of conditions. Conflicting data supporting a reversed generation effect (where the items which were simply read

aloud were better remembered than the generated items) have been reported in a series of experiments by Burns (1986; Burns et al., 1993).

Burns and his colleagues (1993) studied generation effects using measures of immediate and delayed recall and order reconstruction, as well as a final recognition test. Generation was manipulated by using word fragments as stimuli, with a missing letter replaced by a hyphen in the word, such that there was only one legal solution to the fragment. They found results to vary with delay period and distractor difficulty, such that control items were better recalled than generated items under an 'easy' distractor (making odd/even judgements on numbers) and generated items were better recalled than controls under a 'difficult' distractor (recalling numbers), but only after an 80-second delay. They did, however, find a general advantage for generated items over controls on an unexpected final recognition task in each case, as did DeLosh & McDaniel (1996). Patterns such as these have led some researchers (e.g. Burns et al., 1993; Nairne et al., 1991) to conclude that generation has opposite effects on item and order processing.

3.4 Item and Order Processing Tradeoffs in Generation

Partitioning memory for a list of items into item and order processing has been found useful in a number of relatively recent memory studies, and is deemed necessary because different memory tasks as used in the above studies seem to be measuring different aspects of list memory. Immediate serial recall and order reconstruction (where the items from a recently-presented list are re-displayed in a different random or alphabetical order, to be reconstructed into their original order of presentation), for example, involve the necessity to remember list items in their correct presentation order, whereas tasks such as recognition and free recall do not necessarily require the retention of order information, focussing instead on the items themselves.

Order (or relational) processing has been assessed in the above generation studies (Burns et al., 1993; DeLosh & McDaniel, 1996; Engelkamp & Dehn, 2000; Nairne et al., 1991) using an order reconstruction task or free recall. An order reconstruction task would seem to be independent of information specific to the individual items, as all the items from study are provided.

Item-specific processing, in those studies which examined it (Burns et al., 1993; Engelkamp & Dehn, 2000; Nairne et al., 1991), has generally been measured using a recognition test, where words from the studied lists are randomly mixed with distractor items which were not in the experiment. The test is typically given after the conclusion of the experiment, with items either presented individually or on a list, and participants asked to identify whether or not each word had been used in the experiment.

A general pattern has emerged which suggests that generating items has a positive effect on item retention, with a corresponding negative effect on order retention (Engelkamp & Dehn, 2000; Mulligan, 2001). The direction of generation effects seems to be dependent upon the type of recall test used. If memory for order is required, the negative effect is found, but when memory for individual items is tested, the standard generation effect is observed (generated items are remembered more often).

Exceptions extend mainly to situations where recall is required after a long (80 second) delay (Burns et al., 1993) or where the number of generated items within lists was manipulated (Kelley & Nairne, 2001). These results are not necessarily at odds with the general pattern, as the long delay in the Burns et al. (1993) study mirrored the generation advantage found in final recognition tests, while the mixed-list manipulation of Kelley and Nairne is likely to have produced a special case of the von Restorff (isolation) effect, as noted in their discussion.

Notwithstanding the above exceptions, a reasonably robust pattern of generation effects can be predicted in order reconstruction and item recognition tasks. The advantage of generated words over control items in final recognition tests is undisputed to date (Burns et al., 1993; DeLosh & McDaniel, 1996; Engelkamp & Dehn, 2000; Greene, Thapar, & Westerman, 1998; Nairne et al., 1991; Mulligan, 2001). Additionally, the recall advantage for simply read words in order tasks has been found over a range of presentation rates, varying from 500 to 3000 milliseconds, and in lists of six or eight items. These parameters seem eminently suited to be

generalised to experimental designs (such as immediate serial recall) traditionally used in short-term memory research, as noted in the previous word length effect chapter.

In summary of the generation effect, one of the critical assumptions of this dissertation is addressed – that there is a 'difficulty' dimension underlying the generation effect. When considered from a processing perspective, it does not seem unreasonable to suggest that items which must be generated from a cue would require more mental processing than control items which are simply read aloud by participants, as originally suggested by Slamecka and Graf (1978). The additional processing load involved in generating a word from a cue adds a level of complexity relative to an equivalent control item, which is seen by Nairne et al. (1991) as enhancing item-specific information in the memory trace, but at the expense of order-relational information. That is, the above results appear to be showing a tradeoff in item and order processing, to the extent that additional processing of an item relates to a decrement in the available information about its order on a list.

3.5 The Perceptual Interference effect

Although a relatively new phenomenon in the memory literature, the perceptual interference effect, first described by Nairne (1988) is beginning to provide data which support the item-specific/order-relational processing distinction currently being advanced to explain the generation effect. If the perception of a word is interfered with (backward masked) during presentation, improved memory for that word relative to another item displayed without interference tends to result (Mulligan, 2000).

In a typical study of the effect, items in the experimental lists are displayed very briefly (for between 100 and 266 ms) then backward masked with neutral symbols for the remainder of the display period. Control items are displayed in their entirety for the same period, typically 2500 ms. For lists in both conditions, participants are instructed to read the words aloud as they see them. This manipulation has led to improved item recall in the experimental (perceptual

interference)condition across a variety of tasks including, but not limited to, recognition, free recall and cued recall (Mulligan, 2000).

3.6 Processing Tradeoffs in Perceptual Interference

Akin to the effects of generation described earlier, however, there are instances when perceptual interference effects are eliminated or reversed. Perceptual interference has been shown to disrupt order memory, when measured by an order reconstruction test. Mulligan (2000) found that perceptual interference disrupted performance on tests of absolute order in the same manner as it did with generation (Greene et al., 1998). In fact, similarities in memory test performance between perceptual interference and generation are quite striking. Both, under similar conditions, appear to enhance the processing of individual items at the expense of order processing (Nairne, 1988; Westerman & Greene, 1997).

Other similarities between perceptual interference and generation include the observation that both effects are less likely to occur with unfamiliar words and nonwords (Westerman & Greene, 1997), and that both effects are found to be more prevalent in within-subjects than in between-subjects designs (Mulligan, 2000). Such effects are also likely to be stronger in mixed lists (lists containing, for instance, read items as well as generated items) than in pure lists (DeLosh & McDaniel, 1996), and, by extension, may even be a function of list length (Engelkamp & Dehn, 2000). In a similar manner to the generation literature, perceptual interference has been mainly studied from a long-term memory perspective, albeit under conditions which have many features in common with the traditional approach to short term memory studies involving the word length effect previously described. In both cases, the theoretical long-term/short-term distinction is becoming less important as a more generalised processing view is applied.

3.7 Chapter Summary

If two different memory effects reveal a similar pattern of results when viewed from an item/order processing perspective, there is some reason to suspect a general processing view may be able to be taken with further memory effects not previously examined from that perspective.

The following chapter integrates the separate streams of research described above with the word length effect from the previous chapter, to result in a design which will allow the relative contributions to item and order memory of the three memory effects detailed to be examined on a common platform.

Chapter 4 Contents

CHAPTER 4.

Synthesis: Word Length, Generation and Perceptual Interference - A Processing View

4.1 Processing Dissociations in Other Memory Effects

The concept of a tradeoff between item-specific and order-relational processing suggested in the generation effect and perceptual interference literature has also been advanced to explain dissociations in recognition and order reconstruction or free recall performance in a number of other domains. Thus, word frequency (DeLosh & McDaniel, 1996), bizarreness (McDaniel, Einstein, DeLosh, May & Brady, 1995) and enactment (Engelkamp & Dehn, 2000) effects have all been investigated and explained in terms of differential item and order encoding. In each case, it appears that adding any extra processing load associated with difficulty or complexity in one set of items relative to a control set will benefit item processing but diminish order processing.

4.2 A Processing View of Word Length

From the conclusion in Chapter 2, no existing hypothesis can fully and adequately explain the word length effect. Therefore, there is sufficient reason to re-examine it from the perspective of processing, which has become an effective explanation of the generation and perceptual interference effects. To begin, an obvious assumption is that long words necessitate more processing than short words, and justification for the assumption is derived from the lexical memory literature. Across a range of lexical access tasks short words are processed more quickly than long words (Forster & Chambers, 1973; Balota & Chumbley, 1985; Samuels, Laberge & Bremer, 1978). In this area of research, word length is often operationalised in terms of the number of letters and the tasks used are limited to visual presentation. The operational definition of word length and the type of task used to explore the effects are thus very different from those used in the short-term memory domain, however the data are worthy of consideration for two reasons.

Firstly, lexical access variables play an important part in span (Brown & Hulme, 1995; Tehan & Lalor, 2000) so the possibility remains that the word length effects in short-term recall might in some way be related to lexical access differences. Secondly, the item/order processing tradeoff has provided a viable explanation for other effects in the memory literature, and may readily extend to the word length effect if indeed there are processing differences between items of differing length.

Given that there appears to be an item/order trade-off pattern emerging with varying types of 'difficulty' or 'complexity' manipulations exemplified by generation and perceptual interference, it would seem reasonable to extend this hypothesis to the word length effect. If long words are more difficult to process, processed more slowly or require additional item processing relative to short words, fewer resources would subsequently be available for order processing. As such, the processing tradeoff approach predicts that a short word advantage should be apparent with order tasks. However, the extra item processing associated with the long words should produce a long word advantage in item tasks.

4.3 Rationale for the Current Design

In order to directly compare patterns of processing tradeoff effects across word length, generation and perceptual interference, a design is required which permits systematic examination of all three effects within a framework which allows direct comparisons to be made. Disregarding memory models which contain separate components with related memory tasks suited to each, the current approach requires simply a task which measure item processing, combined with one which assesses order processing. These tasks are linked within an experiment whose parameters are suited to exploring all three memory effects while allowing for replication of existing data. The experimental framework developed by Nairne et al. (1991) is considered a useful starting point because it allows for item and order processing to be examined concurrently.

In their (1991) generation effect study, participants were presented with 24 lists, each consisting of eight unique items. On half the lists, each of the items was presented as a word

fragment (with one letter missing, for example 'f_sh') and participants were required to generate a word from that cue (in this case 'fish'). Fragments were carefully constructed so that there was only one possible legitimate English word which could be made from each fragment. The remaining lists contained complete words that were simply read aloud. Participants were instructed that they needed to be able to recall the eight items on each list after a 30 second distractor activity.

After the filled delay, the provided recall cue determined the participant's response. If a line of asterisks was presented, participants were instructed not to respond, but to simply wait for the next list. However, if no asterisks appeared, the items from the list were instead presented in a different random order and participants were requested to rearrange the words into their original presentation order.

As well as the order reconstruction task, the experiment also contained a final surprise recognition test, following the presentation of the lists. Participants were shown a sequence of items one by one, comprising items from the asterisk lists as well as novel distractors and asked to indicate if each had been presented during the experiment. As this was a generation effect study, both the order reconstruction lists and the recognition (asterisk) lists were divided into equal numbers of generation and simply read lists. As noted earlier, the read items were better recalled than the generated items on the test of order memory, but the generate advantage emerged with the item test.

4.4 Modifications to the Nairne et al. (1991) Paradigm

In the above experiment, order reconstruction of relatively long lists after an extended retention interval was used to measure memory for serial order. In the short-term memory studies involving word length effects described in Chapter 2, immediate serial recall of shorter lists has often been used. With the intent to integrate (and replicate) existing word length effects within a common design, order reconstruction will be replaced with immediate serial recall. Immediate serial recall is eminently suited to the objective, as it firstly requires the maintenance of order

information, and secondly it provides a framework within which item processing has to occur within a limited amount of time, typically with item presentation rates of one per second.

In short, the current thesis uses a modified version of the Nairne et al. (1991) procedure. The lists are shorter and order memory is tested by immediate serial recall rather than delayed reconstruction, and the final item recognition test is retained. The filled delay used by Nairne et al. (1991) in the order reconstruction task can remain in the design by the use of the same type of delayed serial recall task as that used in the word length literature.

4.5 Hypotheses of the Current Study

Combining the immediate serial recall task with the existing recognition test in the same basic framework should allow processing tradeoff effects between items of differing complexity to be observed in the same manner as before. When generation and perceptual interference effects are re-examined in the modified paradigm, negative generation and perceptual interference effects should be observed in immediate serial recall, but positive generation and perceptual interference effects should be found in recognition. The first goal of this dissertation is therefore to establish whether or not the generation and perceptual interference effects described will replicate from order reconstruction into immediate serial recall.

Having replaced order reconstruction with serial recall, the word length effect can be easily introduced into the experimental framework. The processing tradeoff approach would suggest that item recognition for long words should be greater than for short words. In order to directly compare patterns of tradeoff effects across word length, generation and perceptual interference, the second goal of this dissertation is to replicate the processing tradeoff found with generation and perceptual interference and extend it to the word length effect.

The hypothesised processing tradeoff effect should be readily observable within such a design. All three effects are therefore assessed using this common platform based upon the modified paradigm of Nairne et al. (1991). From the studies reported in Chapter 2, much is known about the sensitivity of word length to a range of variations on the immediate serial recall task. As

a further avenue for replication, data abound in the word length literature for such variations as delayed recall, irrelevant speech and articulatory suppression. Known effects for word length can be replicated and compared with the unknown effects of generation and perceptual interference when tested under these conditions.

If it can be assumed that a unitary memory system is involved, and that memory performance on different tasks is purely a function of different kinds of processing, then the current series of experiments poses the following questions: Will the word length effect show the same dissociations as the generation and perceptual interference effects; and will the generation and perceptual interference effects show the same sensitivity to immediate recall manipulations as the word length effect? In sum, the question is whether generation, perceptual interference and word length will generate the same pattern of effects in a common framework assessing item and order processing. If so, then parsimony would suggest a common theoretical foundation.

The experiments which follow are organised into three chapters: Those dealing with the word length effect in Chapter 5, the generation effect experiments in Chapter 6, and the perceptual interference effect in Chapter 7. These three effects are examined under conditions of immediate serial recall and recognition which are as much as possible held constant throughout the entire study, and for purposes of replication and extension of existing data, the serial recall component in each effect is tested without a delay, with a delay, and under conditions of irrelevant speech.

Chapter 5 Contents

CHAPTER 5.

Word Length Effect Experiments

5.1 Introduction and Predictions

When word length is defined in terms of the number of syllables or phonemes in a word, the effect is extremely robust in immediate serial recall. The Tehan et al. (2001) results indicate that the word length effect can survive 12 seconds of distractor activity and there is evidence that it will also survive other forms of interference such as background irrelevant speech. In short, the word length effect is a widely studied phenomenon in the short-term memory literature. However, in Chapter 2, it has been argued that there is no universally accepted view of the word length effect. The experiments in the current chapter test a processing account of the word length effect, in which it is assumed that short and long words receive differential amounts of item and order processing while studying a list for ordered recall.

According to this approach, short words receive more order processing than long words, whereas long words receive more item processing than short words. The observation that short words are better remembered than long words in immediate recall, delayed recall and irrelevant speech experiments is explained by the fact that in all these experiments memory is tested via a serial recall task. The key prediction of the processing account is that a long word advantage should be found when only item information is required. The current experiments test this prediction by exploring word length effects in serial recall and item recognition under conditions of immediate recall, delayed recall and irrelevant speech. The expectation is that the standard word length effect will emerge on the serial recall task, but the reverse effects will emerge on a final item recognition task.

In order to replicate and extend the tradeoff effect described in previous chapters, two components to each experiment were necessary: An initial serial recall phase, followed by a later recognition phase. As such the Nairne et al., (1991) procedure was slightly modified for current purposes. As detailed below, although the two phases differed in presentation and instructions,

they utilised the same materials throughout. To link the two phases of each experiment together, the serial recall phase contained a large proportion of lists which were not recalled by participants in serial order, although they were learned during presentation as if they would have to be. Participants were not made aware until after the end of each list whether or not it would have to be recalled in serial order. Recall lists were followed by a series of question marks, and non-recall lists by a series of asterisks. Those lists which were not immediately recalled in order then became part of the recognition test.

5.2 General Method

5.2.1 Participants. Introductory psychology students from the University of Southern Queensland volunteered to participate in each 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. Each person participated in only one experiment.

5.2.2 Materials. For experiments 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 mean Kucera-Francis frequency of the short words was 7.79 ($\underline{SD}$ = 8.14), while for long words the mean frequency was 8.22 ($\underline{SD}$ = 10.43). Mean concreteness ratings for short words were 550.40 ($\underline{SD}$ = 52.74), and for long words 564.86 ($\underline{SD}$ = 64.85).

Each participant in the experiment received a uniquely ordered set of lists. To create the lists, 90 words were randomly selected from each pool, and then randomly assigned to the six serial positions on each list. This effectively produced 15 six-word lists of long words, and 15 six-word lists of short words. The 15 lists in each condition were randomly divided into 10 non-recall (asterisk) lists, and 5 immediate serial recall (question mark) lists. The order of the resulting 30 lists was then

randomised. Resulting computer files were duplicated to provide a hard copy for the experimenter to record participant responses.

The recognition component of the experiment comprised the 10 non-recall, asterisk lists (60 long and 60 short words) from the study phase, as well as filler words which comprised the unused 30 long and 30 short words from the initial pools, again producing a unique word list for each participant. The recognition test was created by randomly-ordering the list of 180 words and arranging them in six columns of 30 words on a single sheet. A matching scoring sheet was also produced for the experimenter which retained serial position information and identified filler words.

5.2.3 Procedure.   Participants were tested individually in a quiet laboratory, in sessions of approximately 35 minutes' duration. Written instructions were provided to participants at the beginning of the session, stating they would be shown lists of words which they were to study silently, then recall in their correct serial order when cued by a series of question marks. To preserve the order of remembered items, they were asked to substitute the word "pass" if they could not remember a particular word. Additionally, participants were told that not all lists would contain the recall cue, and that lists followed by a series of asterisks were not to be recalled aloud; thus when a row of asterisks appeared on the screen, they were not to respond and were to simply wait for the next list to begin. No rationale was given for this instruction. A practice trial of six lists was then conducted, and when the experimenter was satisfied that the participant understood the instructions and could perform the tasks, the test commenced.

The immediate serial recall phase was administered on a Macintosh microcomputer with a monochrome screen using Hypercard software. Each list began with a beep and a "READY" sign appeared on the centre of the computer screen for three seconds. The six words were then presented one at a time, in lowercase, in the centre of the screen, at the rate of one word per second. The end of each list was signified by either a row of question marks or a row of asterisks, which remained on the screen for 2 seconds. 12 seconds were then allowed for participants to

recall the six words, if required, or to wait silently before the next list commenced. The experimenter recorded responses from participants on a hard copy as either correct in position, omitted, transposed or extralist intrusions. Totals for the five long and five short word recall lists (preserving serial position) provided the basis for the measures of immediate recall.

Presentation of the 30-list immediate serial recall phase was followed by the recognition phase of the experiment after a three-minute delay during which administrative details were completed. Participants were given the recognition sheet, and were asked to simply circle or tick any words they remembered seeing in the experiment they had just finished. They were given as much time as they required to complete the task, then were told about the aims of the study.

Numbers of correctly-recognised words from the long and short lists were totalled, preserving original list position, to provide measures of recognition. Filler words incorrectly identified as being from the immediate recall phase (false alarms) were recorded separately.

5.3 Experiment 1. Word Length & Immediate Serial Recall

5.3.1 Participants. 19 introductory psychology students participated in the experiment.

5.3.2 Materials and Procedure. No changes were made to the General Method detailed above.

5.4 Experiment 1 Results

5.4.1 Rationale for Analyses: Two approaches are used to explore the data in each experiment: An individual differences approach and a group means approach.

It is well known that participants can adopt a wide range of strategies in doing the immediate serial recall task (Logie, Della Sala, et al., 1996), from rehearsal to visual encoding, to image chains, to recency based recall. The above hypotheses critically depend upon the person using a serial recall (rehearsal of list items in serial order) strategy when it comes to doing the serial recall task. As such, individual recall protocols were examined firstly to see if the trade-off effect emerged and secondly to determine the strength of that effect. The strength of the trade-off

effect was then examined with reference to whether or not the participant had adopted a forward serial recall output or produced a recency based, backward recall strategy.

Recognition performance was also examined on an individual differences level. Again it is well established in the literature that the criteria for making a "yes" decision varies from person to person and this can show up in differential false alarm rates. Thus, false alarm rates are also reported.

At the level of group means, the design in each experiment is primarily a 2 (word length) x 2 (type of task) x 6 (serial positions) within subjects design. An analysis of variance with this design will produce the three main effects, the three two-way interactions and the three-way interaction. Of these only one is relevant for current purposes. The hypotheses critically depend upon a word length by type of test interaction. The approach adopted here is to initially determine if this interaction is significant and then to do a simple effects analysis by examining separate 2 (word length) x 6 (serial position) analyses for each type of test. For the hypotheses to be supported, a significant main effect for word length must emerge in each of these analyses, and the two main effects must take opposite forms.

5.4.2 Individual Differences. Using the raw data set, estimates were made of the pattern of effects for each individual, and data were appropriately summarised to compare overall treatment group means. From the raw data, the participant's total correct responses from all serial positions from short and long lists were compared for both immediate serial recall and recognition lists. If the cell totals differed by 2 or 3 items, a weak effect was recorded, a moderate effect for a difference of 4 to 6 items, and a strong effect recorded for a difference of 7 items or more.

Directions of effects were recorded as either a short word or a long word advantage, for both immediate recall and recognition data. The expected tradeoff pattern is characterised by a short word advantage at immediate serial recall, switching to a long word advantage at recognition. Of the nineteen participants tested, 11 showed a strong tradeoff effect, a further three showed a weak to moderate effect, two displayed a consistent long word advantage across

memory tasks, and one a consistent short word advantage. Two participants showed no word length effect on the immediate recall test, but a long word advantage at recognition.

Serial position data were assessed to determine direction of recall strategy, using only items recalled in their correct serial positions. For the purposes of heuristic comparison, a forward strategy describes participants who generally could only remember the first few items of a list, and a recency strategy describes participants tending to remember only the last few items on the list. Again cell totals were compared, with effects rated in the same manner as word length above.

Participants' use of strategies was estimated and of the three who appeared to have been exclusively using a recency strategy, all showed the expected tradeoff. Of the six participants exclusively using a forward recall strategy, four showed a tradeoff effect, one a constant short word advantage, and one showed no effect in the immediate serial recall test, but a long word advantage at recognition.

Four participants appeared to vary their strategy with word length on a list-by-list basis, three using forward recall with short words then switching to recency for long lists, and one using a recency strategy with short words, switching to a forward recall technique for the long words. Three of these four showed the expected tradeoff effect, and one (forward to recency) displayed a consistent long word advantage.

Six participants showed no discernible forward or backward preference at recall, and of these four showed the expected tradeoff effect, one showed a constant long word advantage, and one showed no effect at immediate, but a long word advantage at recognition.

The proportions of false alarms shown in Table 1 are likely to have been affected by outliers, with four participants having made more then ten errors. If those four are left out of the data, the proportions of false alarms fall to 0.06 (short) and 0.11 (long words).

5.4.3 Group Data. Summary data from the groups are displayed in Table 1.

Table 1. Summary of group data from Experiment 1.

Mean Recall	Immediate Serial Recall			Recognition	
Probability	Correct	Transpos	Extralist	Correct	F Alarms
Short words	0.46	0.09	0.12	0.35	0.10
Long Words	0.35	0.09	0.05	0.52	0.13

 Note. In Immediate serial recall, 'Correct' refers to proportion of items correct in position across all serial positions while 'Transpos' refers to items remembered out of their correct serial positions. 'Extralist' represents the proportion of responses which were not items from the studied lists. In Recognition 'Correct' is the proportion of correctly recognised items, and 'F Alarms' shows the mean number of filler items incorrectly recognised.

An initial analysis was performed on data from both immediate recall and recognition tasks combined into a 2 (task) x 2 (word length effect) x 6 (serial position) repeated measures ANOVA. The task by word length interaction was found to be significant, F (1,18)= 53.54, MSE= 2.71, $p.$ < .001. This interaction showed a short word advantage at immediate recall, reversing to become a long word advantage in the recognition phase. For clarity and completeness, the individual analyses for each task are reported below.

5.4.3.1 Immediate Serial Recall. Analysis of words recalled in correct serial position from the immediate recall task was undertaken using a 2 (word length) x 6 (serial position) repeated measures ANOVA. A significant main effect was found for word length, F (1,18)= 16.14, MSE= 1.13, $p.$ < .001. A significant main effect was also evident for serial position, F (5,90)= 5.95, MSE= 2.02, $p.$ < .001. The interaction was nonsignificant, F (5,90)= 1.32, MSE= 1.39, $p.$ = .26. The serial recall curves for the immediate recall phase are displayed in Figure 1, and show the expected short word advantage.

Figure 1. Immediate Serial Recall curves from Experiment 1.

Word Length Effect & Immediate Serial Recall: Correct In Position
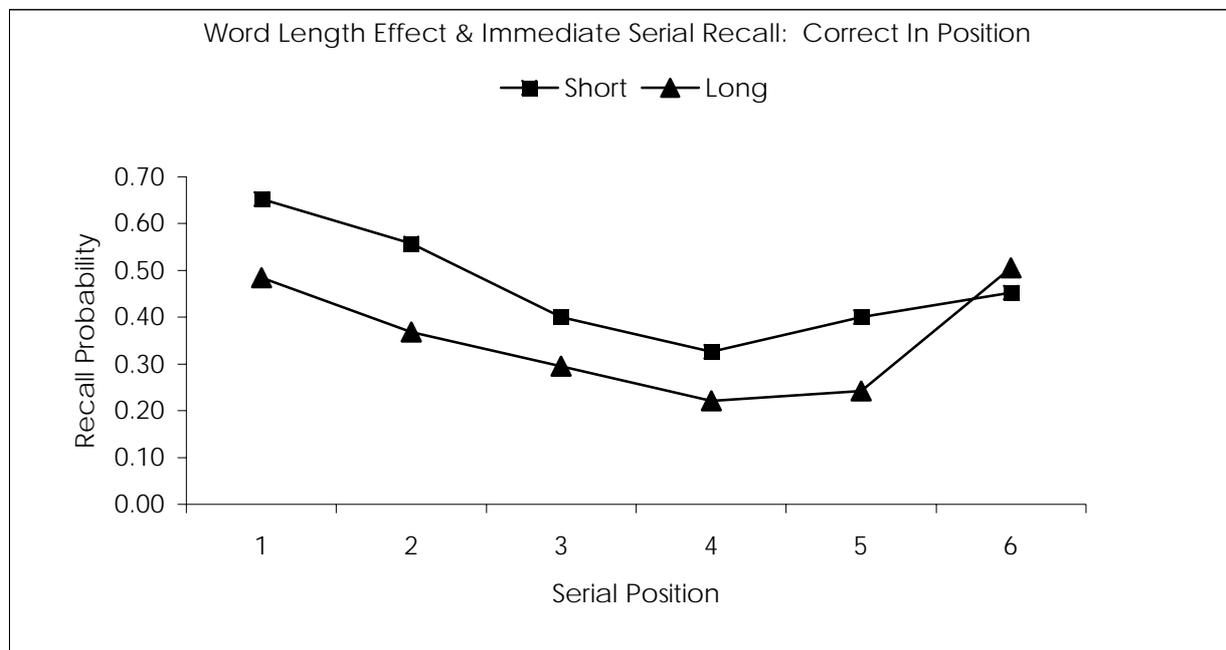
5.4.3.2 Recognition. An identical analysis was performed on data from the recognition test. There was a significant main effect for word length, $F_{(1,18)}= 34.24$, $MSE = 4.77$, $p. < .001$. Figure 2 shows more long words were recognised than short words across all serial positions. A significant main effect for serial position, $F_{(5,90)}= 32.51$, $MSE = 2.98$, $p. < .001$, is also illustrated in Figure 2. There was no significant interaction, $F_{(5,90)} < 1$.

Figure 2. Recognition curves from Experiment 1.
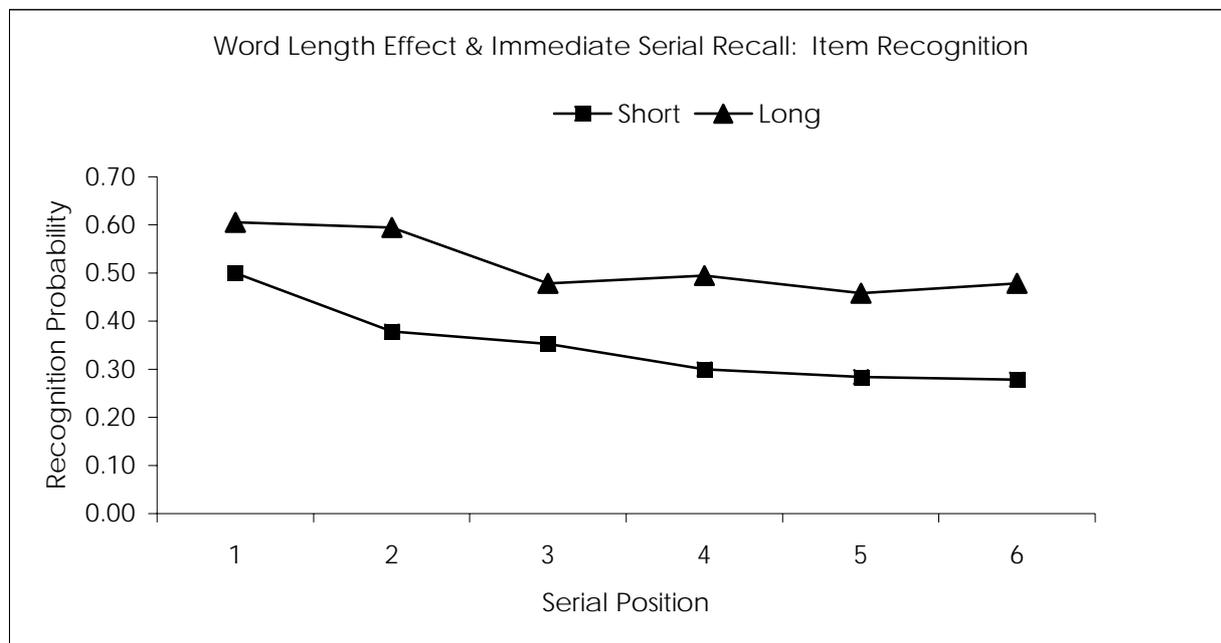
5.5 Experiment 1 Discussion

The individual differences data suggest that within the group, individuals appear to be using varying methods of recalling the items. Regardless of strategy and pattern of effects within individuals, however, the majority showed the expected tradeoff effect, as evidenced by the highly significant interaction found in the combined ANOVA. A short word recall advantage in the immediate serial recall phase reversed to become a long word advantage at the recognition phase.

The serial recall curves shown in Figure 1 conform to predictable shapes, and despite the apparent recency advantage for long items, the lack of a significant interaction shows classic serial recall position effects do not tend to change with the length or difficulty of the stimuli. The curves for recognition, however, show a flatter aspect, with mild yet significant primacy for both item types reflected in the main effect for serial position. Recognition scores were unlikely to have been unduly influenced by errors, as the incidence of false alarms was not considered to be high.

Table 1 reveals that transpositions were effectively the same for both conditions, and the item advantage for long words found at recognition is not previewed at the immediate serial recall phase by an increased number of transpositions in long word lists. In summary, the proposed tradeoff effect appears to be strongly evident with the word length effect and immediate serial recall, and the following experiment explores this further under conditions of delayed recall.

5.6 Experiment 2. Word Length & Delayed Recall

5.6.1 Participants. 20 introductory psychology students from the University of Southern Queensland volunteered to participate in the experiment. None of these had participated in Experiment 1.

5.6.2 Materials and Procedure. The only changes made to the General Method detailed in the above section involved the insertion of a four second, filled delay between the end of each list and the recall cue. Participants read aloud a series of four, four-digit numbers which appeared on the computer screen immediately after the sixth item in the list and before either the immediate recall signal or the asterisks appeared. Instructions and computer files were adjusted accordingly.

5.7 Experiment 2 Results

5.7.1 Individual Differences. Estimates of individual patterns were again made from the raw data set, and data were summarised to compare overall group means. Of the twenty participants, five showed a strong tradeoff effect, a further four showed a weak to moderate effect, three displayed a consistent long word advantage across recall tasks and two showed a consistent short word advantage. Four participants showed no word length effect on the immediate recall test; three of those demonstrated a long word advantage and one a weak short word advantage at recognition. One participant showed the exact reverse of the expected tradeoff, and one showed no word length effect in either task.

The mean numbers of false alarms shown in Table 2 are again likely to have been affected by outliers, with three participants having made more then ten errors. If those three are removed from the false alarm data, the proportions drop to 0.05 (short words) and 0.08 (long words).

Participants' use of strategies was estimated and of the six who appeared to have been exclusively using a forward recall strategy, three showed the expected tradeoff effect, one no effect at all, and two produced no effect at immediate recall but a long word advantage at recognition. No participants appeared to have been using a backward strategy.

Three participants used a forward strategy only with short word lists, and then produced the expected tradeoff effect, and ten did not appear to be using either strategy. Of these ten, three showed the expected tradeoff and one showed a weak reverse tradeoff effect, three demonstrated a consistent long word advantage, one a consistent short word advantage, and two had no effect at immediate, but one showed a long word and one a short word advantage at recognition. The remaining participant's pattern of responses comprised a forward strategy with short lists, switching to a backward strategy for long word lists.

5.7.2 Group Data. Summary data from the groups are displayed in Table 2.

Table 2. Summary of group data from Experiment 2.

Mean Recall	Delayed Serial Recall			Recognition	
Probability	Correct	Transpos	Extralist	Correct	F Alarms
Short words	0.17	0.23	0.04	0.43	0.07
Long Words	0.09	0.22	0.03	0.58	0.12

Note. Values and measures were derived in the same manner as those in Table 1.

As before, an initial analysis was performed on data from both serial recall and recognition tasks combined into a 2 x 2 x 6 repeated measures ANOVA. The task by word length interaction was found to be significant, $F(1,19)= 22.58$, $MSE= 4.67$, $p. < .001$. This interaction again reflected a short word advantage in the immediate recall phase, reversing to become a long word advantage at recognition. The individual analyses for each task are reported below.

5.7.2.1 Delayed Recall. From the delayed recall phase, a significant main effect was found for word length, $F(1,19)= 6.47$, *MSE*= 1.42, *p.* = .020. Figure 3 shows short words were recalled more often than long words. A significant main effect was also evident for serial position, $F(5,95)= 23.19$, *MSE*= 0.76, *p.* < .001. The interaction was significant, $F(5,95)= 5.60$, *MSE*= 0.36, *p.* < .001, indicating the word length effect was strongest in the primacy portion of the curve.

Figure 3. Serial recall curves from Experiment 2.
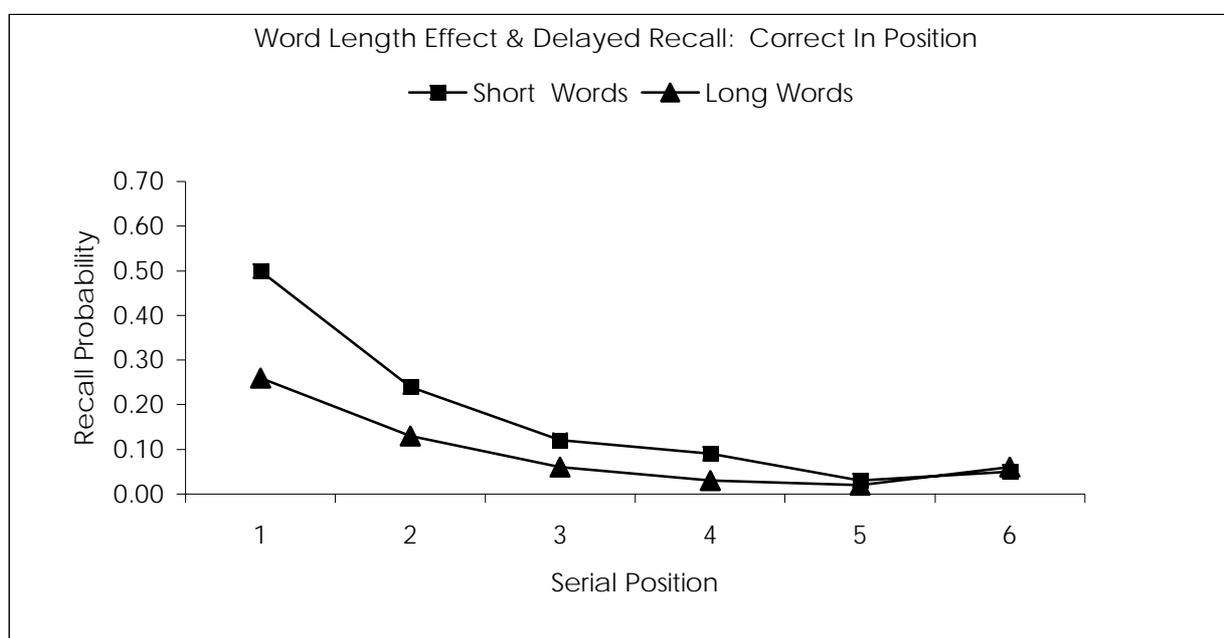
5.7.2.2 Recognition. An identical analysis was again performed on data from the recognition test. There was a significant main effect for word length, $F(1,19)= 20.38$, *MSE* = 6.69, *p.* < .001. Figure 4 shows more long words were recognised than short words across all serial positions. A significant main effect for serial position, $F(5,95)= 2.813$, *MSE* = 2.37, *p.* = .021, is also illustrated in Figure 4, evident in a wider separation in the recency portion of the curves. There was, however, no significant interaction, $F(5,95)= 1.52$, *MSE* = 2.22, *p.* = .191.

Figure 4. Recognition curves from Experiment 2.
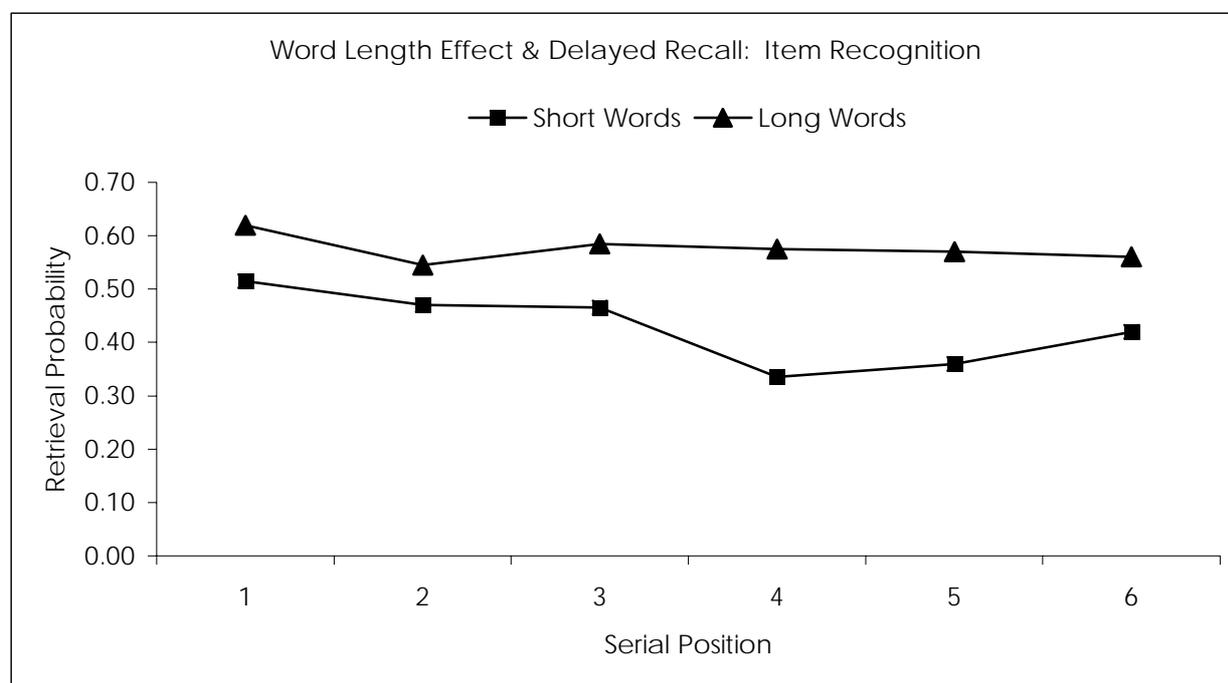
5.8 Experiment 2 Discussion

The individual differences data again suggest that although individuals appeared to be using widely varying strategy combinations, their choice of strategy had no systematic influence on the pattern of results. As before, the majority showed the expected tradeoff effect, the pattern from Experiment 1 recapitulated in Figures 3 and 4, supported by the ANOVA interaction from the combined group data. A short word advantage from the delayed serial recall phase reversed to become a long word advantage at recognition.

Using a filled retention interval generally reduced group mean results in the immediate phase, but did not affect the tradeoff in word length effects which replicated from Experiment 1. Comparing the overall treatment means from Tables 1 and 2, it is noted that the differences between them (word length effect) remain similar under delayed recall conditions. A table of overall effect sizes from all three word length experiments is included at the end of this chapter.

The serial recall curves in Figure 3 show a lack of recency with recall of later list items on floor for both short and long words. The interaction from the ANOVA relates to the presence of a word length effect in the primacy portion of the curve, which then disappears in the later serial positions.

The recognition curves from Figure 4, however, show a clear separation and reverse word length effect. Recognition scores were unlikely to have been unduly influenced by false alarms, as their incidence was generally low for most participants.

The numbers of transpositions and extralist intrusions were again approximately equal. The effects shown in these data are evidently not being influenced by variations in these measures, and the use of alternative scoring techniques is not warranted.

5.9 Experiment 3. Word Length and Irrelevant Speech

5.9.1 Participants. 20 introductory psychology students from the University of Southern Queensland volunteered to participate in the experiment. None of these had participated in Experiments 1 or 2.

5.9.2 Materials and Procedure. The only changes made to the General Method detailed at the start of this section involved the presentation of irrelevant sound during study of lists. News broadcasts in Russian were played at a clearly audible volume from speakers attached to the computer. The speech commenced with presentation of the first item on every list, and continued for six seconds until the recall cue appeared. Instructions stressed that participants were to ignore this speech as best they could.

5.10 Experiment 3 Results

5.10.1 Individual Differences. Estimates of individual patterns were again made from the raw data set, and data were summarised to compare overall group means. Of the twenty participants in Experiment 3, 13 showed a strong tradeoff effect, a further four showed a weak to moderate

effect, one displayed a consistent long word advantage across recall tasks and two participants failed to demonstrate a word length effect in either task.

The false alarm data shown in Table 3 are again likely to have been affected by outliers, with three participants recording more then ten errors. Leaving those three out of the false alarm data, the proportions fall to 0.06 (short words) and 0.08 (long words).

Participants' use of strategies was estimated and of the sixteen who appeared to have been exclusively using a forward recall strategy, 15 showed the expected tradeoff effect, and one showed a consistent long word advantage in both tasks. No participant appeared to have been exclusively using a backward strategy.

Three participants used a forward strategy only with short word lists, two of these demonstrating the expected tradeoff, the other showing no word length effect on either task. No word length effects were evident from the final participant, who demonstrated a weak backward strategy with short lists only.

5.10.2 *Group Data*. Summary data from the groups are displayed in Table 3.

Table 3. Summary of group data from Experiment 3.

Mean Recall	Immediate Serial Recall			Recognition	
Probability	Correct	Transpos	Extralist	Correct	F Alarms
Short words	0.38	0.13	0.09	0.32	0.08
Long Words	0.25	0.15	0.05	0.49	0.11

Note. See notes to Table 1 for measures and derivation of scores.

The initial analysis of combined data from both immediate and recognition tasks was conducted as before, with a significant task by word length interaction found, $F_{(1,19)} = 148.45$, $MSE = 1.16$, $p. < .001$. The overall pattern from both previous experiments replicated even more

strongly with irrelevant speech, showing the same short word advantage in the immediate recall

phase, reversing to become a long word advantage at recognition.

     5.10.2.1 Immediate Recall. From the immediate recall phase, a significant main effect was

found for word length, $F(1,19)= 41.64$, *MSE*= 0.64, $p. < .001$. Figure 5 shows short words were

recalled more often than long words. A significant main effect was also evident for serial position, F

$(5,95)= 41.91$, *MSE*= 1.25, $p. < .001$. Figure 5 indicates high primacy but no recency in the curves,

with a consistent word length effect observed in the first five serial positions, reducing to no effect

at position 6. The interaction between word length and serial position, however, was not found to

be significant, $F(5,95)= 2.09$, *MSE*= 0.89, $p. = .073$.

Figure 5. Immediate serial recall curves from Experiment 3.
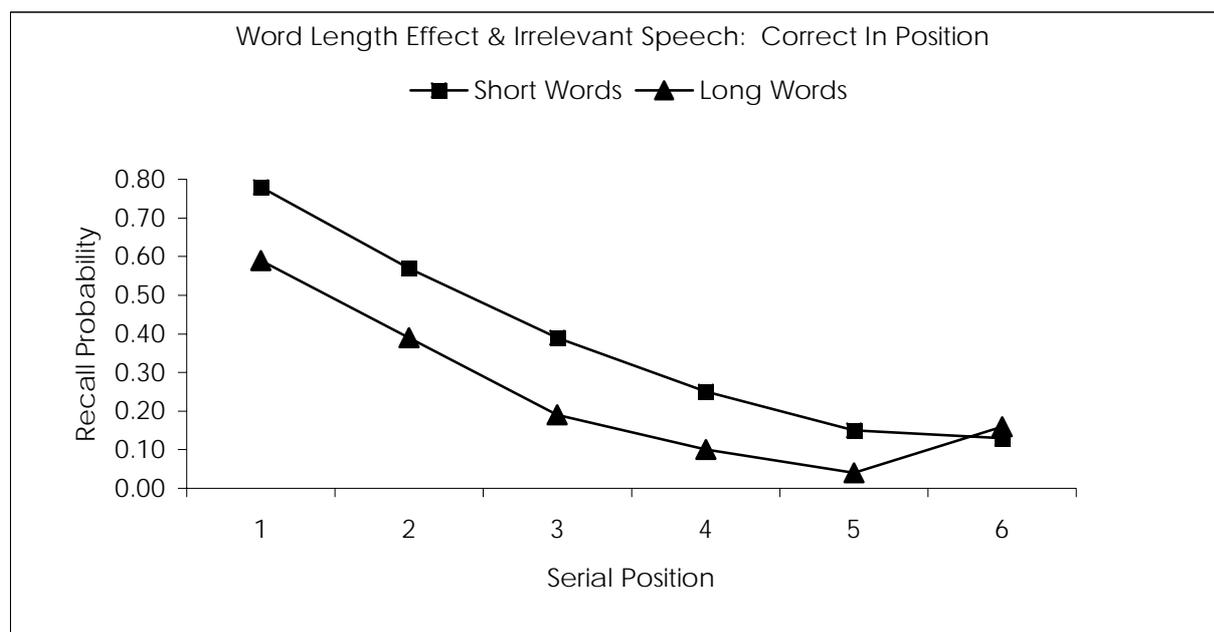
     5.10.2.2 Recognition. Recognition test data were then analysed in the same manner as in

Experiments 1 and 2. There was a significant main effect for word length, $F(1,19)= 76.28$, *MSE* =

2.36, $p. < .001$. Figure 6 shows more long words were recognised than short words across all serial

positions. A significant main effect for serial position, F (5,95)= 12.185, MSE = 2.78, $p.$ < .001, is also

illustrated in Figure 6. Serial position curves for recognition followed the same overall pattern found

in immediate recall. There was no significant interaction between word length and serial position, F

(5,95)= 1.66, MSE = 1.81, $p.$ = .151.

Figure 6. Recognition curves from Experiment 3.
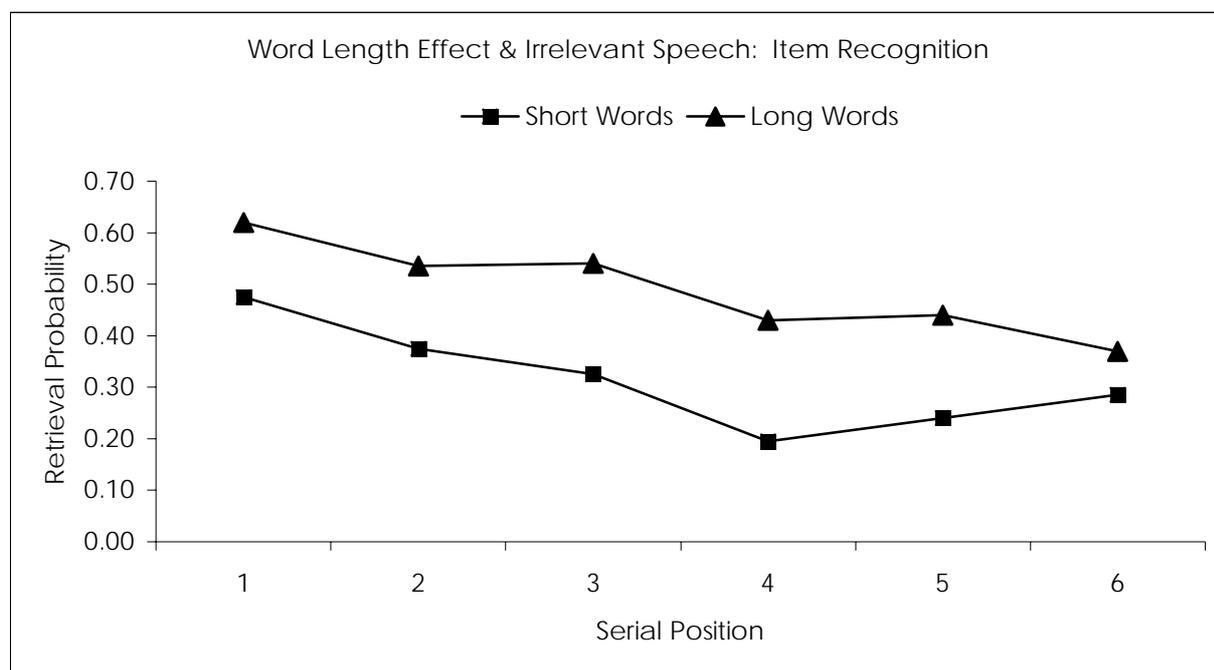
5.11 Experiment 3 Discussion

The individual differences data suggest most of the participants from this experiment were

recalling items using a forward strategy. Again, the majority showed the expected tradeoff effect,

the pattern from Experiments 1 and 2 remaining robust under irrelevant speech conditions. The

ANOVA interaction from the combined data once again supports the processing tradeoff in word

length effects. A short word advantage at the serial recall phase reversed to become a long word

advantage at recognition.

While the presence of irrelevant speech did not change the overall pattern of results for the

immediate serial recall task, the level of performance was generally higher in terms of recall

probabilities than in delayed recall (Experiment 2) but lower than in Experiment 1. That is, irrelevant speech had its expected detrimental effect upon overall levels of performance. Comparing the overall treatment means with those from Tables 1 and 2, it is again noted that the word length effects remain comparable in magnitude for both immediate recall and recognition tasks.

The serial recall curves in Figure 5, as well as the recognition curves at Figure 6, reflect the forward recall strategy adopted by most of the participants, with the word length effects in both tasks attenuating at serial position 6. Recognition scores were again considered reliable due to the relatively low incidence of false alarms.

The proportions of transpositions and intrusions were again equivalent between the word length conditions, repeating the pattern found in Experiments 1 and 2.

5.12 Chapter Summary

Short words were better recalled than long words in the serial recall component of the three experiments. That is, the standard word length effect emerged in immediate serial recall, delayed serial recall and under irrelevant speech conditions. Likewise, the serial position curves derived from these tasks correspond to those normally found in serial recall tasks with visual presentation, with pronounced primacy and very little recency. These results clearly replicate previous findings and allow for some confidence in the more exploratory aspects of the experiments. These more exploratory aspects deal with the reverse word length effect on the item recognition task. Across all three experiments, long words were better recognised than short words. As such the data strongly conform to the pattern predicted by the item-order tradeoff perspective.

5.12.1 Effect Size Estimates. The robust pattern of effects throughout this group of experiments is reflected in the effect size estimates. Eta-squared effect sizes from the two components of each word length experiment are shown at Table 4. One notable feature of these effects is the relationship between the effect sizes for the different tasks. Small effect sizes in serial recall

correspond with relatively small effect sizes in recognition. Likewise large effects in serial recall produce large effects in item recognition. This pattern is consistent with the assumptions of the tradeoff approach. In short, the complete pattern of results is consistent with the item/order tradeoff explanation of the word length effect.

Table 4. Summary of effect sizes from word length experiments 1 to 3.

Eta-squared Effect sizes	Immediate Serial Recall – Short Word Advantage	Recognition – Long Word Advantage
Immediate Recall Exp. 1	.473	.655
Delayed Exp. 2	.254	.518
Irrelevant Speech Exp.3	.687	.801

In terms of the individual differences data, the estimates of forward or primacy and backward or recency strategies used by participants tend to shed no further light upon the pattern of tradeoff effects, and there appears to be no value in examining strategy estimates for the remaining experiments. A breakdown of the patterns and directions of effects within the groups will still be shown for completeness – as it is evident that not all participants will produce the same pattern of tradeoff effects under the conditions tested so far.

The results of the current experiment suggest that the word length effect may well be considered another instance of the item-order tradeoff framework. This case would be more compelling if other well established accounts of the tradeoff perspective would provide data showing the same pattern of dissociations when subjected to the same experimental conditions as used in the current experiments. In the following chapter the experiments described and reported above are replicated using the generation effect.

Chapter 6 Contents

CHAPTER 6

Generation Effect Experiments

6.1 Rationale and predictions

The item-order tradeoff account had its genesis in the dissociation of the generation effect across serial recall and item recognition tasks (Nairne et al., 1991). Chapter 3 outlined the relevant research that has subsequently confirmed the tradeoff perspective. It would thus seem unnecessary to examine this phenomenon again. It is done so, however, because in the previous literature the serial recall component has typically not been examined under standard serial recall conditions. That is, immediate serial recall of short lists presented at rapid rates has not been employed, and there has been no reason to examine the effects under irrelevant speech. The closest methodology has employed longer lists, presented at two to three second rates, and memory has been tested via delayed order reconstruction.

Thus, while word length effects under standard short-term serial recall conditions were well established and recognition performance was unknown, the reverse is true here. The effects of generation on item recognition are well documented, but its effects on short-term serial recall have not so far been studied. However, assuming that short-term serial recall conditions do not radically differ from those previously employed, the expectations are clear. The generated words should not be as well recalled as control words in the serial recall task, but should be better recognised than the control items.

6.2 General Method

As with the word length experiments in the previous chapter, the following generation effect experiments contained an initial serial recall phase, followed by a later recognition phase. Structurally, there were no differences between the following experiments and those involving word length, however different word pools were assembled for generation. The generation manipulation involved creating a fragment from a word by removing a letter and leaving a blank

space (Nairne et al., 1991). Participants then had to re-create the word at study from the fragment by including the missing letter when reading it aloud.

To relate the following series of experiments to the previous word length effect section, the less difficult control lists previously comprised short words, the more difficult lists contained long words. In the following generation effect series the less difficult control lists contained items which were simply read aloud, and the more difficult lists required generation of the items from fragments.

6.2.1 Participants. 20 introductory psychology students from the University of Southern Queensland volunteered to participate in each 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. For the word length experiments as well as those following, a different 20 people were tested in each.

6.2.2 Materials. For experiments involving the generation effect, a single word pool was created from the MRC Psycholinguistic Database (Quinlan, 1992) comprising 240 words. All words contained six phonemes, and were selected such that if a single letter was replaced with a blank space, only that one letter could be inserted to produce a legitimate English word (i.e. there was only one possible legal solution to the word fragment).

Under the same design as the word length experiments, each participant received a unique set of lists. For each participant, the 240 words were randomly assigned to either the generation condition (made into fragments) or the read control condition (left as entire words). Two pools of 120 words each resulted, from which 90 words were randomly selected, and then randomly assigned to the six serial positions on each list in the immediate serial recall component of the experiments. As with the word length lists, this produced 15 six-word lists of word fragments, and 15 six-word lists of read (control) words. The 15 lists in each condition were randomly divided into 10 non-recall (asterisk) lists, and 5 recall (question mark) lists. The order of the resulting 30 lists was then randomised. As before, computer files were duplicated to provide a hard copy for the experimenter to record participant responses.

The recognition component of the task involved the presentation of 180 intact words. The recognition test was comprised of the 10 asterisk lists (60 read and 60 generated words) from the asterisk trials, as well as filler words which comprised the unused 30 read and 30 generated words from the initial pools. The fragments that were used in the study phase were replaced by the words from which they were derived. The test sheet was created by randomly ordering the list of 180 words and arranging them in six columns of 30 words on a single A4 page.

6.2.3 Procedure Participants were tested individually, in sessions of approximately 35 minutes' duration. The only difference in procedure from the experiments involving word length was the requirement to read aloud the words from each list as they appeared on the screen, thus providing the experimenter with a measure of generation accuracy. In the control or read condition this was straightforward, and in the generation condition the instructions stressed that each fragment had to be completed, and then recalled in serial order. Recall procedures were the same as in the previous experiments, with participants instructed to recall list words in order if they saw the series of question marks after the list, or to wait for the next list if there were asterisks. A practice trial of six lists was then conducted, and when the experimenter was satisfied that the participant understood the instructions, including the requirement to generate words from fragments, the test commenced.

Computerised presentation of lists in the following experiments was identical to that used in the previous word length experiments. During presentation, as the generated words were read aloud, the experimenter recorded any errors or omissions made in generating words from the fragments (read errors). At recall, the experimenter recorded responses from participants on a hard copy as either correct in position, omitted, transposed or extralist intrusions. Totals for generated and control lists (preserving serial position) provided the basis for the measures of immediate recall, to be later modified using scoring conditional upon correct generation at study.

The recognition phase, commencing after a three minute delay, again differed little from the previous word length experiments, with participants simply being asked to tick or circle any

words they remembered seeing in the previous phase of the experiment. They were given as much time as they required to complete the task, then were debriefed about the purpose of the experiment.

6.2.4 Conditional Scoring. Numbers of correctly-recognised words from the generation and read lists were totalled to provide measures of recognition. Filler words incorrectly identified as being from the immediate recall phase (false alarms) were recorded separately. Alternate recognition scores were calculated where performance was conditionalised upon correct generation at study.

The use of conditional scoring represented a minor departure from the methodology used in the previous chapter, and involved expressing individual total scores as proportions rather than as whole number counts. In both cases, however, proportions were calculated for the group data to be used in the graphs, and the only effective difference to the statistical analyses was an observable reduction of the size of mean squares.

6.3 Experiment 4. Generation & Immediate Serial Recall

6.3.1 Participants. 20 introductory psychology students from the University of Southern Queensland volunteered to participate in the experiment.

6.3.2 Materials and Procedure. For immediate serial recall, no changes were made to the General Method detailed above.

6.4 Experiment 4 Results

6.4.1 Individual Differences. Using the raw data set, estimates were made of the pattern of tradeoff effects within the group using the same methodology as in the previous chapter. Tradeoff effects, comprising an advantage for read items at immediate serial recall, reversing to become an advantage for generated items at recognition, were produced by six of the participants, three of these were relatively weak at immediate recall. One participant showed the exact reverse pattern.

A further five participants demonstrated a preference for generated items in both tasks, one of these in a relatively weak manner, and one participant produced a strong read item preference in the immediate task, but no effect at recognition. The other seven participants did not show any discernible effects over both tasks.

6.4.2 Group Data. Scoring of the group data was carried out in the same manner for generation as it was for the Chapter 5 word length experiments, with the only change being the use of conditional scoring. As participants read aloud the list items at study, generation failures were recorded, so that the proportion correct in each list could be expressed using the number of correctly generated items as the denominator. Likewise, similar proportions were calculated for recognition based on recorded generation failures in the non-recall lists.

The Figures which follow illustrate the difference between face value and conditional scoring, which was deemed necessary as generation failures averaged 11% for the group. There were no errors made by any participant in the control lists in which items were simply read aloud. A summary of recall data for the whole group is shown at Table 5.

Table 5. Summary data from Experiment 4.

Recall Probability	Immediate Serial Recall			Recognition	
	Correct	Transposed	Extralist	Correct	F Alarms
Read	0.32	0.15	0.00	0.47	0.06
Generated	0.29	0.12	0.00	0.51	
Conditional	0.33			0.56	

Note. Conditional data were derived from the Generate scores as described above. Only a single total was available for false alarms.

Initial analysis of words correctly recalled from lists in Experiment 4 was conducted on conditional data using a 2 x 2 x 6 repeated measures ANOVA, with the only F-ratio of theoretical interest being the task by effect interaction, which would demonstrate the hypothesised

processing tradeoff between immediate serial recall and recognition tasks. The interaction was significant, $F(1,19) = 9.195$, $MSE = .024$, $p. = .007$.

6.4.2.1 Immediate Serial Recall. 2 x 6 ANOVAs were conducted on data from each task individually, to produce the following effects with generation effect and serial position as within-subjects variables. For the immediate test, no significant main effect was found for generation, $F(1,19) < 1$. A significant main effect was found for serial position, $F(5,95) = 5.04$, $MSE = .083$, $p. <$ .001. The serial position curves in this experiment appeared to be more bowed that in the experiments in the previous chapter. Here, primacy and recency effects are roughly equal. This may reflect an enhanced use of a recency based strategy for recall. The interaction was not significant, $F(5,95) < 1$. Figure 7 shows a slight recall advantage for read words in the early list positions, with a very weak advantage for generated words at recency. For comparative purposes, both unconditional and conditional curves are shown for generation.

Figure 7. Immediate serial recall curves from Experiment 4.
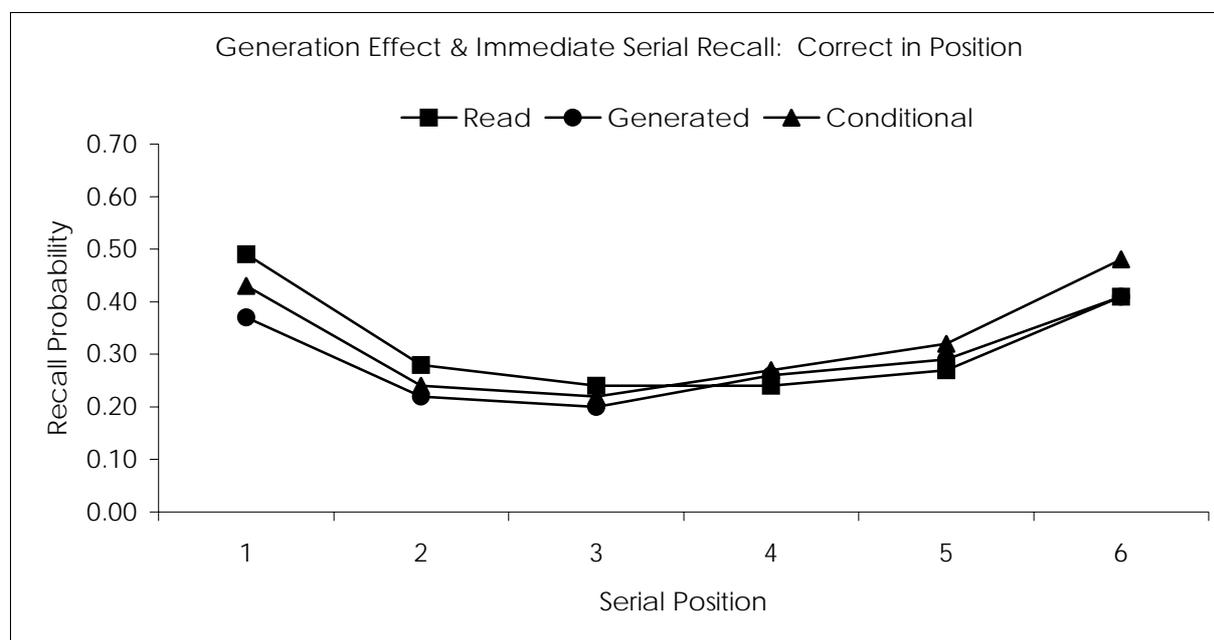
6.4.2.2 Recognition. Conditional data from the recognition test were analysed in the same manner. There was a significant main effect generation, $F(1,19) = 17.43$, $MSE = .028$, $p. = .001$.

More generated items were recognised than control items. A significant main effect for serial position, F (5,95) = 2.94, MSE = .027, $p.$ = .016, is also shown in Figure 8, with generated words recognised more often in later list positions. The interaction, however, was not significant, F (5,95) < 1.

Figure 8. Recognition curves from Experiment 4.
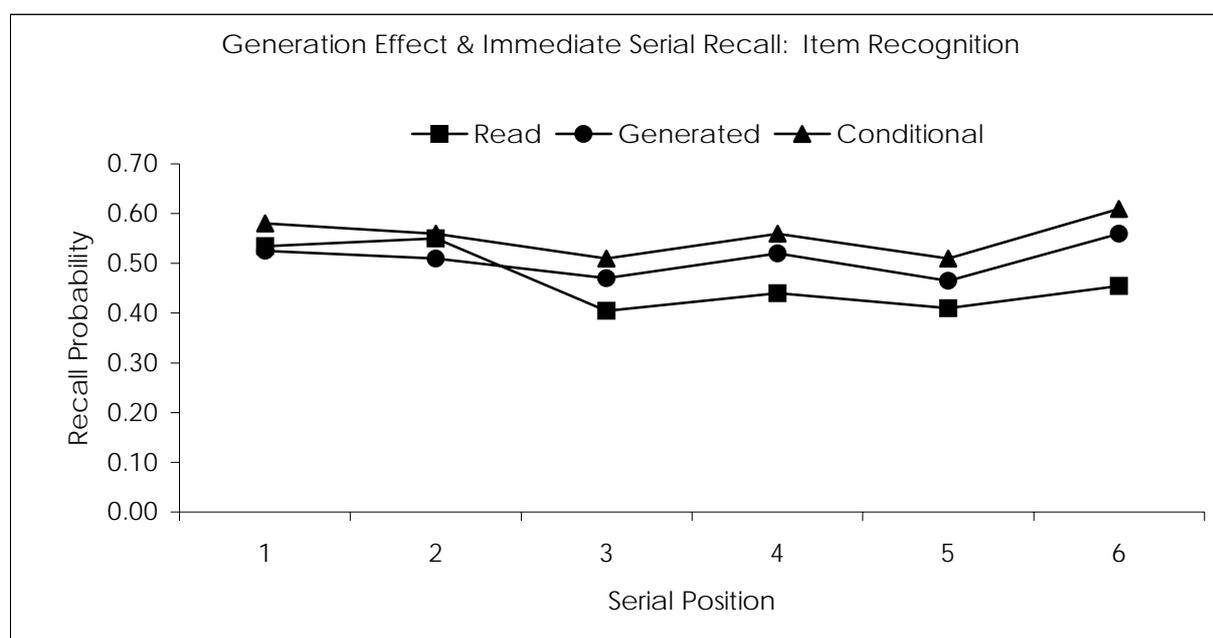
6.5 Experiment 4 Discussion

While a significant interaction between the tasks was evident from the overall analysis in this experiment, it did not fully support the hypothesised tradeoff effect. Although there was a significant recognition advantage for generated items, groups were approximately equivalent at immediate serial recall. The generated item advantage at recognition is attenuated when compared with the previous results, and the overall picture suggests there may have been issues with the sensitivity of the experiment.

Continuing from the word length experiments in the previous chapter, values were again equivalent for transpositions and extralist intrusions, supporting the notion that the two levels of generation were behaving in a similar manner. Under immediate serial recall conditions, a higher

level of performance would generally be expected, but if a lack of sensitivity was responsible for the effects in these data, it would indeed be reflected in depressed scores across the whole experiment. It remains the case that six of the participants, or 30%, still showed the hypothesised tradeoff pattern.

6.6 Experiment 5. Generation Effect & Delayed Recall

6.6.1 Participants. 20 introductory psychology students from the University of Southern Queensland volunteered to participate in the experiment.

6.6.2 Materials and Procedure. In the same manner as the delayed recall word length experiment in Chapter 5, the only change made to the General Method detailed above was the insertion of a four second filled delay, using the same mechanism as Experiment 2, between the end of the list and the recall cue. Participants read aloud a series of four, four-digit random numbers which appeared on the computer screen before either the asterisks or the question marks were displayed. Instructions to participants and computer files were altered to reflect the delay modification.

6.7 Experiment 5 Results

6.7.1 Individual Differences. Estimates were made as before of the pattern of tradeoff effects within the group. Tradeoff effects, comprising an advantage for read items at the delayed recall phase, reversing to become an advantage for generated items at recognition, were observed in three of the participants, while one participant showed the reverse pattern, albeit weakly.

A single participant demonstrated a preference for generated items in both tasks, and three produced a strong read item preference in both. Five participants did not show a preference in either task. Of a further six who failed to demonstrate a generation effect at delayed recall, four showed a generated item advantage at recognition, two a read item advantage. The remaining participants produced an advantage at delayed recall for read items, but no effect at recognition.

Conditional scoring of the group data was carried out as it was in the previous experiment, with generation failures at study again averaging 11%. There were no errors made by any participant in the read item lists. A summary of recall data for the group is shown at Table 6.

Table 6. Summary data from Experiment 5.

Recall	Delayed Serial Recall			Recognition	
Probability	Correct	Transposed	Extralist	Correct	F Alarms
Read	0.28	0.14	0.01	0.48	0.06
Generated	0.20	0.17	0.01	0.49	
Conditional	0.23			0.54	

6.7.2 Group Data. Initial analysis of conditional data from Experiment 5 was conducted using a 2 x 2 x 6 repeated measures ANOVA, once again with the only F-ratio of theoretical interest being the task by generation effect interaction. The interaction was significant, $F(1,19) = 9.07$, $MSE = .037$, $p. = .007$.

6.7.2.1 Delayed Recall. Individual 2 x 6 ANOVAs were then carried out for each task, and in the serial recall phase, the main effect for generation was not found to be significant, $F(1,19) = 2.54$, $MSE = .063$, $p. = .127$. A significant main effect was found for serial position, $F(5,95) = 9.35$, $MSE = .055$, $p. < .001$. Again, bow shaped serial position curves are apparent in the data and match those found in the previous experiment. The interaction was significant, $F(5,95) = 2.59$, $MSE = .031$, $p. = .031$. Figure 9 shows a recall advantage for read words in the early list positions, with a very weak advantage for generated words at the final list position. Both unconditional and conditional curves are shown.

Figure 9. Delayed recall curves from Experiment 5.
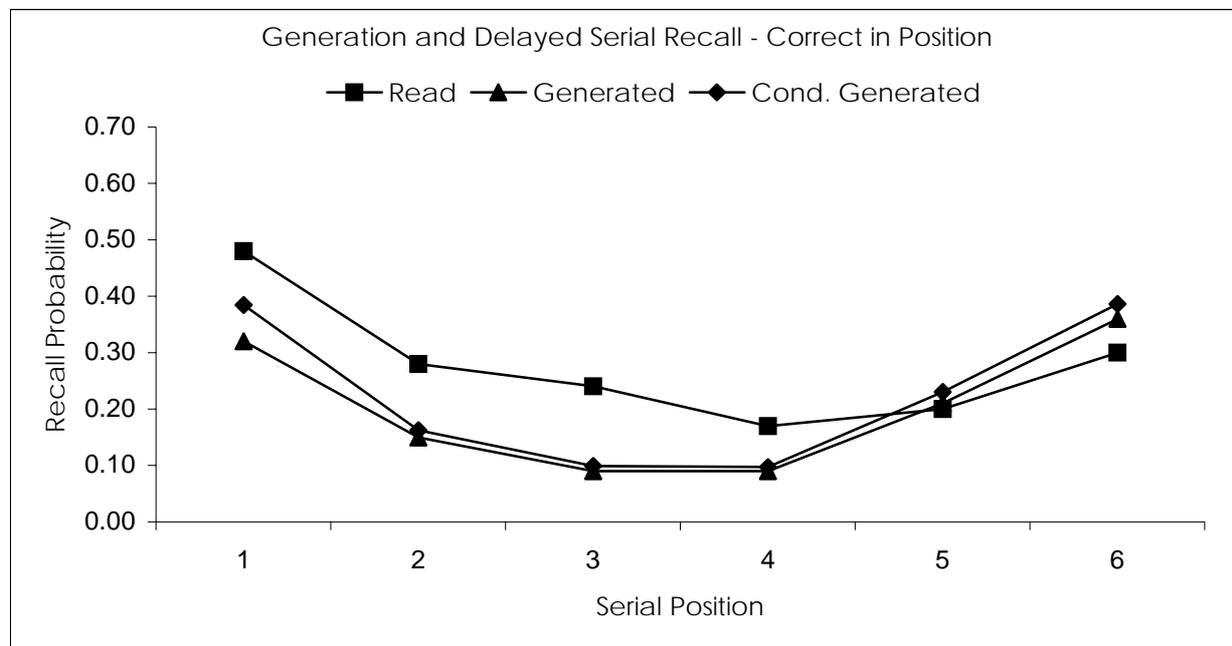
6.7.2.2 Recognition. Data from the recognition test, analysed in the same manner, provided a main effect for generation which approached significance, F (1,19) = 3.32, *MSE* = .053, *p.* = .084. A significant main effect for serial position, F (5,95) = 3.85, *MSE* = .023, *p.* = .003, is also shown in Figure 10, and is associated with greater recognition of generated words at list position 6. The interaction, however, was not significant, F (5,95) = 1.73, *MSE* = .022, *p.* = .135.

Figure 10. Recognition curves from Experiment 5.
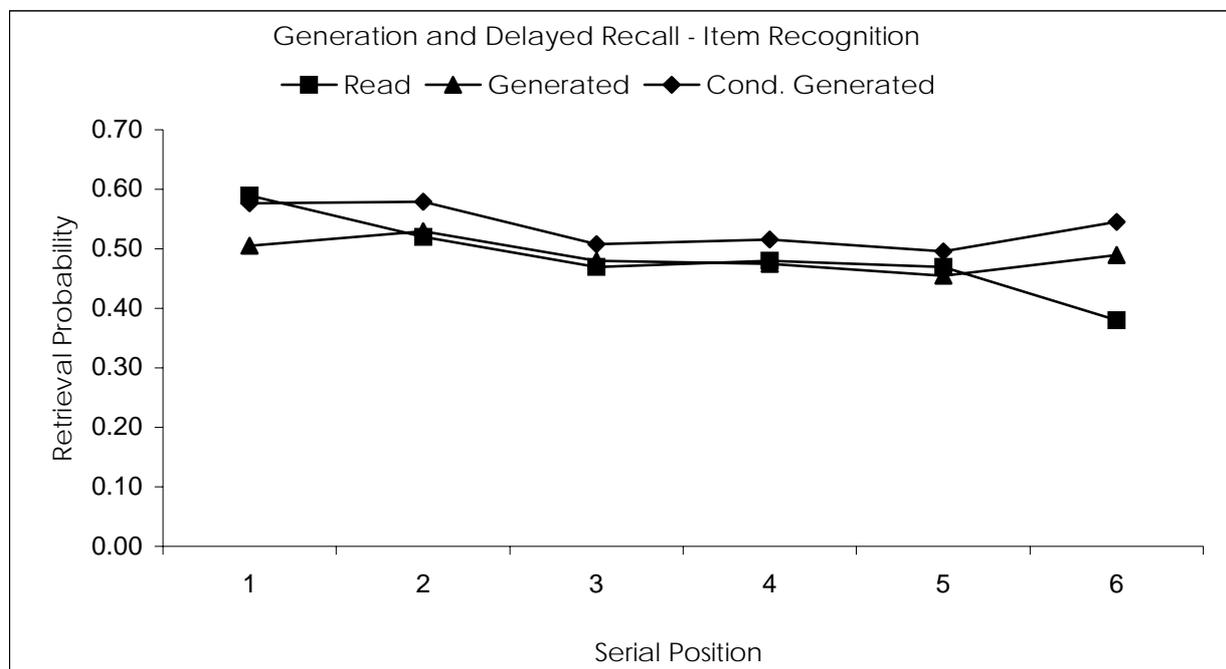
6.8 Experiment 5 Discussion

Again, using conditional scoring, the overall interaction reveals a difference in effects between tasks, of which the direction is as hypothesised, but not statistically significant in either task. Under conditions of delayed serial recall, performance in both tasks is attenuated, and the individual differences data reveal a wide variation in individual patterns of recall performance.

The values in Table 5 again show remarkable consistency in the levels of transpositions and intrusions between conditions, it is once again evident that these measures are not contributing to any systematic variation between groups, and the recognition scores and false alarm levels are consistent with the previous experiments. Indeed, as the experiments appear to increase in relative complexity, the number of false alarms at recognition diminishes rather than increases. Outliers did not affect the recognition false alarm score as much as in previous experiments, and if the two participants with over 10 errors are removed from the data, the proportion drops to 0.04.

6.9 Experiment 6. Generation & Irrelevant Speech

6.9.1 Participants. 20 introductory psychology students from the University of Southern Queensland volunteered to participate in the experiment.

6.9.2 Materials and Procedure. The only change made to the General Method detailed above was the addition of recorded news broadcasts in Russian, which were played during study at a clearly audible volume from speakers attached to the computer. Following list presentation and simultaneous irrelevant sounds, either the asterisks or the question marks were displayed as before. In fact, presentation was identical in all ways but the materials to that used in Experiment 3. Instructions to participants and computer files were adjusted to reflect the change.

6.10 Experiment 6 Results

6.10.1 Individual Differences. Estimates of the pattern of tradeoff effects within the group were again made from the raw data. Tradeoff effects were observed in just two of the participants, while one participant showed the reverse pattern.

Five participants demonstrated a preference for generated items in both tasks, and two produced a strong read item preference in both. Four participants did not show a discernible effect in either task. Of a further five who failed to demonstrate a generation effect at the immediate recall phase, four showed a generated item advantage at recognition, and one a read item advantage. The remaining participant produced a read word advantage at immediate recall only.

Conditional scoring of the group data was again carried out as it was in the previous experiment, with generation failures at study averaging 10.5%. There were no errors made by any participant in the read item lists. A summary of recall data for the group is shown at Table 7.

Table 7. Summary data from Experiment 6.

Recall	Immediate Serial Recall			Recognition	
Probability	Correct	Transposed	Extralist	Correct	F Alarms
Read	0.21	0.29	0.06	0.41	0.08
Generated	0.20	0.25	0.04	0.48	
Conditional	0.22			0.52	

6.10.2 Group Data. Initial analysis of conditional data from Experiment 5 again began with the combined data from both tasks, with the only F-ratio of interest being the task by effect interaction. The interaction was significant, $F(1,19) = 7.40$, $MSE = .043$, $p = .014$, but once again it reflected equivalent group performance at immediate recall, changing to a generated item advantage at the recognition phase.

6.10.2.1 Immediate Serial Recall. Individual 2 x 6 ANOVAs were conducted for each task, and in the immediate serial recall phase, the main effect for generation was not found to be significant, $F(1,19) < 1$. A significant main effect was found for serial position, $F(5,95) = 8.63$, $MSE = .070$, $p. < .001$, and again bow shaped serial recall curves are evident. The interaction was not significant, $F(5,95) < 1$. Figure 11 shows little if any separation across all list positions, reflecting the absence of a main effect for generation. Both unconditional and conditional curves are shown.

Figure 11. Immediate serial recall curves from Experiment 6.
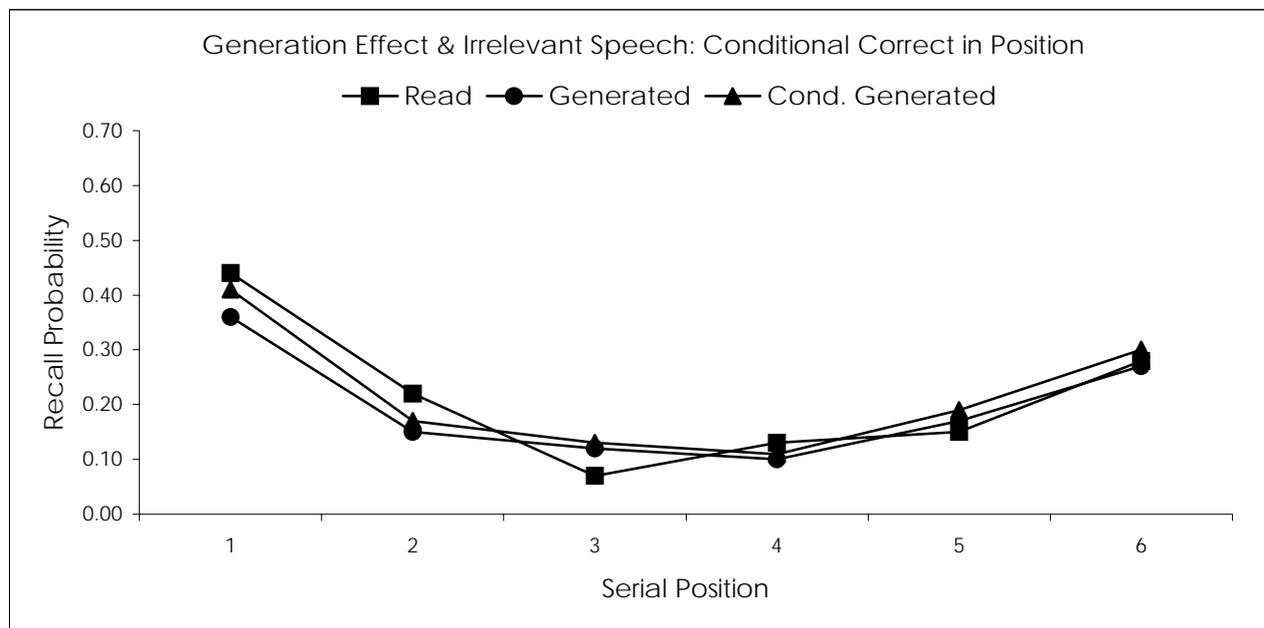
6.10.2.2 Recognition. Data from the recognition test, analysed in the same manner, revealed a significant main effect for generation, $F(1,19) = 13.88$, $MSE = .049$, $p. = .001$. A significant main effect was not evident for serial position, $F(5,95) = 1.65$, $MSE = .023$, $p. = .155$, and the interaction was not significant, $F(5,95) = 1.78$, $MSE = .030$, $p. = .125$. The recognition curves shown in Figure 12 demonstrate greater recognition of generated words at all list positions except position 1.

Figure 12. Recognition curves from Experiment 6.
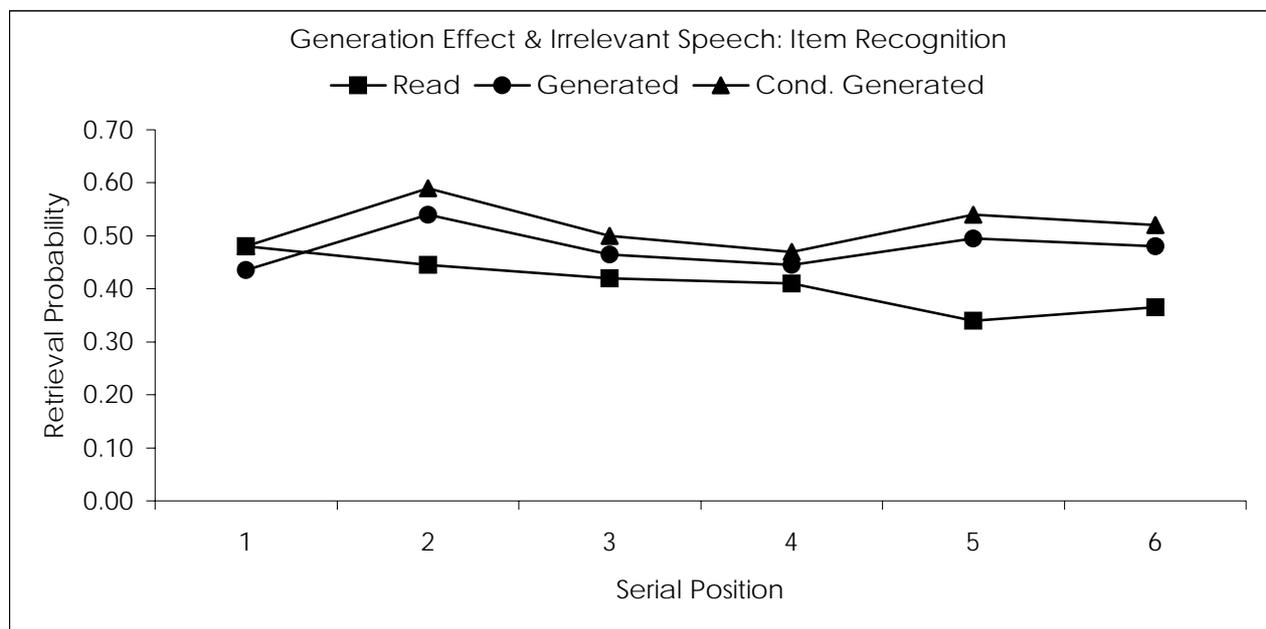
6.11 Experiments 4-6 Discussion

Throughout the generation effect experiments detailed above, trends in the data reveal an advantage for generated items at recognition, which represents only half of the hypothesised tradeoff effect. The expected advantage at immediate recall for simply read items simply did not emerge with the above manipulation of generation. Effect sizes (Eta-squared values from SPSS output) for the three generation effect experiments in this chapter are shown at Table 8.

Table 8. Summary of effect sizes from generation experiments 4 to 6.

Eta-squared Effect sizes	Immediate Serial Recall – Read Word Advantage	Recognition – Generated Word Advantage
No Delay Exp. 4	.001	.478
Delayed Exp. 5	.118	.149
Irrelevant Speech Exp. 6	.001	.422

The fact that the read advantage did not emerge in the serial recall component of the current experiments was unexpected given that such advantages have been observed in other serial recall tasks, albeit order reconstruction tasks that have been tested over a substantial delay. The conclusion here is that either there are fundamentally different processes in the short-term realm - or that the current task parameters render the task insensitive to the manipulation. The effect sizes for the various conditions across the experiment are presented in Table 8. Compared with the word length experiments all effect sizes are relatively weak.

6.12 Chapter Summary

An initial sensitivity argument is that there was insufficient statistical power in each experiment, although the nature of the design used and the numbers of participants in each of the above three experiments would suggest an effect would be found if it was there. To confirm the pattern of results in Table 8, the data from the three experiments were combined and re-analysed using a repeated-measures t-test for both the serial recall phase and the recognition phase. With the increased power from this manipulation, results did not change. For serial recall, $t(59) = 0.86$, $p.$ = .39 and for recognition, $t(59) = 5.38$, $p. < .001$. Thus, the traditional finding that generation produces an advantage in item recognition has been replicated. The expected generation decrement in serial recall has failed to emerge.

The consistency of design between these generation experiments and those involving word length in the previous chapter allows, however, for some inferences to be drawn from the overall pattern. It seemed to be the case from the word length experiments and their associated effect sizes summarised in Table 4 that a strong reverse effect at recognition is related to a weaker effect at immediate recall. Given that the trend is for a weak to moderate reverse effect at recognition in these experiments, it seems likely that an effect at immediate recall may actually exist, but the effect is too weak to be significant.

The differences in serial position curves between the word length and generation experiments suggest that there are other differences between the experiments. That is, in the word

length effect experiments, the serial position curves in the serial recall data were those that are typically found in the immediate serial recall literature. In the current generation experiments, the same serial position curves are much more bow shaped. The likely explanation here is that more participants are relying upon a recency or backward recall strategy to do the task. As such, serial order encoding may not be as prevalent as in the first experiment and as such one would expect that the tradeoff effect may be reduced in strength.

Finally, it is entirely possible that the manipulation of generation in these experiments, specifically in the materials used, has led to a reduction in sensitivity. Due to the requirement of a word fragment having only one legal solution, the words by nature must be moderately long. As long words were considered a 'difficulty' manipulation in the previous set of experiments, the requirement to generate from a long fragment can be seen as compounding the difficulty of all the generated items in the experiments.

The compound difficulty exists because the items are all long words to begin with, then there is a further processing strain with the added requirement to generate. The additional processing required for the less difficult (simply read) items is further intensified when the requirement to generate is added, and so the ability of such a structure to show an effect is compromised. Sensitivity is therefore an issue which will be raised and dealt with in Chapter 8, where a stronger manipulation of generation (using shorter items) is expected to show patterns of effects which are more consistent with those already noted for word length.

Before any additional manipulations of the generation effect can be considered, however, the perceptual interference effect is examined under the same conditions as word length and generation in the next chapter.

Chapter 7 Contents

CHAPTER 7

Perceptual Interference Effect Experiments

7.1 Rationale and Predictions

The item-order tradeoff was readily observed in the experiments dealing with word length, however the effects were not as strong with the experiments involving a generate/read manipulation. The perceptual interference effect is another phenomenon that has been explained in terms of an item-order processing tradeoff, as discussed in the third chapter. However, like the generation effect, perceptual interference has not been studied under what would traditionally be considered short-term memory conditions. The expectation with the following series of experiments is that perceptual interference effects will generalise to the short-term domain and that the processing tradeoff found in previous research will be apparent.

7.2 General Method

In the same manner as all previous experiments from Chapters 5 and 6, the following perceptual interference effect experiments comprised an initial serial recall phase, followed by a later recognition phase. There were no differences in design between the following experiments and those involving word length and generation, however different word pools were used. The perceptual interference manipulation was purely one of presentation – in the control or less difficult condition participants were presented with items at the rate of one per second, which they simply read aloud in the same way as they did in the generation experiments. In the more difficult or 'fast' presentation condition, the items on these lists were only displayed on the screen for 100 ms, after which a 900 ms mask was applied, constructed from characters from the uppercase numbers (@#$!^%&*) on a keyboard. The overall presentation rate (of one item per second) was thus maintained throughout the series of experiments.

7.2.1 Participants. 20 introductory psychology students from the University of Southern Queensland volunteered to participate in each 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. A different 20 people were tested in

each experiment, which did not include participants who had been involved in any previous

experiments from this study.

7.2.2 Materials. For experiments involving the perceptual interference effect, a single word pool

was created from the MRC Psycholinguistic Database (Quinlan, 1992) comprising 240 words. All

words contained five phonemes, and ranged from five to eight letters.

In accordance with the overall design used in the previous experiments, each participant

received a unique set of lists. For each participant, the 240 words were randomly assigned to

either the control condition or the fast presentation condition. Two pools of 120 words each

resulted, from which 90 words were randomly selected, and then randomly assigned to the six serial

positions on each list in the immediate serial recall component of the experiments. This produced

15 six-word lists of control items and 15 six-word lists for fast presentation. The 15 lists in each

condition were randomly divided into 10 non-recall (asterisk) lists, and 5 serial recall (question mark)

lists. The order of the resulting 30 lists was then randomised. As before, computer files were

duplicated to provide a hard copy for the experimenter to record participant responses.

The recognition component of the perceptual interference experiments comprised the 10

non-recall lists (60 long and 60 short words) from the asterisk trials, as well as filler words(the unused

60 words from the initial pools). Test sheets were created by randomly-ordering the list of 180 words

and arranging them in six columns of 30 words on a single A4 page. An additional scoring sheet

allowed the experimenter to preserve the original serial positions of recognised items.

7.2.3 Procedure. Participants were tested individually, in sessions of approximately 35 minutes'

duration. The requirement to read aloud the words from each list as they appeared on the screen

was retained from the generation effect experiments, providing a measure of reading accuracy.

Recall procedures were the same as in the previous experiments, with participants instructed to

recall list words in order if they saw the series of question marks after the list, or to not respond but

wait for the next list if there were asterisks. A practice trial of six lists was then conducted, and when the experimenter was satisfied that the participant understood the instructions and could perform the tasks, the test commenced.

Computerised presentation of lists in the following experiments was identical to that used in the previous word length experiments. During presentation, as the words from each list were read aloud, the experimenter recorded any errors or omissions made in reading the control and fast presentation items. At recall, the experimenter recorded responses from participants on a hard copy as either correct in position, omitted, transposed or extralist intrusions. Totals for fast presentation and control lists (preserving serial position) provided the basis for the measures of immediate recall, to be later modified using scoring conditional upon correct item production.

The recognition phase, commencing after a three minute delay during which administrative details were completed, did not differ from the previous experiments, with participants simply being asked to tick or circle any words they remembered seeing in the previous phase of the experiment. They were given as much time as they required to complete the task, then were debriefed about the nature of the study.

Numbers of correctly-recognised words from the control and fast presentation lists were totalled to provide measures of recognition. Filler words incorrectly identified as being from the immediate recall phase (false alarms) were recorded separately. Recognition scores were again conditionalised upon correct item production at study. The experiments in this chapter follow the same pattern as those for word length and generation, commencing with immediate serial recall, then delayed recall, then irrelevant speech.

7.3 Experiment 7. Perceptual Interference & Immediate Serial Recall .

7.3.1 Participants. 20 introductory psychology students from the University of Southern Queensland volunteered to participate in the experiment.

7.3.2 Materials and Procedure. No changes were made to the General Method described above.

7.4 Experiment 7 Results

7.4.1 Individual Differences. Using the raw data set, estimates were made of the pattern of tradeoff effects within the group using the same methodology as before. Tradeoff effects, comprising an advantage for control items at immediate serial recall, reversing to become an advantage for fast presentation items at recognition, were produced by just one participant, yet five produced the exact reverse pattern (three strongly).

A further seven participants demonstrated a preference for control items in both tasks, one of these in a relatively weak manner, and one participant produced a fast item preference in both tasks. Another two showed no effect in the immediate task, but a control item advantage at recognition. The other four participants did not show any discernible effects in either phase of the experiment. As in the previous experiments, there was a wide variation in the patterns of individual effects within the overall group.

Conditional scoring proportions were again calculated, but unfortunately, in this experiment only, read error data from the asterisk (recognition) lists were not collected by the experimenter due to a misunderstanding of instructions. Conditional data are therefore available only for the immediate serial recall phase, and did not change the pattern or magnitude of scores. A summary of recall data for the whole group is shown at Table 9.

Table 9. Summary data from Experiment 7.

Recall	Immediate Serial Recall			Recognition	
Probability	Correct	Transposed	Extralist	Correct	F Alarms
Regular	0.43	0.10	0.03	0.39	0.08
Fast	0.40	0.09	0.03	0.32	
Conditional	0.41			N/A	

Note. Conditional data were derived from the Fast presentation scores as described above.

7.4.2 Group Data. Initial analysis of words correctly recalled from lists in Experiment 7 was conducted on conditional data using a 2 x 2 x 6 repeated measures ANOVA, again with the only F-ratio of theoretical interest being the task by effect interaction, which would relate to any processing tradeoff between immediate serial recall and recognition tasks. The interaction was not significant, *F* (1,19) = 1.017, *MSE* = .049, *p* = .326.

7.4.2.1 Immediate Serial Recall. Individual analyses were then conducted using 2 x 6 ANOVAs, to produce the following effects with perceptual interference and serial position as within-subjects variables. For the immediate test, no significant main effect was found for perceptual interference, *F* (1,19) < 1. A significant main effect was found for serial position, *F* (5,95) = 25.70, *MSE* = .077, *p.* < .001, and relates to the descending shapes of the immediate recall curves, which show a slight recency effect at position 6 only. The interaction was not significant, *F* (5,95) < 1. Figure 13 shows no clear recall advantage for either condition, with the curves crossing each other several times at different list positions. In this case, conditional values are shown for the Fast condition. The unconditional curve was virtually identical.

Figure 13. Immediate serial recall curves from Experiment 7.
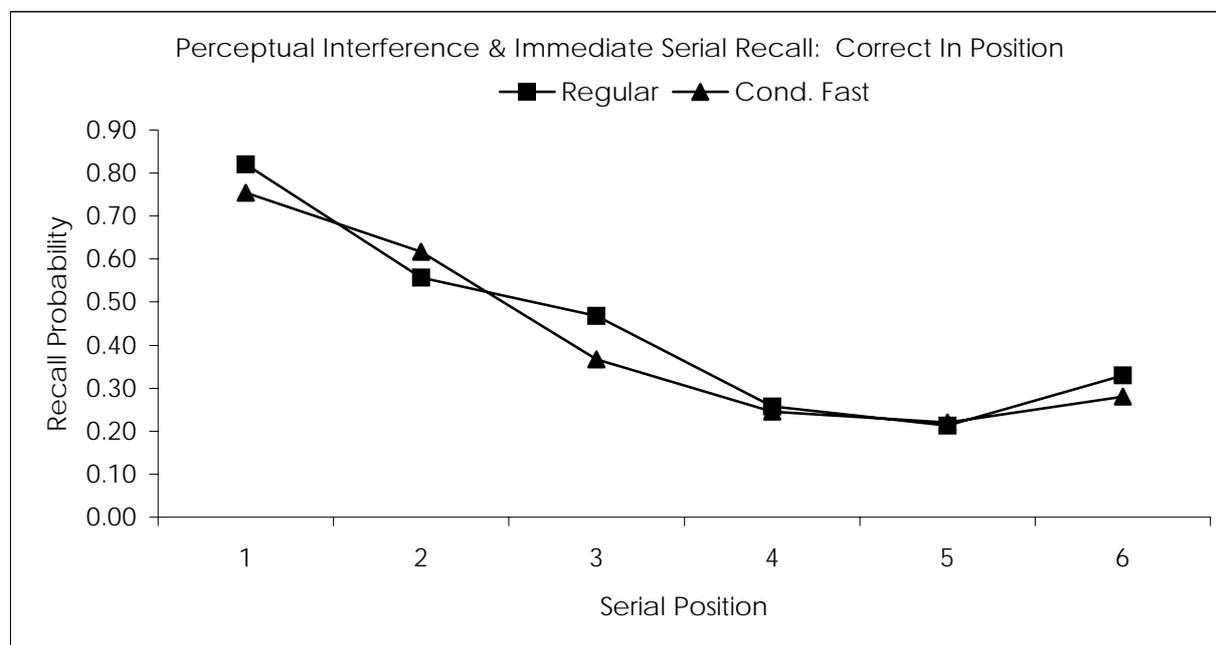
7.4.2.2 *Recognition*. Data from the recognition test were analysed in the same manner.

Perceptual interference produced a significant main effect, $F(1,19) = 16.80$, *MSE* = .016, *p.* = .001,

with control items recognised more often than those which had been presented at a faster rate.

There was no significant main effect for serial position, $F(5,95) < 1$. Figure 14 shows a tendency

toward flat lines, with a slight advantage for control items in early list positions. The interaction was

also nonsignificant, $F(5,95) = 1.51$, *MSE* = .019, *p.* = .193.

Figure 14. Recognition curves from Experiment 7.

7.5 Experiment 7 Discussion

The data from Experiment 7 do not support the idea of a processing tradeoff when the perceptual interference effect is studied under immediate serial recall conditions. The hypothesised control item advantage in the immediate serial recall phase did not emerge, and neither did the expected fast presentation advantage at recognition, instead a small but significant advantage for the control items was evident.

This reversal of the expected effect was supported in the individual differences data, with more participants producing the reverse effect than the hypothesised fast item advantage. In comparison to the generation experiments in the previous chapter, the null effect at the immediate serial recall phase remains, but there is now no evidence to support the idea of a processing tradeoff.

7.6 Experiment 8. Perceptual Interference & Delayed Recall

7.6.1 Participants. 20 introductory psychology students from the University of Southern Queensland volunteered to participate in the experiment.

7.6.2 Materials and Procedure. In the same manner as the delayed recall word length and generation experiments, the only change made to the General Method detailed above was the insertion of a filled delay, using the same mechanism as Experiments 2 and 5. Instructions to participants and computer files were adjusted accordingly.

7.7 Experiment 8 Results

7.7.1 Individual Differences. Estimates were made as before of the pattern of tradeoff effects within the group. Tradeoff effects, comprising an advantage for control items at the delayed recall phase, reversing to become an advantage for fast items at recognition, were observed in two of the participants, while none produced the reverse pattern.

One participant demonstrated a preference for fast items in both tasks, and six produced a moderate to strong control item preference in both. Of the five who failed to show a preference in the delayed recall task, four produced a control item advantage at recognition, and one a fast item advantage. Two participants had no effect at recognition, and of these one recalled more control items in the immediate test, and one recalled more fast items. Four participants did not have a discernible preference in either task.

Conditional scoring of the group data was carried out as it was in the previous experiment, but effectively made no difference to any of the proportions or recall curves shown below. A summary of recall data for the group is shown at Table 10. False alarms for this experiment were not trimmed, as all participants made fewer than ten errors.

Table 10. Summary data from Experiment 8.

Recall	Immediate Serial Recall			Recognition	
Probability	Correct	Transposed	Extralist	Correct	F Alarms
Control	0.21	0.22	0.03	0.43	0.06
Cond. Fast	0.15	0.21	0.05	0.40	

7.7.2 Group Data. Initial analysis of conditional data from Experiment 8 was conducted using a 2 x 2 x 6 repeated measures ANOVA, and the relevant task by effect F-ratio was not found to be significant, $F(1,19) = 1.50$, *MSE* = .022, *p.* = .236.

7.7.2.1 Delayed Recall. Individual 2 x 6 ANOVAs were then carried out for each task, and in the immediate (delayed recall) phase, the main effect for perceptual interference was significant, $F(1,19) = 9.07$, *MSE* = .024, *p.* = .007, with more control items recalled than was the case for those presented at a fast rate. A significant main effect was also found for serial position, $F(5,95) = 9.68$, *MSE* = .053, *p.* < .001, and again classic serial recall curves are evident. The interaction was not significant, $F(5,95) < 1$. Figure 15 shows a recall advantage for control items across all list positions except position 6.

Figure 15. Delayed recall curves from Experiment 8.

Perceptual Interference & Delayed Recall: Correct In Position
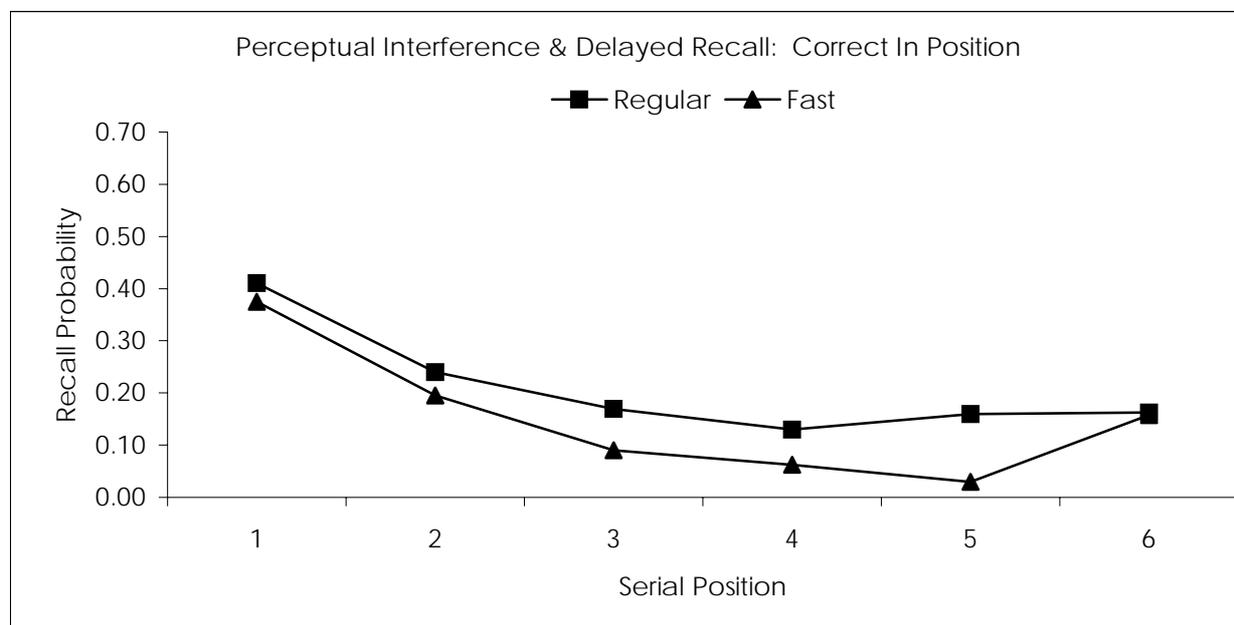
7.7.2.2 *Recognition*. Analysis of data from the recognition test did not show a significant main effect for perceptual interference, $F(1,19) = 1.44$, *MSE* = .031, *p.* = .244. A significant main effect was evident for serial position, $F(5,95) = 3.59$, *MSE* = .029, *p.* = .005, with the recognition curves shown in Figure 16 below, and is associated with a slight recognition advantage for control items at list position 1. The interaction was not significant, $F(5,95) < 1$.

Figure 16. Recognition curves from Experiment 8.

Perceptual Interference & Delayed Recall: Item Recognition
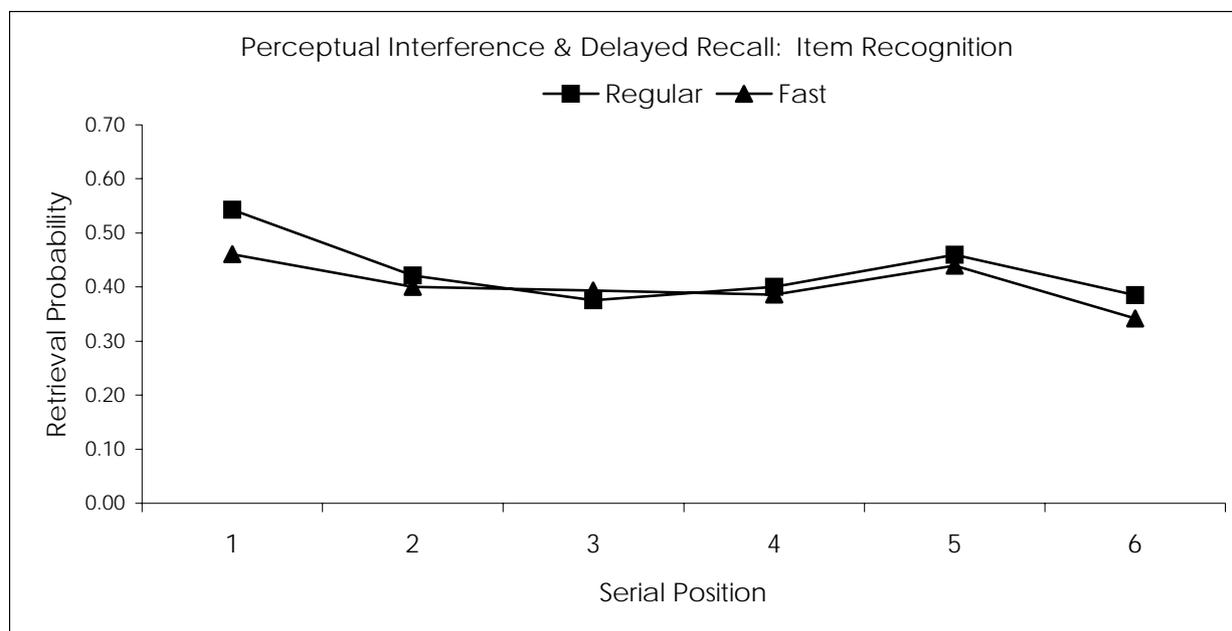
7.8 Experiment 8 Discussion

Under conditions of delayed recall, the pattern of effects from Experiment 7 has reversed, with an overall recall advantage for control items in the immediate recall phase, but no significant effect at recognition. Recall performance in the immediate phase is generally lower than in the previous experiment, being close to floor in later list positions.

The values in Table 5 again show remarkable consistency in the levels of transpositions and intrusions between conditions, it is once again evident that these measures are not contributing to any systematic variation between groups, and the overall level of recognition scores and false alarm levels are consistent with the previous experiments. The recognition false alarm scores were unaffected by outliers, and can be considered reliable in this experiment.

7.9 Experiment 9. Perceptual Interference & Irrelevant Speech

7.9.1 Participants. 20 introductory psychology students from the University of Southern Queensland volunteered to participate in the experiment.

7.9.2 Materials and Procedure. The only change made to the General Method detailed above was the addition of recorded news broadcasts in Russian, which were utilised in the same manner as in Experiments 3 and 6. Instructions to participants and computer files were adjusted to reflect the change.

7.10 Experiment 9 Results

7.10.1 Individual Differences. Estimates of the pattern of tradeoff effects made from the raw data again show a wide variation. A tradeoff effect was observed in just one of the participants, whilst two participants showed the reverse pattern.

Four participants demonstrated a preference for fast items in both tasks, and three produced a control item preference in both. Of a further seven who failed to demonstrate a perceptual interference effect at the immediate recall phase, four showed a fast item advantage at recognition, and three a control item advantage. The remaining three participants did not show a discernible effect in either task.

Conditional scoring of the group data was carried out, and under irrelevant speech conditions made a slight difference to the patterns of scores. Read errors were made in all types of lists, but not by all participants. A summary of group recall data, including conditional scoring where applicable, is shown at Table 11.

Table 11. Summary data from Experiment 9.

Recall Probability	Immediate Serial Recall			Recognition	
	Correct	Transposed	Extralist	Correct	F Alarms
Control	0.29	0.26	0.05	0.40	0.07
Fast	0.27	0.24	0.05	0.33	
Cond. Control	0.29			0.41	
Cond. Fast	0.30			0.36	

7.10.2 Group Data. Initial analysis of conditional data from Experiment 9 used the combined data from both tasks, with the task by effect interaction of interest found not to be significant, $F(1,19) = 2.82$, $MSE = .043$, $p = .110$.

7.10.2.1 Immediate Serial Recall. Individual 2 x 6 ANOVAs were again conducted for each task, and in the immediate serial recall phase, the main effect for perceptual interference was not found to be significant, $F(1,19) < 1$. A significant main effect was found for serial position, $F(5,95) = 17.17$, $MSE = .051$, $p. < .001$, reflected in the typical serial recall curves shown below. The interaction was not significant, $F(5,95) = 1.09$, $MSE = .044$, $p. = .373$. Figure 17 shows little if any separation across all list positions, with a slight advantage for control items at the initial serial position. Both unconditional and conditional curves are shown for fast presentation items.

Figure 17. Immediate serial recall curves from Experiment 9.

Perceptual Interference and Irrelevant Speech - Immediate Recall

.
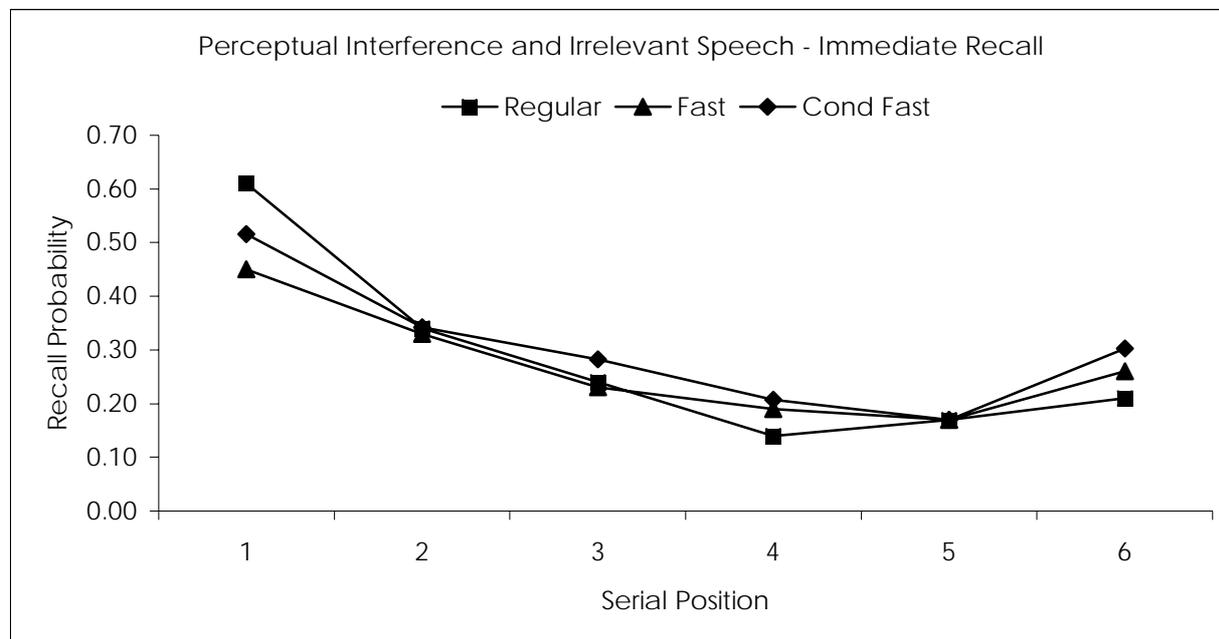
7.10.2.2 Recognition. Data from the recognition task were analysed in the same manner, and a significant main effect was found for perceptual interference, $F (1,19) = 5.80$, *MSE* = .023, *p.* = .026. The main effect for serial position approached significance at the 0.05 level, $F (5,95) = 2.09$, *MSE* = .021, *p.* = .073, and the interaction was not significant, $F (5,95) = 1.34$, *MSE* = .025, *p.* = .253. The recognition curves shown in Figure 18 demonstrate greater recognition of control items in the first three list positions.

Figure 18. Recognition curves from Experiment 9.

Perceptual Interference and Irrelevant Speech - Recognition
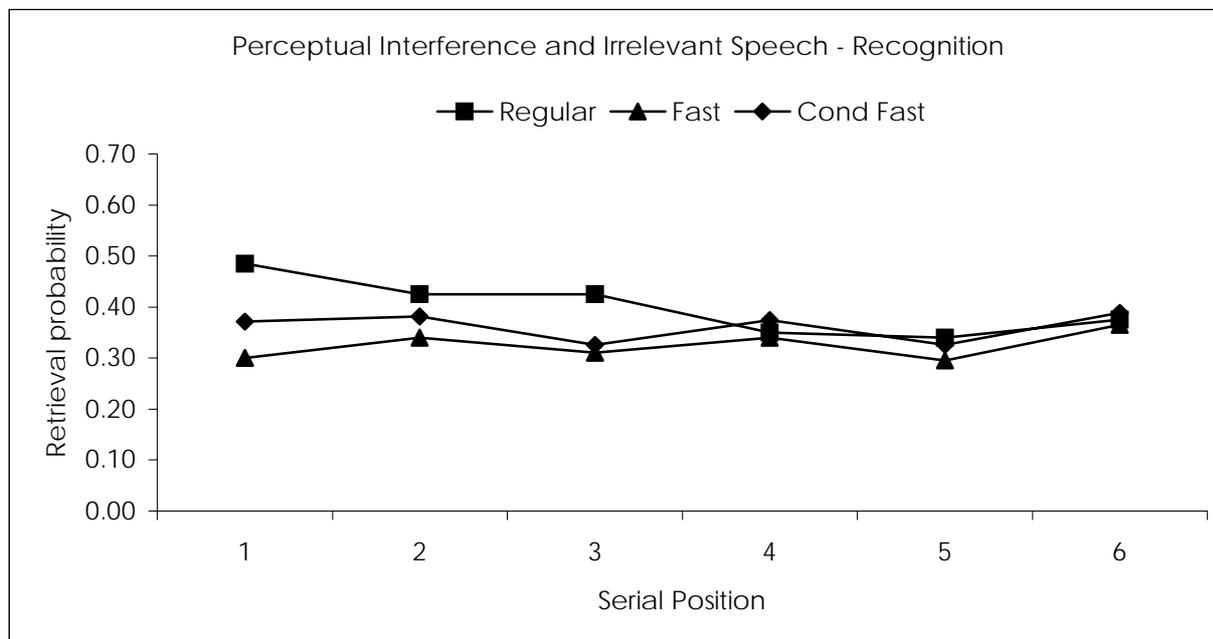
7.11 Experiment 9 Discussion

Under irrelevant speech conditions, the perceptual interference effect was again not in evidence at the immediate recall phase, but an advantage for control items was found at recognition. False alarms in the recognition data were slightly affected by one outlier, and deleting it from the calculation dropped the overall proportion only marginally from 0.07 to 0.06. The expected processing tradeoff was once again absent from the data, but processing differences were apparent in the differing patterns of recall performance between the experimental tasks. Overall levels of recall performance, transpositions and extralist intrusions were again equivalent, and reflected the patterns found in the previous experiments.

7.12 Chapter Summary

When the perceptual interference effect is studied within the common paradigm used in this series of experiments, the hypothesised tradeoff effect was not evident under any of the recall conditions used. The only occasion where recall followed the expected pattern was in the

immediate phase of the delayed recall experiment, and generally the recognition data showed the opposite of the expected fast word advantage where an effect was evident. The data from experiments 7 to 9, while following a generally consistent pattern, did not appear to relate to trends in the data from word length and generation, indicating that the perceptual interference effect may not generalise to the current paradigm. To summarise the above findings, the Eta-squared effect sizes for the three perceptual interference experiments in this chapter are shown at Table 12.

Table 12. Summary of effect sizes from perceptual interference experiments 7 to 9.

Eta-squared Effect sizes	Immediate Serial Recall – Control Item Advantage	Recognition – Control Item Advantage
No Delay Exp. 7	.026	.469
Delayed Exp. 8	.323	.071
Irrelevant Speech Exp. 9	.009	.234

From the overall pattern of effect sizes above, as well as the results for each experiment in this chapter, it would appear that the perceptual interference effect becomes unstable when brought into the classic short-term memory arena. Furthermore, at no stage did the current data replicate effects in the literature which suggest that words which are interfered with at presentation would be better recalled in a final recognition task. It may well be the case that such effects are found when the rate of presentation at study is slowed to 2500 ms, but are not evident with a more rapid form of presentation.

It is beyond the scope and timeframe of the current series of experiments to further explore this effect under differing conditions, and to reduce the rate of presentation in this paradigm to a level where the effect is likely to be observed would invalidate the current experimental paradigm, making direct comparisons between word length and perceptual interference effects impossible.

Chapter 8 Contents

CHAPTER 8

Further Experiments in Word Length and Generation

8.1 Issues Arising from Chapter 6 Experiments

This chapter addresses issues arising from the general attenuation of tradeoff effects in the Chapter 6 generation experiments relative to those found in the Chapter 5 word length experiments. It was suggested in Chapter 6 that weak generation effects may well have been the result of a sensitivity problem. Possible problems were associated with the types of items used in lists and in the method used to generate an item from the given stimulus fragment candidate. With regards to item characteristics, it appeared that the items selected for the generation difficulty manipulation were already all difficult items relative to the less difficult short items used for word length. That is, all the items were multi-syllabic such that both control (read) and generated words for the generation experiments were functionally equivalent to the difficult (long) words in the word length experiments.

If the items from each set of experiments differ in baseline complexity, then the experiments may not be considered fully comparable due to the decreased sensitivity of the generation experiments. There were two approaches taken in response to this question, and the following (Experiments 10 and 11) address the baseline difference by increasing the overall difficulty of the word length experiments, and by using an alternative generation manipulation with shorter words to reduce the overall difficulty of the generation experiments. Thus, the attempt was made to make word length resemble the generation experiments (weak tradeoff effects) and to make generation resemble the word length experiments (strong tradeoff effects). The increase in overall difficulty of the word length effect experiments was attained by the use of articulatory suppression during presentation in Experiment 10. The decrease in overall difficulty of the generation effect experiments was achieved by using shorter items with a corresponding alteration of the generation manipulation in Experiment 11.

8.2 Experiment 10. Word Length & Articulatory Suppression

In the short-term memory literature, word length effects for visually presented material are often eliminated if participants are required to suppress articulation during list learning (Baddeley, et al., 1975). That is, if participants continually repeat a word like "the" during presentation of a list, overall recall for short and long lists is depressed and the word length effect disappears. Assuming that this effect can be replicated, given the results of the generation experiments, one might expect that a weak reverse word length effect could still be evident in the item recognition task.

8.2.1 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. None of these had participated in the previous experiments.

8.2.2 Materials and Procedure. The only changes made to the General Method detailed in Chapter 5 involved the participants undertaking an articulatory suppression task during study of lists. The volunteers were required to rapidly articulate "the the the" about twice every second continuously during presentation of lists. Instructions and computer files were adjusted accordingly.

8.3 Experiment 10 Results

8.3.1 Individual Differences. Estimates of individual patterns were again made from the raw data set, and data were summarised to compare overall group mean recall probabilities. Summary data from the groups are displayed in Table 13.

Table 13. Summary of group data from Experiment 10.

Mean Recall Probability	Immediate Serial Recall			Recognition	
	Correct	Transpos	Extralist	Correct	F Alarms
Short words	0.25	0.12	0.04	0.31	0.10
Long Words	0.26	0.10	0.02	0.39	0.10

Of the twenty participants in Experiment 10, seven showed a weak to moderate tradeoff effect, comprising a weak short word advantage in immediate recall and a moderate long word advantage at recognition. Another seven displayed a consistent long word advantage across both tasks, five a long word advantage at recognition only, and one a moderate long word advantage at immediate recall only.

The mean numbers of false alarms shown in Table 13 are again likely to have been affected by outliers, with three participants recording ten or more errors. If those three are trimmed from the false alarm data, the mean false alarm probability falls to 0.06.

8.3.2 Group Data. Initial analysis of words correctly recalled from lists in Experiment 10 was conducted using a 2 x 2 x 6 repeated measures ANOVA, with the only F-ratio of theoretical interest being the task by effect interaction, which would demonstrate the hypothesised processing tradeoff between immediate serial recall and recognition tasks. The interaction was significant, $F (1,19) = 10.828$, *MSE* = 1.524, $p. = .004$, but did not fully demonstrate the tradeoff effect. Individual analyses shown below for each task clarify this result.

8.3.2.1 Immediate Serial Recall. Analysis of words recalled in correct serial position from the immediate recall task was undertaken using a 2 x 6 repeated measures ANOVA, in the same manner as in the previous experiments. From the immediate recall phase, no main effect was found for word length, $F (1,19) < 1$. A significant main effect was evident for serial position, $F (5,95) = 8.93$, *MSE* = 2.77, $p. < .001$. The interaction was nonsignificant, $F (5,95) < 1$. The serial recall curves for the immediate recall component are shown at Figure 19.

Figure 19. Immediate serial recall curves from Experiment 10.
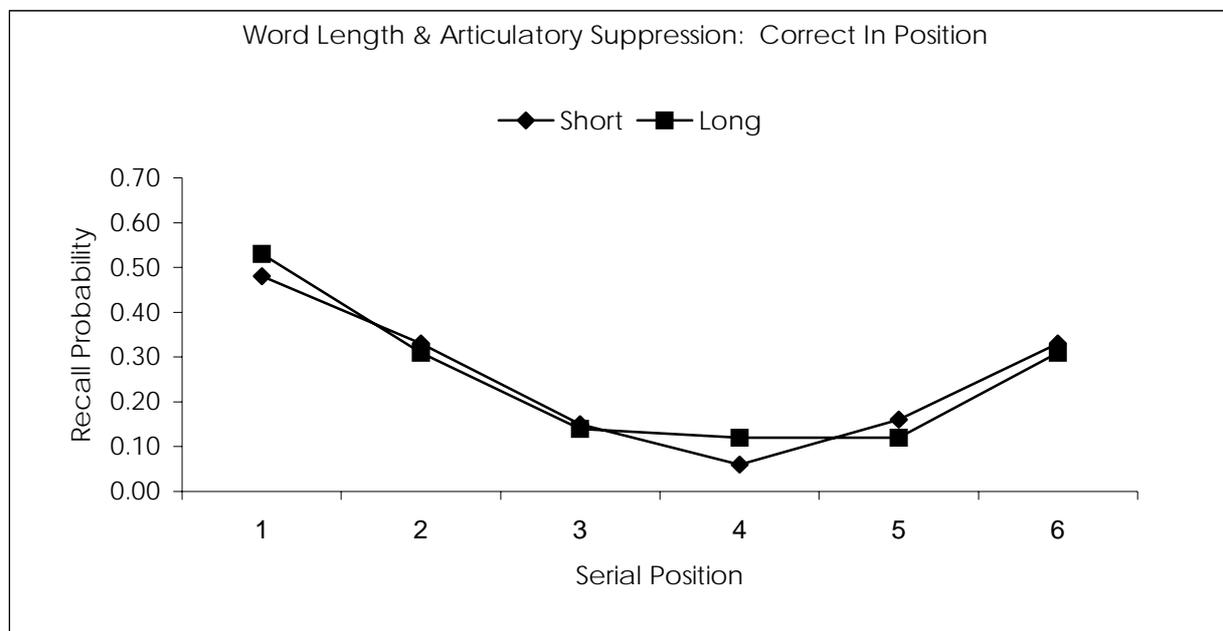
8.3.2.2 Recognition. An identical analysis was performed on data from the recognition test. There was a significant main effect for word length, $F(1,19) = 15.22$, $MSE = 2.27$, $p. = .001$. Figure 20 shows more long words were recognised than short words across all serial positions. A significant main effect for serial position, $F(5,95) = 13.49$, $MSE = 2.72$, $p. < .001$, is also illustrated in Figure 20. There was no significant interaction, $F(5,95) < 1$.

Figure 20. Recognition curves from Experiment 10.

Word Length & Articulatory Suppression: Item Recognition
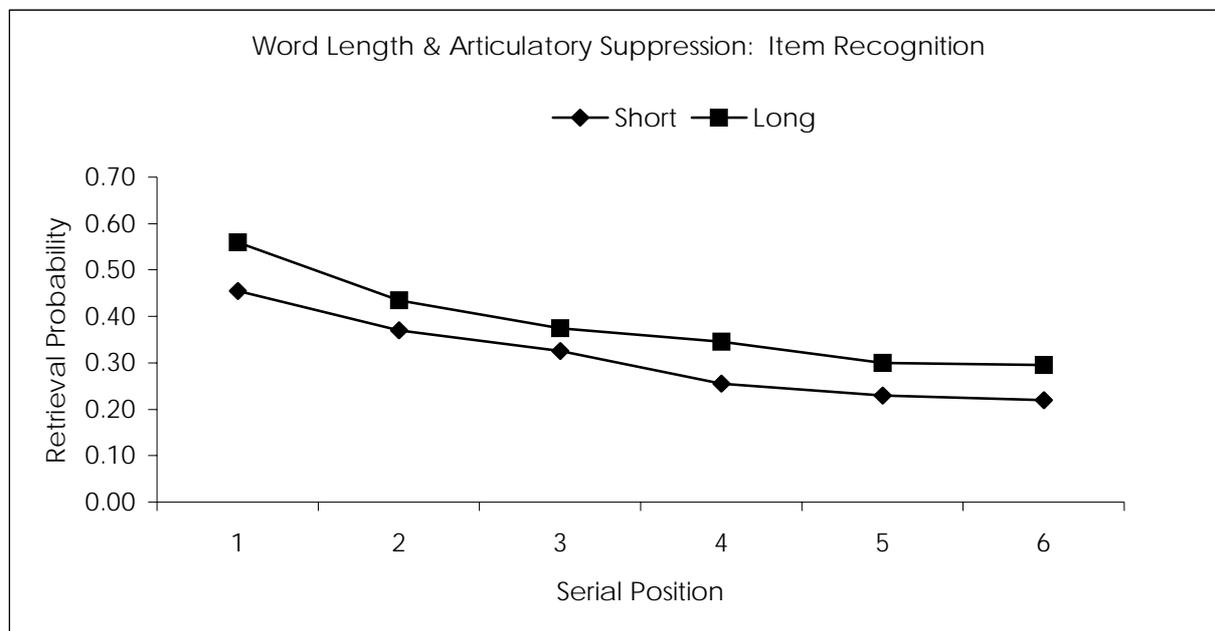
8.4 *Experiment 10 Discussion*

The word length effect found in immediate serial recall in the earlier experiments has disappeared when lists were studied with articulatory suppression, thus replicating many previous studies (Baddeley et al., 1975; LaPointe & Engle, 1990; Tehan et al., 2001). Moreover, the overall difficulty of the task is increased with the addition of articulatory suppression. In short, the immediate serial recall data conform to previous experimental results.

The absence of any word length effect in the immediate phase was not reflected in the recognition task results, which displayed an observable and significant long word advantage. Although there was no serial recall effect, evidence of a word length effect remained in the recognition data, even when rehearsal had been prevented by the disrupting effect of articulatory suppression. At the very least, the above data and those in Chapter 5 suggest that the word length effect remains observable outside its traditional domain of immediate serial recall.

In sum, recall patterns from both phases tends to look very similar to the Chapter 6 generation effect patterns. There are nonsignificant differences in the serial recall component but weak reverse effects in the recognition component. Table 14 below compares the effect sizes from Experiments 10 and 11 with those from the Chapter 6 experiments.

Table 14. Comparison of Eta-squared effect sizes.

Experiment Effect Sizes	Immediate Phase	Recognition	Interaction
Exp.10 Word Length A/S	.001	.445	.363
Exp.4 Generation ISR	.001	.478	.326
Exp.5 Generation DEL	.118	.149	.323
Exp.6 Generation I/S	.001	.422	.280
Exp.11 Generation ISR	.206	.399	.336

Note. All effect sizes shown above are the Eta-squared values from SPSS output. Effect sizes from Experiments 4, 5 and 6 are reprinted here for convenience.

8.5 Experiment 11. Generation and Immediate Serial Recall

In order to reduce the overall task difficulty of the generation effect experiments described in Chapter 6, two approaches were considered. An obvious solution to the problem as described in Chapter 6 was to simply make the items shorter, thus providing a control group equivalent to the short words used in Chapter 5. Another approach would be to change the generation mechanism – many different methods of generation have been researched, some of which were described earlier.

8.5.1 An Alternative Generation Manipulation. One problem prevented the simple substitution of existing materials with short items such as those used in Chapter 5. The requirement in the Chapter 6 generation experiments to generate list items from word fragments, to provide a replication of the Mulligan (1999) study or the Nairne et al. (1991) data, could not be used for items which were shorter, due to the constraint on the original items that there was only one possible legal solution to each word fragment. These unique solutions were shown in their entirety for the recognition

component of the experiments, and so the current experimental paradigm was not appropriate for shorter item stems with multiple solutions.

It was therefore necessary to find an alternate method of manipulating generation using short duration item stems with only one possible legal generated solution. A further study by Mulligan (2002) provided another manipulation of generation which satisfied both the above criteria. Items were presented with the initial two letters swapped around, and participants instructed to generate the item by reversing the order of the first two letters, for example, the item "acr" would be generated as "car". This generation manipulation allowed item stems of any length to be used, with always just one legal solution. Experiment 11, therefore, replicates Experiment 4 in all ways except the (shorter) materials and the different method of generation.

8.5.2 Participants. 20 introductory psychology students from the Australian Catholic University 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. None of these had participated in the previous experiments.

8.5.3 Materials and Procedure. The only changes made to the Experiment 4 Method detailed in Chapter 6 involved changing the materials to shorter items, and altering the generation mechanism so that participants were required to swap the initial two letters of items from experimental lists to create the generated words. All items from control (read) and experimental (swap) lists were read aloud during presentation as before. The immediate serial recall task was unchanged, as was the recognition test following completion of the first phase. Instructions and computer files were modified in accordance with the different generation method and materials. Scoring was again conditionalised upon correct generation at study.

8.6 Experiment 11 Results

8.6.1 Individual Differences. Estimates of individual patterns were again made from the raw data set, which was summarised to compare overall group mean recall probabilities. Summary data from the groups are displayed in Table 15.

Table 15. Summary of group data from Experiment 11.

Mean Recall Probability	Immediate Serial Recall			Recognition	
	Correct	Transpos	Extralist	Correct	F Alarms
Read words	0.35	0.14	0.06	0.22	0.15
Swap Words	0.21	0.16	0.03	0.25	
Cond. Swap	0.29			0.28	

Of the twenty participants in Experiment 11, eight produced an item/order tradeoff effect, comprising a read item advantage in immediate recall with an associated generated word recall advantage at recognition. Four displayed a consistent read item advantage across both tasks, two a read item advantage at immediate recall only, and one a read item advantage at recognition only.

One participant demonstrated a recall advantage for generated items at recognition only, and the remaining four participants showed no discernible effect of generation in either task.

The mean proportion of false alarms shown in Table 15 is again likely to have been affected by outliers, with seven participants recording ten or more errors (five having more than ten errors). If these five are trimmed from the false alarm data, the mean false alarm probability falls to 10.1 per cent, similar to the corresponding value from Experiment 10.

8.6.2 Group Data. Initial analysis of words correctly recalled from lists in Experiment 11 was conducted using a 2 x 2 x 6 repeated measures ANOVA, with the main focus of theoretical interest again being the task by effect interaction, which would demonstrate the hypothesised processing

tradeoff between immediate serial recall and recognition tasks. The interaction was significant, F (1,19) = 9.61, MSE = .052, $p.$ = .006, and reflected the tradeoff effect as hypothesised. Individual analyses for each task are shown below.

8.6.2.1 Immediate Serial Recall. Analysis of words recalled in correct serial position from the immediate recall task was undertaken using a 2 x 6 repeated measures ANOVA, using conditionalised data in the same manner as in the previous experiments. From the immediate recall phase, a significant main effect was found for generation, F (1,19) = 4.93, MSE = .059, $p.$ = .039. There was an immediate recall advantage found for the control (read) items over the generated items. A significant main effect was evident for serial position, F (5,95)= 15.16, MSE = .084, $p.$ < .001, and the interaction was also significant, F (5,95) = 5.15, MSE = .030, $p.$ < .001. Figure 21 displays the serial recall curves for the immediate recall component, including conditional data.

Figure 21. Immediate serial recall curves from Experiment 11.
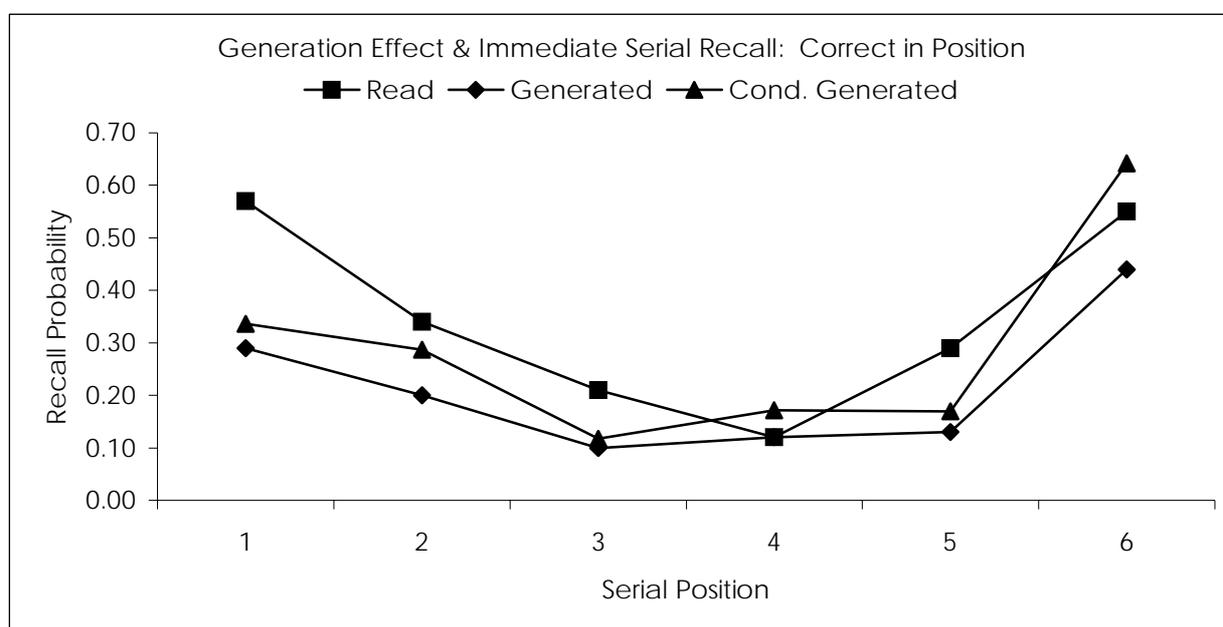
8.6.2.2 Recognition. An identical analysis was performed on data from the recognition test. There was a significant main effect for generation, $F(1,19)= 12.62$, *MSE* = .017, *p.* = .002. Figure 22 shows more generated words were recognised than control items across all serial positions except the first. No significant main effect was found for serial position, $F(5,95) < 1$, and the interaction was not found to be significant, $F(5,95) = 1.20$, *MSE* = .011, *p.* = .32.

Figure 22. Recognition curves from Experiment 11.
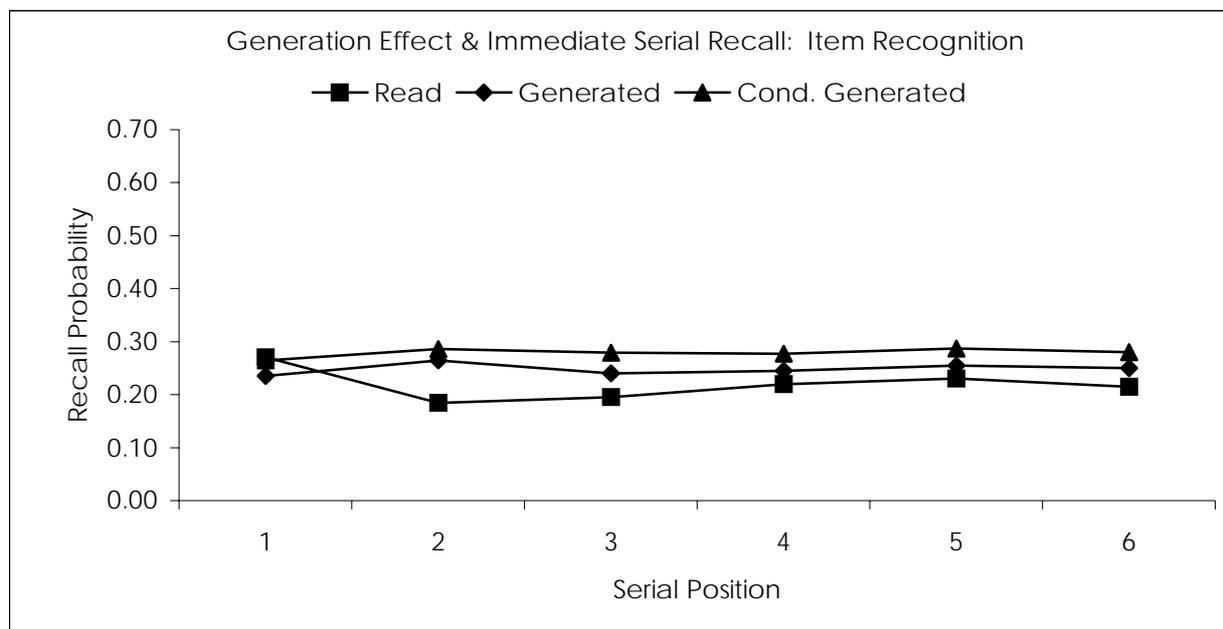
8.7 Experiment 11 Discussion

The manipulation of item length and generation task has produced a different pattern of effects from those found in the earlier generation experiments. While the recognition data appear similar to those in Experiments 4, 5 and 6, a generation effect was found in the immediate phase where none had been found previously. The overall pattern of effects in this experiment supports the notion of a processing tradeoff previously found in long-term generation studies becoming observable in a short-term memory based experimental paradigm.

8.8 Chapter Summary

This last pair of experiments has provided sufficient evidence to suggest the processing tradeoff effect will be found in both word length and generation effects, when investigated using the Nairne et al. experimental paradigm with sufficiently sensitive groups of items. The suspicion in Chapter 6 that the relative length of the test items was leading to a lack of sensitivity in the experimental design seems to have been supported by the results of Experiments 10 and 11.

In experiment 10, the robust word length tradeoff effect found in the preceding Chapter 5 experiments has become seriously weakened by a manipulation which arguably increased the overall difficulty of the experiment, the requirement for participants to undertake concurrent articulatory suppression at study. There was no word length effect found in the immediate phase of the experiment, but a significant long word advantage at recognition.

Table 14 effect sizes for the immediate and recognition phases from Experiment 10 show an almost identical pattern and magnitude to those from Experiment 4 (generation and immediate serial recall) and Experiment 6 (generation and irrelevant speech). It seems reasonable to suggest that an increase in the overall difficulty of the experiment has led to reduced sensitivity, which specifically manifests as no effect in the immediate phase, but a significant recall advantage for difficult items at recognition. Half the expected tradeoff is observed in each case, whether the increase in overall difficulty is due to articulatory suppression or the use of long items in a control condition.

Experiment 11, where the generation manipulation was changed, demonstrates the reverse approach. Thus when the generation manipulation was made 'easier' by reducing the length of the words, there was a read advantage in the serial recall component, and the expected tradeoff was observed. With this more sensitive design, the expected generation tradeoff effect emerged.

The overall results of these last two experiments produce the expected outcomes. The word length effect results look like the generation results in Chapter 6 and the generation results

look like the word length effects in Chapter 5. Overall the full set of experiments suggests that an item/order processing tradeoff is observable for both word length and generation.

Chapter 9 Contents

CHAPTER 9.

General Discussion

The aim of this dissertation was to test an item and order processing tradeoff account of the word length effect. The methodology employed a modified version of the Nairne et al., (1991) paradigm in which lists were learnt for serial recall and were then tested either by a serial order recall task or by a single item recognition task. In addition to word length, the rationale was to test generation and perceptual interference effects under the same experimental conditions. The latter two were chosen because they have widely been interpreted within the item-order processing tradeoff perspective. Two sources of evidence were proposed to support the item-order tradeoff. The first was that the word length effect experiments should dissociate in the direction of their effects between serial order and item recognition tasks. On the proviso that this pattern would replicate in the generation and perceptual interference effect experiments, a compelling case could be made that all three effects had a common foundation, that being the tradeoff in encoding of item and order information during list learning.

9.1 Summary of Experiments

The first source of evidence was readily apparent. The word length experiments in Chapter 5 all demonstrated a short word recall advantage in serial recall and a long word advantage in item recognition. The second source of evidence was less convincing. In the generation experiments there were signs of the expected pattern in the item recognition component, but no differences were found in the serial recall component. With the perceptual interference experiments there was no sign of any dissociations in any of the experiments. On the basis of the current data, therefore, a compelling case for a common underpinning of the three effects cannot be convincingly made.

The experiments in Chapter 8 were an attempt to salvage the notion of a common foundation. Changes were made to the procedures to make the word length effect resemble the generation effect; and to make the generation effect resemble the word length effect. In both

instances, results consistent with the processing tradeoff perspective emerged. That is, it was possible to observe an attenuation of the word length effect in serial recall complemented by a weak long word advantage in item recognition under suppression. Likewise, the predicted dissociation across serial order and item tasks was observed with a finer manipulation of generation. The results from Chapter 8 provide some support for the idea that a common set of processes may underlie both the word length and generation effects, but the argument cannot be made as strongly as had been hoped.

9.2 Item Manipulations Vs Encoding Manipulations

In a sense it is not surprising that the results emerged as they did, in that there are some key differences between the word length experiments and the generation and perceptual interference experiments. The word length effect boils down to being a materials effect where the key manipulation involves selecting items with specific characteristics. In the study phase of these experiments the actual encoding manipulation was equivalent for the two types of materials, in that all the words were studied silently. In the case of generation and perceptual interference conditions, the selection of materials was largely irrelevant. The crucial manipulation was the encoding condition. Participants read items, or generated them or viewed them under masked conditions. So in a sense, while it is argued that all three effects can be explained in terms of an item-order processing tradeoff, two are the result of encoding manipulations while the other is due to item selection manipulations.

9.3 Word Length and Word Frequency

If comparing item manipulations with encoding manipulations is less than optimal, a better comparison might be to compare performance on two item manipulation effects. DeLosh and McDaniel (1996) argued that the word frequency effect could also be interpreted from an item-order tradeoff perspective. As they say "For lists consisting of common, run-of-the-mill items, we argue that order information tends to be encoded... In contrast, we assume that (with) lists of uncommon or unusual items... the learner's resources are lured to processing and interpreting the

individual and idiosyncratic features of the unusual items, leaving fewer resources for encoding of order information." (DeLosh & McDaniel, 1996, p. 1137).

The basic components of the item-order tradeoff are already well established in the literature. Common words are better recalled than low frequency words in short-term serial recall (Tehan & Humphreys, 1988; Watkins, 1977). In item recognition tasks it has long been known that single item recognition is better for low frequency words than high frequency words. So at face value there is a similar pattern between word length effects and word frequency effects. The similarity between word length and word frequency effects is even more apparent when the DeLosh and McDaniel experiments are considered.

Their (1996) experiments dealt with two issues, the comparison of free recall with serial recall, and the use of mixed and pure lists. Regarding free recall, they argued that with short lists (8 words) participants were likely to use order information in performing the free recall task. As such, they argued that there should be a high frequency word advantage in both free recall and in serial recall. The use of mixed lists was predicated on the finding that the generation effect (Serra & Nairne, 1993), the bizarreness effect (McDaniel et al, 1995) and the perceptual interference effect (Mulligan, 2001) all disappear when mixed lists are used. The argument here is that with mixed lists, the order encoding of high frequency items is disrupted by the presence of low frequency items relative to pure lists and that for low frequency items, order encoding is facilitated relative to pure lists. The results were as expected. With pure lists there was a high frequency advantage in both serial recall and free recall. With mixed lists, there was evidence of reduced order encoding across the board and the word frequency effect disappeared.

Similar effects can be found in the word length literature. For example, Russo and Grammatopoulou (2003) explored word length effects in free recall of short lists. Across a number of experiments they found a reliable short word advantage. Similarly, Hulme, Surprenant, Bireta, Stuart and Neath (2004) explored word length effects in pure and mixed lists requiring immediate

serial recall. With pure lists they found the short word advantage but with mixed lists that effect disappeared.

In sum, there are quite striking similarities between the word frequency and word length effects across a number of tasks and a number of encoding conditions. Given the explanation of the word frequency effect in terms of an item-order tradeoff, it is possible that the word length effect may well be due to the same processes.

9.4 Applying the Item/Order account to other word length effects

The above discussion of free recall indicates that the word length effect extends beyond immediate serial recall. If the item-order processing hypothesis is correct, a short word advantage should be evident in any task where order information is being used. The current approach thus readily accounts for the findings that a short word advantage has been found in complex span tasks where items and distractor activity alternate with each other – as well as in the Brown-Peterson task where a 12 second filled retention interval was utilised (Tehan et al., 2001). Furthermore, word length effects in backward recall (Cowan, Wood & Borne, 1994) are explainable given the assumption that backward recall is accomplished via a series of forward recalls (Thomas, Milner & Haberlandt, 2003).

However, word length effects have been observed in probed recall tasks where ordered recall of multiple items is not required (Avons, Wright & Pammer, 1994; Henry, 1991; Henry, Turner, Smith & Leather, 2000). As has been demonstrated with free recall, participants can adopt a serial recall strategy even though the instructions do not stress the use of order information, and to the extent that this is so, one would expect the standard word length effect. A probe recall study by Henry et al. (2000) illustrates this point. They examined the emergence of word length effects as a function of age. In their first experiment, serial recall was requested and child participants in the three age groups used (4-year-old, 7-year-old and 10-year-old) all produced reliable word length effects. In a second experiment, probe recall was required rather than serial recall. In the case where only one item had to be recalled, the 4-year-olds did not exhibit word length effects. In fact

there was a tendency for reverse word length effects to emerge. With the two older groups, word length effects were present in this task but were stronger for those students who reported using a serial recall strategy. Those who reported using a simple naming strategy produced very weak effects. Thus, even though the probe recall task does not require serial order to be utilised, participants still adopted a serial rehearsal strategy. Interestingly, as implied by the item-order processing tradeoff perspective, word length effects were not strong when serial rehearsal was not required.

Finally, word length effects have been observed in a serial recognition task (Baddeley, Chincotta, Stafford & Turk, 2002). In this task, participants were presented a list of items for study, followed by a probe item which contained either the list presented again in its original serial order, or with two of the items in the list transposed. The task was to indicate if the probe list maintained the items in their original positions or not. Again, since the task required participants to utilise order information, the item/order hypothesis would have predicted a short word advantage.

The one instance of word length effects in the literature that the item/order hypothesis would not be able to explain is the results of a study by Cowan, Nugent, Elliott and Geer (2000). They manipulated output speed by having participants articulate their responses either at a fast speed or an exaggeratedly slow speed. Given that the same items were involved in each case, the deficit for the slowly articulated responses suggests that forgetting was occurring during output. However, this research has not been free from criticism on the basis of different attentional demands (Service, 2000).

Another potentially problematic finding is that reported by Cowan, Wood and Borne (1994) where six-item lists were studied for backward recall. They studied word length effects under immediate serial recall and under continuous distractor conditions where each list item was preceded and followed by 15 seconds of distractor activity. They found the standard short word advantage on the immediate test, but the advantage was reversed on the continuous distractor condition. On the basis of these findings they argued that separate short and long term memory

systems were required. From the item/order perspective, the continuous distractor task is one that required serial order, yet given all the previous evidence that order information is lost relatively quickly (transposition errors rapidly become omission errors), it seems reasonable to suppose that with a one and a half minute retention interval for the early list items that such order information is rapidly lost. As such even though the task requires serial information, participants may be relying predominantly upon item information with the result that a long word advantage emerges. This argument is the mirror of that used to explain standard word length effects in free recall. In that task it was suggested that participants use order information although the task does not demand it. With the continuous distractor task, participants are required to use serial order information, but the task may be so difficult that item information is used to a greater extent than order information.

9.5 Theoretical Implications

The word length effect has been one of the cornerstones of theorising about short-term recall and has been one of the prime pieces of evidence used to support the notion that short-term memory traces decay rapidly. While the decay explanation of the word length effect has become less plausible, alternative explanations of the word length effect have also been problematic. The current research here adopted a processing approach to the problem to investigate whether the item-order perspective that has successfully been applied to the generation and perceptual interference effects might generalise to the word length effect. The results were consistent with this expectation, in that word length produced the dissociative effects on item and order tasks that are typically observed with generation and perceptual interference.

The fact that word length effects can be observed across a wide range of tasks is also problematic for most alternative models. While standard short-term decay explanations might apply to the immediate serial recall, it is less likely that they apply to word length effects in complex span (Tehan et al., 2001), delayed recall tasks (Tehan et al, 2001; Current experiment 2), free recall (Russo & Grammatopoulou, 2003), probe recall (Avons et al, 1994) and serial recognition,(Baddeley et al., 2002). The more parsimonious explanation is that the word length effect occurs in all these

situations because long words do not receive as much order processing as short words. If the item-order explanation of the word length effect can be confirmed, the foundations of the structural models of short-term memory that are based upon decay as the source of forgetting will be seriously undermined. Rather, the data point more to a unitary view of memory in which the type of processing and type of test employed determine in large part what is recalled. Of course, this latter idea is not new (Craik & Lockhart, 1972; Einstein & Hunt, 1980).

Finding word length effects in a number of short-term memory tasks may be problematic for decay based accounts of the word length effect, but they are probably not so problematic for the word complexity accounts (Caplan, Rochon & Waters, 1992; Service, 1998). The locus of word length effects in these models is in long-term lexical memory, and to the extent that lexical memory is involved in the various tasks, one might reasonably predict that word length effects should generalise across tasks. Thus the item complexity account would probably predict that standard word length effects should be observed in the range of tasks described above. However, it is only the item-order processing tradeoff account that makes the prediction that reverse word length effects should be observed on item recognition tests. The results of Experiments 1 to 3 and 10 are the crucial and reasonably compelling data for the item-order account of the word length effect.

While the word length data are consistent with the item-order explanation, there may be some problems with this approach. The fundamental assumption made in this paper is that long words take longer and/or are more difficult to process at the item level than short words and that this reduces the amount of subsequent order processing. There are two aspects of this assumption that may be problematic for understanding word length effects in immediate serial recall. Firstly, while there is probably some face validity to the idea of differential item processing for the generation and perceptual interference effects, it is not as self-evident that long words receive more item processing than short words. Secondly, the explanation assumes that item processing and subsequent identification is a key determinant of short-term memory performance.

As mentioned in the introductory chapters, word length effects are observed in lexical access tasks such as word naming, perceptual identification and lexical decision (Balota & Chumbley, 1985; Forster & Chambers, 1973; Samuels, Laberge & Bremer, 1978). Thus, there is direct evidence that long words take longer to identify and respond to than short words. While this lexical access literature would thus appear to support the item-order trade-off account, there is no direct evidence that supports the notion that increased identification time results in enhanced item processing.

With regards to item identification in immediate memory, the individual differences literature has indicated that one of the prime determinants of immediate memory span is the speed at which items can be identified. For instance, in a review of the literature to that time, Dempster (1981) examined ten possible sources of individual differences in memory span. His review indicated that item identification speed was the most reliable source of individual differences in span among children. The relationship between item identification time and span has since been demonstrated on a number of occasions with both children (Case, Kurland & Goldberg, 1982; Hitch, Halliday & Littler, 1989; 1993), and adults (Tehan & Lalor, 2000) as participants (but see Henry & Millar, 1993, for an alternative view).

The Tehan and Lalor (2000) data are relevant here in that they demonstrated that performance on lexical decision, word naming and other visual word decoding tasks made a significant contribution to individual differences in span performance; a contribution that was more important that traditional rehearsal and output time measures. Importantly, the tasks they used to establish the relationship between span and lexical access were the same tasks that were showing word length differences in the lexical access literature and the same tasks that are used in the item identification literature. Thus, the available literature does support the notion that item processing speed is important in immediate serial recall and that short words are processed faster than long words. Thus, the argument that the word length effect in serial recall is due to differences in item processing does have some support.

9.6 Problems and Limitations.

There are some potential weaknesses of the current experiments, in that while the standard word length effect can be observed across a range of short-term and long-term order memory tasks, the reverse word length effect has only been demonstrated in a single item memory task, namely long-term recognition. The item-order tradeoff account would be more compelling if reverse effects could be observed across a range of item memory tasks. Long-term free recall has often been used as a measure of item memory, but as indicated above this task is prone to differential strategy usage. Cued recall may well be a potential candidate to further investigate the item/order processing tradeoff.

The second problem is that in modifying the Nairne et al. (1991) procedure, a potential confound has been introduced. The serial recall component of the task is tested immediately, while item recognition is tested after quite a substantial delay. To some readers, this would equate with a short-term memory test followed by a long-term memory test. Introducing this confound was unavoidable if the goal of marrying the word length effect in immediate recall with the standard item-order methodology were to be achieved. However, to have a complete account, one would need to deconfound retention interval with type of memory task. As such, what is needed is to demonstrate that the standard word length effect could be observed in a typical long-term serial order task like serial list learning. Likewise, reverse word length effects would need to be demonstrated in, for instance, a short-term item recognition or cued recall task. These latter two may be problematic because in the short-term domain, participants can utilise order information irrespective of the type of memory task that is employed (Beaman & Jones, 1997).

9.7 Conclusion

In conclusion, the current results, while consistent with the item-order processing approach, have a correlational flavour to them. Because the pattern of word length effects seems to mimic generation and perceptual interference effects, it is assumed that the same processes underlie all three phenomena. Whether this is in fact the case will require further investigation. For instance,

there are many factors which influence the trade-off in item-order processing in generation and perceptual interference; these same factors would need to influence word length in the same way. Vice versa, there are many variables that have an impact upon word length in the short-term domain. Again, generation and perceptual interference effects would need to parallel word length effects in the short-term tasks if common processes are to be assumed. However, the current paradigm does provide a novel and potentially useful approach to understanding word length effects in the short-term domain.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. R. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.

Avons, S. E., Wright, K. L., & Pammer, K. (1994). The word length effect in probed and serial recall. *Quarterly Journal of Experimental Psychology*, **20**, 249-264.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, **18**, 363-365.

Baddeley, A. D. (1978). The trouble with levels: A re-examination of Craik and Lockhart's framework for memory research. *Psychological Review*, **85**, 139-152.

Baddeley, A. D. (1990). *Human memory: Theory and practice*. Boston: Allyn and Bacon.

Baddeley, A. D. (1996). *Human memory: Theory and practice*. (2nd Ed.) Hove, UK: Psychology Press.

Baddeley, A. D., & Andrade, J. (1994). Reversing the word length effect: A comment on Caplan, Rochon & Waters. *Quarterly Journal of Experimental Psychology*, **47A**, 1047-1054.

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, A.,* **55A (2)**, 353–369.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 8, pp. 47-89). New York: Academic Press.

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.

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(1)**, 89-106.

Beaman, C. P., & Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **23(2)**, 459-471.

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

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, **10,** 12-21.

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

Burns, D. J. (1986). *Differential rates of forgetting: The exception or the rule?* Unpublished doctoral dissertation, State University of New York at Binghamton.

Burns D. J., Curti, E. T., & Lavin, J. C. (1993). The effects of generation on item and order retention in immediate and delayed recall. *Memory & Cognition*, **21**, 846-852.

Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word-length effects in span tasks. *Quarterly Journal of Experimental Psychology*, **45A**, 177-192.

Caplan, D., & Waters, G. S. (1994). Articulatory length and phonological similarity in span tasks: a reply to Baddeley & Andrade. *Quarterly Journal of Experimental Psychology*, **45A**, 177-192.

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.

Coltheart, V., & Langdon, R. (1998). Recall of short word lists presented visually at fast rates: Effects of phonemic similarity and word length. *Memory & Cognition,* **26(2)**, 330-342.

Cowan, N., Baddeley, A. D., Elliott, E. M., & Norris, J. (2003). List composition and the word length effect in immediate recall: A comparison of localist and globalist assumptions. *Psychonomic Bulletin and Review,* **10(1)**, 74-79.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, **31**, 1-17.

Cowan, N., Nugent, L. D., Elliott, E. M., & Geer, T. (2000). Is there a temporal basis of the word length effect?: A response to Service (1998). *Quarterly Journal of Experimental Psychology*, **53A**, 647-700.

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

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

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.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin*, **89**, 63-100.

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organisation: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 588-598.

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, **71**, 43-52.

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26(3)**, 671-682.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behaviour*, **12(6)**, 627-635.

Gathercole, S. E. (1997). Models of verbal short-term memory. In Conway, M. A. (Ed.) *Cognitive Models of Memory*. Guildford, UK: Biddles.

Greene, R. L., Thapar, A., & Westerman, D. L. (1998). Effects of generation on memory for order. *Journal of Memory and Language*, **38**, 255-264.

Henry, L. A. (1991). The effect of word length and phonemic similarity in children's short term memory. *Quarterly Journal of Experimental Psychology*, **43A**, 35-52.

Henry, L. A., & Millar, S. (1993). Why does memory span improve with age? A review of the evidence for two current hypotheses. *European Journal of Cognitive Psychology*, **5(3)**, 241-287.

Henry, L. A., Turner, J. E., Smith, P. T., & Leather, C. (2000). Modality effects and the development of the word length effect in children. *Memory*, **8(1)**, 1-17.

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.

Hitch, G. J., Halliday, M. S., & Littler, J. E. (1993). Development of memory span for spoken words: The role of rehearsal and item identification processes. *British Journal of Developmental Psychology*, **11**, 159-169.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, **30**, 685-701.

Hulme, C., Surprenant, A. M., Bireta, T. J., Stuart, G., & Neath, I. (2004). Abolishing the word-length effect. *Journal of experimental Psychology: Learning, Memory & Cognition*, **30(1)**, 98-106.

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, **38**, 241-253.

Kelley, M. R., & Nairne, J. S. (2001). von Restorff revisited: Isolation, generation, and memory for order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **27(1)**, 54-66.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

LaPointe, L. B., & Engle, R. W. (1990). Simple and complex spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 1118-33.

Logie, R. H., Della Salla, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition*, **24(3)**, 305-321.

Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *Quarterly Journal of Experimental Psychology*, **53A**, 1-22.

Lovatt, P., Avons, S. E., & Masterson, J. (2002). Output decay in immediate serial recall: Speech time revisited. *Journal of Memory and Language*, **46(1)**, 227-243.

McDaniel, M. A., Einstein, G. O., DeLosh, E. L., May, C. P., & Brady, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 422-435.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behaviour*, **2**, 1-21.

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

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.

Mulligan, N. W. (2000). Perceptual interference and memory for order. *Journal of Memory & Language*, **43**, 680-697.

Mulligan, N. W. (2001). Generation and hypermnesia. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27(2)**, 436-450.

Mulligan, N. W. (2002). The generation effect: Dissociating enhanced item memory and disrupted order memory. *Memory & Cognition*, **30(6)**, 850-861.

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition*, **16**, 343-352.

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., Neath, I., & Serra, M. (1997). Proactive interference plays a role in the word-length effect. *Psychological Bulletin and Review,* **4**, 541-545.

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.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review,* **7**, 403-423.

Neath, I., & Nairne, J S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin and Review*, **2(4)**, 429-441.

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

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

Russo, R., & Grammatopoulou, N. (2003). Word length and articulatory suppression affect short-term and long-term recall tasks. *Memory & Cognition*, **31(5)**, 728-737.

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(6)**, 715-720.

Schweickert, R., Guentert, L., & Hersberger, L. (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science*, **1**, 74-77.

Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **12**, 419-425.

Searleman, A., & Herrmann, D. (1994). *Memory from a broader perspective*. New York: McGraw-Hill.

Serra, M. & Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition*, **21**, 34-40.

Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology*, **51A**, 283-304.

Service, E. (2000). Phonological complexity and word duration in immediate recall: Different paradigms answer different questions – A comment on Cowan, Nugent, Elliott, and Geer. *Quarterly Journal of Experimental Psychology*, **53A**, 661-665.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 592-604.

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(4/5/6)**, 333-348.

Tehan, G., & Humphreys, M. S. (1988). Articulatory Loop explanations of memory span and pronunciation rate correspondences: A cautionary note. *Bulletin of the Psychonomic Society*, **26**, 293-296.

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.

Tehan, G., & Turcotte, J. (2002). Word length effects are not due to proactive interference. *Memory*, **10(2)**, 139-149.

Thomas, J. G., Milner, H. R., & Haberlandt, K. F. (2003). Forward and backward recall: Different response time patterns, same retrieval order. *Psychological Science,* **14(2)**, 169-174.

von Restorff, H. (1933). On the effect of field formations in the trace field: 1. Analysis of processes in the trace field. *Psychologische Forschung*, **18**, 299-342.

Westerman, D. L., & Greene, R. L. (1997). The effects of visual masking on recognition: Similarities to the generation effect. *Journal of Memory and Language*, **37**, 584-596.

Watkins, M. J. (1977). The intricacy of memory span. *Memory & Cognition*, **5**, 529-534.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, **72**, 89-104.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. R. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, **18**, 363-365.

Baddeley, A. D. (1990). *Human memory: Theory and practice*. Boston: Allyn and Bacon.

Baddeley, A. D. (1996). *Human memory: Theory and practice*. (2nd Ed.) Hove, UK: Psychology Press.

Baddeley, A. D., & Andrade, J. (1994). Reversing the word length effect: A comment on Caplan, Rochon & Waters. *Quarterly Journal of Experimental Psychology*, **47A**, 1047-1054.

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, A.,* **55A (2)**, 353–369.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 8, pp. 47-89). New York: Academic Press.

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.

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(1)**, 89-106.

Beaman, C. P., & Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **23(2)**, 459-471.

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

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, **10,** 12-21.

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

Burns, D. J. (1986). *Differential rates of forgetting: The exception or the rule?* Unpublished doctoral dissertation, State University of New York at Binghamton.

Burns D. J., Curti, E. T., & Lavin, J. C. (1993). The effects of generation on item and order retention in immediate and delayed recall. *Memory & Cognition*, **21**, 846-852.

Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word-length effects in span tasks. *Quarterly Journal of Experimental Psychology,* **45A**, 177-192.

Caplan, D., & Waters, G. S. (1994). Articulatory length and phonological similarity in span tasks: a reply to Baddeley & Andrade. *Quarterly Journal of Experimental Psychology*, **45A**, 177-192.

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.

Coltheart, V., & Langdon, R. (1998). Recall of short word lists presented visually at fast rates: Effects of phonemic similarity and word length. *Memory & Cognition,* **26(2)**, 330-342.

Cowan, N., Baddeley, A. D., Elliott, E. M., & Norris, J. (2003). List composition and the word length effect in immediate recall: A comparison of localist and globalist assumptions. *Psychonomic Bulletin and Review,* **10(1)**, 74-79.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, **31**, 1-17.

Cowan, N., Nugent, L. D., Elliott, E. M., & Geer, T. (2000). Is there a temporal basis of the word length effect?: A response to Service (1998). *Quarterly Journal of Experimental Psychology*, **53A**, 647-700.

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

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.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin*, **89**, 63-100.

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organisation: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory,* **6**, 588-598.

Ellis, N. C., & Hennelley, R. A. (1980). A bilingual word length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, **71**, 43-52.

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26(3)**, 671-682.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behaviour*, **12(6)**, 627-635.

Gathercole, S. E. (1997). Models of verbal short-term memory. In Conway, M. A. (Ed.) *Cognitive Models of Memory*. Guildford, UK: Biddles.

Greene, R. L., Thapar, A., & Westerman, D. L. (1998). Effects of generation on memory for order. *Journal of Memory and Language*, **38**, 255-264.

Henry, L. A. (1991). The effect of word length and phonemic similarity in children's short term memory. *Quarterly Journal of Experimental Psychology,* **43A**, 35-52.

Henry, L. A., & Millar, S. (1993). Why does memory span improve with age? A review of the evidence for two current hypotheses. *European Journal of Cognitive Psychology,* **5(3)**, 241-287.

Henry, L. A., Turner, J. E., Smith, P. T., & Leather, C. (2000). Modality effects and the development of the word length effect in children. *Memory*, **8(1)**, 1-17.

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.

Hitch, G. J., Halliday, M. S., & Littler, J. E. (1993). Development of memory span for spoken words: The role of rehearsal and item identification processes. *British Journal of Developmental Psychology*, **11**, 159-169.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, **30**, 685-701.

Hulme, C., Surprenant, A. M., Bireta, T. J., Stuart, G., & Neath, I. (2004). Abolishing the word-length effect. *Journal of experimental Psychology: Learning, Memory & Cognition,* **30(1)**, 98-106.

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, **38**, 241-253.

Kelley, M. R., & Nairne, J. S. (2001). von Restorff revisited: Isolation, generation, and memory for order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **27(1)**, 54-66.

LaPointe, L. B., & Engle, R. W. (1990). Simple and complex spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 1118-33.

Larsen, J. D., & Baddeley, A. D. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression and manual tapping: Do they have a common source? *Quarterly Journal of Experimental Psychology*, **56A(8)**, 1249-1268.

Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *Quarterly Journal of Experimental Psychology*, **53A**, 1-22.

McDaniel, M. A., Einstein, G. O., Delosh, E. L., May, C. P., & Brady, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 422-435.

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

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.

Mulligan, N. W. (2000). Perceptual interference and memory for order. *Journal of Memory & Language*, **43**, 680-697.

Mulligan, N. W. (2001). Generation and hypermnesia. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27(2)**, 436-450.

Mulligan, N. W. (2002). The generation effect: Dissociating enhanced item memory and disrupted order memory. *Memory & Cognition*, **30(6)**, 850-861.

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition*, **16**, 343-352.

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., Neath, I., & Serra, M. (1997). Proactive interference plays a role in the word-length effect. *Psychological Bulletin and Review,* **4**, 541-545.

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.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review,* **7**, 403-423.

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

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(6)**, 715-720.

Schweickert, R., Guentert, L., & Hersberger, L. (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science*, **1**, 74-77.

Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **12**, 419-425.

Searleman, A., & Herrmann, D. (1994). *Memory from a broader perspective*. New York: McGraw-Hill.

Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology*, **51A**, 283-304.

Service, E. (2000). Phonological complexity and word duration in immediate recall: Different paradigms answer different questions – A comment on Cowan,

Nugent, Elliott, and Geer. *Quarterly Journal of Experimental Psychology*, **53A**, 661-665.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 592-604.

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(4/5/6)**, 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.

Tehan, G., & Turcotte, J. (2002). Word length effects are not due to proactive interference. *Memory*, **10(2)**, 139-149.

Thomas, J. G., Milner, H. R., & Haberlandt, K. F. (2003). Forward and backward recall: Different response time patterns, same retrieval order. *Psychological Science,* **14(2)**, 169-174.