Re-Examining the Phonological Similarity Effect in Immediate Serial Recall: The Roles of Type of Similarity, Category Cueing, and Item Recall

Prahlad Gupta
John Lipinski
Emrah Aktunc

Department of Psychology
University of Iowa
Iowa City, IA 52242
USA

Address correspondence to:

Prahlad Gupta
Department of Psychology
University of Iowa
Iowa City, IA 52242
USA

Phone: +1 319-335-2908
Fax: +1 319-335-0191
Email: prahlad-gupta@uiowa.edu

Running head: PHONOLOGICAL SIMILARITY EFFECT REEXAMINED

[[MEMORY & COGNITION, in press]]

Abstract

Study of the phonological similarity effect (PSE) in immediate serial recall (ISR) has produced a conflicting body of results. Five experiments tested various theoretical ideas that together may help integrate these results. Experiments 1 and 2 tested alternative accounts that explain the effect of phonological similarity on item recall in terms of either feature overlap, linguistic structure, or serial order. In each experiment, participants' ISR was assessed for rhyming, alliterative, and similar nonrhyming/nonalliterative lists. The results were consistent with the predictions of the serial order account, with item recall being higher for rhyming than alliterative lists, and higher for alliterative than similar nonrhyming/nonalliterative lists. Experiments 3 and 4 showed that these item recall differences are reduced when list items repeat across lists. Experiment 5 employed rhyming and dissimilar one-syllable and two-syllable lists to demonstrate that recall for similar (rhyming) lists can be better than for dissimilar lists even in a typical ISR task using words, providing a direct reversal of the classic PSE. These and other previously published results are interpreted and integrated within a proposed theoretical framework that offers an account of the PSE.

Re-Examining the Phonological Similarity Effect in Immediate Serial Recall: The Roles of Type of Similarity, Category Cueing, and Item Recall

A classic finding in the study of immediate serial recall of lists of verbal materials is that recall is poorer for lists consisting of phonologically similar items such as *{cad, map, man, cap, mad}* than for lists consisting of phonologically dissimilar items such as *{pit, day, pen, bar, few}* (e.g., Baddeley, 1966; Conrad, 1964). This is known as the *phonological similarity effect* (PSE); it has come to be viewed as one of the key phenomena characterizing immediate serial recall of verbal lists (e.g., Baddeley, 1992; Baddeley & Hitch, 1994; Baddeley, Gathercole, & Papagno, 1998). It also formed the basis of the theoretical proposal that the verbal short-term memory component of working memory employs a phonological code (Baddeley & Hitch, 1974; Baddeley, 1986); that is, the PSE constitutes key evidence for positing the very existence of what has come to be known as the *phonological loop*. It is thus a phenomenon of central importance to the influential working memory model of short-term memory.

The standard interpretation of the PSE is that it arises as a result of interference between similar phonological memory traces in the phonological store (Baddeley, 1986). The classic PSE is a robust effect, and has been replicated numerous times (e.g., Baddeley, Lewis, & Vallar, 1984; Poirier & Saint-Aubin, 1996; Watkins, Watkins, & Crowder, 1974; Wickelgren, 1965). Other investigations have indicated, however, that phonological similarity may have no detrimental effect on immediate serial recall (e.g., Fallon, Groves, & Tehan, 1999, Experiment 1), and that it may even facilitate memory for item identity (as opposed to order) in immediate serial recall (e.g., Gathercole, Gardiner, & Gregg, 1982), and memory for order (as opposed to item identity) in order reconstruction tasks (Nairne & Kelley, 1999). This research has highlighted the importance of distinguishing between the typical serial, item-in-position criterion of correctness, whereby a list item is scored as correctly recalled only if it is recalled in the correct serial position ("strict serial recall"), and an order-free item criterion of correctness ("item recall"), whereby a list item is scored as correctly recalled if it is produced during recall of a list, whether or not it was produced in the correct serial position.

For example, Watkins et al. (1974) compared serial recall of phonologically similar and phonologically dissimilar lists. When performance was assessed using the strict serial recall measure, it was better for the phonologically distinct lists, demonstrating the classic PSE. However, when performance was assessed using the item recall measure, it was no different for the phonologically similar versus dissimilar lists. Similarly, Gathercole et al. (1982) compared serial recall of phonologically similar and phonologically dissimilar lists. Using the strict serial recall measure, performance was better for the phonologically distinct lists; however, item recall was actually better for the phonologically similar lists than for the phonologically dissmilar lists (Gathercole et al., 1982, p.180). Along similar lines, a study by Poirier and Saint-Aubin (1996) examined serial recall of lists of 2-syllable words. Strict serial recall was better for the phonologically distinct lists, but item recall was no different for the phonologically similar versus dissimilar lists.

However, there have also been studies that obtained the classic PSE (better recall for dissimilar than similar lists) using both the strict serial recall and item recall scoring criteria. For example, Coltheart (1993) found that recall was better for phonologically dissimilar than for phonologically similar lists in terms of both the strict serial recall and item recall measures (Coltheart, 1993, Experiment 1). Similarly, Drewnowski (1980) found that recall was better for the dissimilar than for the similar lists in terms of both strict serial recall and item recall (Drewnowski, 1980, Experiment 3).

In an insightful analysis, Fallon et al. (1999) noted that the differing effects of phonological similarity obtained in different investigations appear to be related to how the notion of phonological similarity had been operationalized in the studies. The studies that obtained a facilitatory effect of phonological similarity at the item level employed rhyming list items in their phonologically similar lists (Gathercole et al., 1982; Wickelgren, 1965). The studies that obtained a detrimental effect or no effect of phonological similarity at the item level operationalized phonological similarity in terms of list items that shared high phonemic overlap but that did not all rhyme within a list (Coltheart, 1993; Drewnowski, 1980). Fallon et al. (1999) suggested that the former type of phonological similarity (rhyming similarity) can act as an effective category cue, and therefore facilitates item recall; but that the latter type of phonological similarity (phonological overlap without rhyme) does not provide an effective category cue, and therefore does not facilitate item recall.

Fallon et al. (1999) tested these hypotheses by examining immediate serial recall of rhyming lists, phonologically similar but nonrhyming lists, and phonologically dissimilar lists. In their first experiment, they found that item recall was indeed greater for rhyming than for phonologically dissimilar lists, consistent with the hypothesis that rhyme similarity can produce a category cuing effect. Item recall for phonologically dissimilar lists was greater than for the similar nonrhyming lists, consistent with the hypothesis that a category cuing effect is obtained only with rhyme similarity. Stric serial recall was equivalent for the similar rhyming lists and the dissimilar lists. However, strict serial recall was higher for the dissimilar lists than for the similar nonrhyming lists, a replication of the classic PSE. These results were consistent with the hypothesis that for similar rhyming lists, a detrimental effect of phonological similarity on order recall is offset by the beneficial category cuing effect of rhyme on item recall, leading to strict serial recall that is equivalent to that for dissimilar lists. For similar nonrhyming lists, however, there is no beneficial category cuing effect to facilitate item recall, but there still is a detrimental effect of phonological similarity on order information, so that strict serial recall is worse than for dissimilar lists, yielding the classic PSE.

Fallon et al. (1999) also tested the hypothesis that the effectiveness of category cuing for rhyming lists is a function of the uniqueness of the cue. The results just described were from their first experiment, in which list items did not recur across lists – that is, list items were drawn from what may be termed an *open set*. In a second experiment, Fallon et al. (1999) again examined immediate serial recall of dissimilar, similar rhyming, and similar nonrhyming lists, but the lists in each condition were now drawn from a *closed set*. That is, in each condition, lists were drawn from a small set of items, so that list items did recur across lists within a condition. Under these circumstances, item recall for the rhyming lists was equivalent to that for dissimilar lists (rather than greater, as in their first experiment), consistent with the hypothesis that the effectiveness of rhyme as a category cue is affected by the uniqueness of the rhyme. Strict serial recall was greater for dissimilar than for rhyming lists, consistent with the hypothesis that the detrimental effect of phonological similarity on order information was not offset by a beneficial category cuing effect on item recall.

There is at least one important issue that remains unresolved. Fallon et al. (1999) showed that item recall in phonologically similar rhyming lists was greater than in phonologically similar nonrhyming lists, and hypothesized that it is the presence of a rhyme category that is critical. However, the rhyming-nonrhyming manipulation in their experiments was confounded with a difference in the degree of within-list phonological overlap. Each member of similar rhyming lists such as *{mat, fat, sat, rat, hat, bat}* shared two phonemes, whereas each member of similar nonrhyming lists such as *{rat, map, tab, fad, can, gag}* shared on average only one phoneme. The difference in item recall for these two types of lists could therefore have been due to cuing by the degree of phonemic overlap rather than by rhyme category, a possibility that the authors acknowledged in discussing transposition errors (Fallon et al., 1999, p. 303). The question therefore is whether the effects obtained by Fallon et al. (1999) were due to cuing by rhyme category or cuing by the degree of phonemic overlap.

An obvious way to address this would be to compare item recall for lists of rhyming versus non-rhyming stimuli with a controlled degree of phonemic overlap. However, the question makes contact with issues of considerably broader significance than simply controlling for a confound, because different theoretical accounts of the phonological similarity effect make different predictions about the relative effects of rhyme versus phonemic overlap. Let us consider these alternative accounts. In doing so, we will consider their predictions with regard to three types of lists. One type of list is comprised of phonologically similar rhyming stimuli that share a certain degree of phonemic overlap, say two phonemes, and this overlap is in the vowel and final consonant (e.g., *{mat, fat, sat, rat, hat, bat}*). Let us refer to these as *rhyming* lists. A second type of list is comprised of phonologically similar non-rhyming stimuli that share as much phonemic overlap as the rhyming lists, in this case two phonemes, but overlap in the first consonant and phoneme (e.g., *{cat, cab, cad, can, cap}*). Let us refer to these as *alliterative* lists. Comparison of item recall for lists of these two kinds would address the question of whether cuing in the Fallon et al. (1999) study was a result of the degree of phonemic overlap or of the rhyme category. Let us also, however, consider a third type of phonologically similar list, in which the list items share some overlap, but the overlap is not consistently in the same place, and the total overlap within a list is not as much as in the other types of similar lists (e.g., *{cad, cat, map, can, man}*). Let us refer to these as *canonically similar* lists, as acknowledgment that this is the type of similarity utilized in some of the seminal studies that originally demonstrated the classic PSE (e.g., Baddeley, 1966). Let us now consider the predictions that various theoretical accounts would make for item recall of such lists.

One kind of theoretical account that has been proposed is a feature model. In feature models in general, memory traces are represented as vectors of features. The feature model of Nairne (1990; Nairne & Kelley, 1999; Nairne & Neumann, 1993) incorporates representations of this type. In this model, the effect of phonological similarity in serial memory arises from overlap of the feature vectors that represent the phonologically similar list items. Phonological similarity makes it difficult to recover an item's correct position within a list because there are overlapping features; however, common phonological features among list items can be used to discriminate the list as a whole from other lists, thus aiding item recall (e.g., Nairne & Kelley, 1999, p. 49). This latter aspect constitutes a means for phonological commonality to act as what Fallon et al. (1999) termed a category cue. Implicit in the feature account is the notion that it is the degree of overlap that matters for these effects. The feature model would therefore predict that item recall for lists comprised of phonologically similar rhyming stimuli should be equivalent to that for lists comprised of phonologically similar non-rhyming stimuli such as alliterative lists, if the degree of phonological overlap is controlled. In addition, it would predict that, based on the degree of phonological overlap, item recall should be greater for alliterative than for canonically similar lists. We will refer to this as the *feature account*.

An alternative account is derived from linguistic theory. According to linguistic analysis, a syllable has two constituents, an *onset* and a *rime*. The onset contains any consonants that precede the vowel, e.g., *c* in *cat*. The rime contains the vowel and any following consonants, e.g., *at* in *cat*[1]. Thus, according to linguistic theory, for a consonant-vowel-consonant (CVC) syllable, the combination of the second and third segments corresponds to a theoretically defined linguistic category (the rime), whereas the combination of the first and second segments does not correspond to a linguistic category. An account of phonological similarity effects based on this linguistic analysis would view a two-phoneme overlap between the rimes of words in a list as being quite different in nature from a two-phoneme overlap between the first two segments of the words; and would predict that the potential for a category cuing effect should only arise in the case of rime overlap, which corresponds to linguistic category overlap (e.g., Nimmo & Roodenrys, 2004). This predicts that item recall for lists comprised of phonologically similar rhyming stimuli should be better than for lists comprised of phonologically similar non-rhyming stimuli such as alliterative lists, even if the degree of phonological overlap is controlled. Further, it predicts no difference in item recall between alliterative and canonically similar lists, because in neither case is there systematic linguistic category overlap between list items. We will refer to this as the *linguistic structure account*.

A third account is derived from consideration of the computational requirements of a processing system that performs list recall in the verbal domain (e.g., Gupta, 1996; Gupta & MacWhinney, 1997). One point highlighted by computational analysis is that list recall requires the processing of serial order at both the level of lists (i.e., the serial order of items within lists), and at the level of words (i.e., the serial order of the constituents of individual words, which are, after all, phonological sequences). This raises the question of what effect, if any, the serial ordering within words might be expected to have on serial recall of the lists containing them. In a theory in which the serial order of phonemes *within* words is explicitly represented (as in the Gupta & MacWhinney, 1997 account), there is a basis for considering how serial order at this level might affect serial ordering at the next level up. How might it play a role? One suggestion comes from work by Gupta and Dell (1999), who noted that a sequence of words such as *{cat, cab}* that share overlap at the beginning is more difficult to produce than a sequence of words such as*{cat, mat}* that share overlap at the end, as shown by Sevald and Dell (1994). Following this earlier work, Gupta and Dell (1999) suggested that this is because of the serial order of phonemes within word forms. The idea is that words can be thought of as dynamic trajectories over time, in phonological space. Trajectories that start similarly but end differently (i.e., words that share overlap at the beginning but not at the end) are more difficult to discriminate than trajectories that start differently but end similarly (i.e., words that share overlap at the end but not at the beginning). Applying these ideas to the current issue, this suggests that cuing effects in list recall should be sensitive, not merely to the degree of overlap between list items, but to the serial position of the overlap. This predicts that there will be a difference between recall of lists whose items share overlap at their beginings versus at their ends. Specifically, it predicts that item recall for lists comprised of phonologically similar rhyming stimuli (in which the overlap is at the ends of the words) should be better than that for lists comprised of phonologically similar non-rhyming stimuli such as alliterative lists in which the overlap is at the beginnings of the words, even if the degree of phonological overlap is controlled. It also predicts that item recall will be greater for alliterative than for canonically similar lists, because of greater phonemic overlap in the former. We will refer to this as the *serial order account*. It is worth noting that the serial order account incorporates important elements of the feature account. In particular, the degree of feature overlap does form the basis of cuing effects, just as in the feature account. However, the serial order account additionally posits that the location of the overlap matters, thus leading to different predictions regarding item recall for rhyming versus alliterative lists. Thus the serial order account can be seen as building on and extending the feature account; the proposal that the serial order *within* list items is relevant is, nevertheless, an important difference.

Thus the three accounts differ in their overall sets of predictions. The item recall predictions of the feature account are $Rhyming = Alliterative > Canonical$. The item recall predictions of the linguistic structure account are $Rhyming > Alliterative = Canonical$. The item recall predictions of the serial order account are $Rhyming > Alliterative > Canonical$.

The first goal of the present work was to test these predictions and thus discriminate between these theoretical accounts, by comparing item recall in immediate serial recall of rhyming and alliterative lists with the same degree of overlap, as well as immediate serial recall of canonically similar lists, with lesser overlap; Experiments 1 and 2 addressed this goal. A second goal was to test the hypothesis proposed by Fallon et al. (1999) and Nairne and Kelley (1999) that a category cue will be more effective when using an open rather than a closed set; this aim was addressed by Experiments 3 and 4. A third goal, addressed in Experiment 5, was to examine whether item recall could be boosted sufficiently to lead to greater strict serial recall for similar than dissimilar lists. A fourth overarching goal was to articulate a theoretical framework that can serve to interpret and integrate the varied findings regarding the PSE in immediate serial recall into a coherent and unified account; in the General Discussion, we offer such a framework, and attempt to show that it can relate various findings to each other.

Experiment 1

Experiment 1 aimed to directly test the predictions of the three alternative accounts outlined above. To this end, participants were presented with five-item lists for immediate serial recall in each of four conditions: a rhyming condition, an alliterative condition, a canonically similar condition, and a phonologically dissimilar condition. The main question of interest was the pattern of relationship between item recall in the rhyming, alliterative, and canonically similar conditions, regarding which the theoretical accounts make differing predictions. The phonologically dissimilar condition was included as a baseline to investigate the presence of a classic PSE. Lists in each condition were drawn from an open set of words, so that no words were repeated across lists.

Method

Participants

Participants in this and all other experiments reported here were undergraduate students at the University of Iowa, who participated for course credit. A total of 24 participants engaged in Experiment 1. No participant took part in more than one of the present experiments.

Materials and Design

Each list consisted of five three-phoneme words presented auditorily. No word appeared more than once in a list; thus, the lists in each condition (and in fact for the experiment as a whole) were drawn from an open set. Ten five-item lists were created for each of the rhyming, alliterative, canonically similar and phonologically dissimilar conditions. Two additional practice lists were created for each condition. The mean frequency of list items was controlled so that none of the conditions differed significantly from any other in mean Kucera-Francis frequency of the list stimuli (Kucera & Francis, 1967). The lists used are shown in Appendix A.

For each of the canonically similar and phonologically dissimilar conditions, ten lists were generated by random selection without replacement from the set of 50 phonologically similar and 50 phonologically dissimilar words used by Coltheart (1993). Rhyming and alliterative lists were created individually. Rhyming lists were constructed through use of several rhyming dictionaries (Bogus, 1991; Merriam-Webster, 1987, 1995; Mitchell, 1996). A total of 10 differently rhyming 5- item lists were created. All words in a rhyming list differed phonologically only in their initial phoneme, and orthographically only in their first grapheme. Alliterative lists were generated with the assistance of an online dictionary (Merriam-Webster, 2001). All items within a list were matched on initial consonant and following vowel. Stimuli within a list were also matched orthographically, differing only in their post-vocalic graphemes.

Each word to be used in the lists was spoken by a female native speaker of American English and recorded as 16-bit digitized sound at a sampling rate of 44.1 kHz on a Macintosh computer using the SoundEdit software program produced by Macromedia, Inc (SoundEdit 16 Users Guide [Computer software], 1997).

Procedure

Several aspects of the procedure were common to all the experiments reported here. In Experiment 1 and in all other experiments, each participant performed immediate serial recall under each of four conditions. Lists were presented at the rate of one word per second by a Macintosh computer running the Psyscope experiment control system (Cohen, MacWhinney, Flatt, & Provost, 1993). At the end of each list presentation, a cross appeared on the display, at which time the participant was required to recall the list. Each response was scored for the number of items correctly recalled (item recall) and for the number of items correctly recalled in the correct serial position (strict serial recall).

In Experiment 1, list presentation was blocked by condition. The order of presentation of conditions was counterbalanced across participants. The lists in each condition were presented in random order. One trial consisted of auditory presentation of one of the lists. Recall was performed by writing on an answer sheet. The answer sheet contained rows of five underlines, each row corresponding to a list of five items. The participant's task was to write down the words that had appeared in the list, indicating their order by writing each word in the correct underline on the row.

Results and Discussion

Results are summarized in Table 1. Analyses were conducted both by subjects (i.e., with subjects as the random variable) and by items (with lists as the random variable). For conciseness, only the results of the analysis by subjects are shown in Table 1. However, all results discussed below were obtained in both the subject and item analyses, except where otherwise noted.

As shown in Table 1, the main effect of similarity condition on item recall was significant in an ANOVA. A Tukey test for pairwise comparisons between means using a family confidence coefficient of 0.95 indicated a significant difference between every pair of conditions except the dissimilar and alliterative. Thus, as Table 1 shows, item recall was higher in the rhyming condition than in the alliterative condition, which was higher than the canonically similar condition.

[Table 1 about here.]

In terms of strict serial recall also (Table 1), the main effect of similarity was significant. A Tukey test conducted as previously described indicated that in the subject analysis, all pairwise differences were significant, except those between rhyming and dissimilar lists, and between alliterative and canonically similar lists. The Tukey test on the item analyses yielded the same result except that additionally, the difference between dissimilar and alliterative lists was not significant. The reader may find it useful to consult the relevant parts of Table 3, which summarizes item recall and strict serial recall results in words.

Of primary interest for present purposes, the item recall results are $Rhyming > Alliterative > Canonical$. This is in accordance with the predictions of the serial order account, but not the other two accounts, providing preliminary support for the serial order account. Before discussing the implications of this finding, let us consider the strict serial recall results. These are also interesting in that there is a classic PSE ($Dissimilar > Canonical$), but no detrimental effect of rhyming similarity ($Dissimilar = Rhyming$). This replicates the pattern of results in Fallon et al. (1999, Experiment 1). Taken together, the item recall and strict serial recall results suggest that different kinds of similarity have different effects on item recall: as compared with dissimilar lists, canonical similarity is detrimental, whereas rhyming similarity is facilitatory and alliterative similarity no worse. However, despite rhyming similarity being facilitatory at the item level, strict serial recall for rhyming lists is only equivalent to (not higher than) that for dissimilar lists. Despite alliterative similarity being equivalent at the item level, strict serial recall for alliterative lists is lower than that for dissimilar lists. These findings are consistent with the notion that phonological similarity, including rhyming and alliterative similarity, has a detrimental effect on retention of order information, which may be offset if there is facilitation at the item level.

This can be further examined in terms of order accuracy, measured as the number of items that were recalled in correct serial position divided by the number of items that were correctly recalled in total; it is computed as the ratio of strict serial recall to item recall (Fallon et al., 1999). An ANOVA indicated a significant effect of phonological similarity on order accuracy (Table 1). A Tukey test indicated that order accuracy was significantly higher for dissimilar than for rhyming and alliterative lists, and significantly higher for canonically similar than alliterative lists. Thus for rhyming lists, the benefit in item recall relative to dissimilar lists offsets the decrement in order accuracy relative to dissimilar lists, leading to equivalent strict serial recall for rhyming and dissimilar lists. For alliterative lists as compared to dissimilar lists, the equivalent item recall does not offset the decrement in order accuracy, and so strict serial recall is lower for the alliterative than dissimilar lists. For canonically similar lists as compared to dissimilar lists, the (non-significant) decrement in order accuracy is compounded by a significant decrement in item recall (there is no facilitatory category cuing effect), leading to lower strict serial recall. The results suggest that performance in immediate serial recall, as measured by strict serial recall, is a function of how well the identity of the items is recalled, and how well the order of the items is recalled, and that these can trade off against each other, an idea that has been present in the writing of several investigators (e.g., Drewnowski, 1980; Fallon et al., 1999; Nairne & Neumann, 1993; Nairne & Kelley, 1999; Poirier & Saint-Aubin, 1996; Wickelgren, 1965).

The measure of strict serial recall described above was based on items recalled in correct serial position. However, another possible measure of strict serial recall is based on the number or proportion of lists correct. The list-based measure entails a binary scoring, in that a list is either correct or incorrect, based on whether all its items were recalled in correct serial position; a partially correct list is scored as completely incorrect. This measure is quite common, and has been used by a number of investigators, including some of the seminal studies that established the classic PSE, as well as in very recent studies (e.g., Baddeley, 1966; Service & Maury, 2003).

Recall was therefore also assessed in terms of this measure, primarily to verify that the classic PSE ($Dissimilar > Canonical$) was in fact obtained using this classic measure. As shown in Table 1, a planned comparison indicated that the classic PSE was significant using this list recall measure. Of lesser interest, the ANOVA on list recall also indicated an overall effect of similarity.

Let us return to the item recall results as they relate to distinguishing between the feature, linguistic structure, and serial order accounts. The significantly higher item recall for rhyming over alliterative lists indicates that facilitation at the item recall level is not simply a matter of extent of phonemic overlap, consistent with the prediction of the serial order account, but not the feature account. The significantly higher item recall for alliterative over canonically similar lists suggests that facilitation at the item recall level is not simply a matter of the presence or absence of linguistic category overlap, consistent with the prediction of the serial order account, but not the linguistic account. Overall, these results provide clear support for the serial order account over the other two accounts.

Additionally, these results serve to clarify Fallon et al.'s (1999) finding of greater item recall for rhyming than for nonrhyming similar lists (which in our terms were canonically similar). As we noted earlier, in that study, the difference between rhyming and nonrhyming similar lists was confounded with the degree of overlap between items within those lists; the result could therefore have been due to cuing either by rhyme category or by degree of feature overlap. The present results indicate that it was neither rhyme alone that was critical (in effect, the linguistic structure account), nor feature overlap alone (in effect, the feature account). If presence or absence of a rhyme category alone had been the critical factor underlying Fallon et al.'s (1999) findings, then in the present experiment, item recall should have been equivalent for the alliterative and canonically similar lists, upholding the linguistic structure account. If degree of phonemic overlap alone had been the critical factor underlying Fallon et al.'s (1999) findings, then in the present experiment, item recall in the present experiment should have been equivalent for rhyming and alliterative lists, upholding the feature account. Neither of these was true in the present experiment; rather, the serial order account was supported. This in turn suggests that

Fallon et al.'s (1999) results arose from the fact that their rhyming lists incorporated greater feature overlap than the similar nonrhyming lists, *and* from the within-word serial location of that overlap.

The present results thus speak to the theoretical alternatives as well as to Fallon et al.'s (1999) previous results. But how robust are the present findings? The distinction between rhyming, alliterative, and canonically similar lists has not previously been examined directly; it would therefore seem appropriate to investigate its replicability. Additionally, the present results were obtained under blocked presentation, and could conceivably have been an artifact of blocking. Experiment 2 aimed to address these issues.

Experiment 2

Experiment 2 examined whether the pattern of results obtained in Experiment 1 would replicate under interleaved rather than blocked presentation of conditions. In Experiment 2, as in Experiment 1, a total of 24 participants each performed immediate serial recall of auditorily presented lists in a rhyming condition, an alliterative condition, a canonically similar condition, and a phonologically dissimilar condition. The materials, design, and procedure were exactly as in Experiment 1, except that presentation of lists was not blocked by condition.

Results and Discussion

All results discussed below were obtained in both the subject and item analyses. As shown in Table 1, the effect of phonological similarity on item recall was significant. Exactly as in Experiment 1, the Tukey test indicated a significant difference between every pair of conditions except the dissimilar and alliterative conditions. Thus, as in Experiment 1, item recall was higher in the rhyming condition than in the alliterative condition, which was higher than the canonically similar condition, but did not differ from the dissimilar condition.

For strict serial recall also, phonological similarity had a significant effect. The Tukey test by subjects indicated exactly the same pattern of results as in Experiment 1: all pairwise differences were significant, except those between rhyming and dissimilar lists, and between alliterative and canonically similar lists. The Tukey test by items yielded the same result. Phonological similarity also had a significant effect on order accuracy, while the Tukey test indicated that order accuracy was significantly higher for dissimilar than for rhyming and alliterative lists, and also significantly higher for canonically similar than for alliterative lists (all exactly as in Experiment 1). Finally, a planned comparison on list recall indicated that the classic PSE was significant, and the ANOVA indicated an overall effect of similarity.

To summarize, the item recall results were $Rhyming > Alliterative > Canonical$ in both subject and item analyses, as in Experiment 1, replicating the critical result that distinguishes between alternative accounts. Overall, the pattern of item recall, strict serial recall, and order accuracy results was identical to that in Experiment 1 in the analysis by subjects, and the pattern of results in the item analysis was identical to that in the subject analysis.

Overall, three conclusions can be drawn from Experiment 2. First, the results confirm the finding of $Rhyming > Alliterative > Canonical$ for item recall, supporting the serial order account. Second, the close replication of Experiment 1 indicates the robustness of this finding. Third, the present results indicate that blocking of similarity conditions does not greatly affect the results that are obtained; blocking therefore does not appear to be a major factor in determining the effects of phonological similarity in immediate serial recall.

Experiment 3

The results of Fallon et al. (1999, Experiment 2) discussed previously suggest that the effectiveness of a category cue in facilitating item recall of a list is determined by its uniqueness, i.e., by whether the cue characterizes only one or many lists in the stimulus set, and that such uniqueness is dependent on whether the lists are drawn from an open or closed set of items (Fallon et al., 1999; Nairne & Kelley, 1999). If this is correct, then the item recall advantage for rhyming over dissimilar lists, and for alliterative over canonically similar lists obtained in Experiments 1 and 2 should be reduced or eliminated when the lists are drawn from a closed set of stimuli. Experiment 3 tested this prediction.

A total of 24 participants were each auditorily presented with five-item lists, twelve in each of the rhyming, alliterative, canonically similar and phonologically dissimilar conditions. Each list consisted of five three-phoneme words. The stimulus lists for each condition were drawn randomly without replacement from a closed set or pool of items. The pool for the rhyming lists consisted of the words *set, bet, yet, wet, met, get, let, net*. The pool for the alliterative lists was *bud, but, buff, bug, buzz, buff, bun, buck*. The pools for the canonically similar and dissimilar lists were those used by Baddeley (1966), which were also used by Fallon et al. (1999) for their closed sets; these were *can, mad, cap, man, cad, cat, map, mat* and *cow, day, bar, few, hot, pit, pen, sup* respectively. The mean Kucera-Francis frequency of items did not differ across the pools used for each condition. As in Experiment 1, lists were blocked by condition, and the order of presentation of conditions was counterbalanced across participants. Participants were not informed that lists would be drawn from a small set of words nor were they pre-familiarized with these words.

Results and Discussion

All results discussed below were obtained in both the subject and item analyses. As shown in Table 1, the effect of similarity on item recall was significant. The Tukey test indicated a significant difference between every pair of conditions except the alliterative and canonically similar conditions. Item recall was highest in the dissimilar condition, which was higher than the rhyming condition, which was higher than the alliterative condition, which did not differ from the canonically similar condition. The effect of similarity on strict serial recall was also significant. The Tukey test indicated that all pairwise differences were significant, except between the alliterative and canonically similar lists. Similarity also had a significant effect on order accuracy. The Tukey test indicated that all pairwise differences were significant, except between the alliterative and canonically similar lists, and the canonically similar and rhyming lists. Finally, for list recall, a planned comparison indicated that the classic PSE was significant and an ANOVA indicated an overall effect of similarity.

The focus of interest is the item recall findings for rhyming and alliterative lists as compared with canonically similar and dissimilar lists. In contrast with Experiments 1 and 2, which used open sets, there was no item recall advantage for alliterative over canonically similar lists. Also in contrast with Experiments 1 and 2, there was no advantage for rhyming lists over dissimilar lists, and in fact item recall was significantly greater for dissimilar than for all other types of lists. This provides support for the theoretical view that category cuing effects are reduced or eliminated when lists are drawn from a closed set.

However, as this is the first investigation of item recall for rhyming, alliterative, and dissimilar lists using closed sets, it seemed appropriate to investigate the replicability of this result. Additionally, as in the extension of Experiment 1 by Experiment 2, we wished to investigate the role that blocked presentation might have played in obtaining these results. Experiment 4 aimed to address these issues.

Experiment 4

Experiment 4 examined whether the pattern of results obtained in Experiment 3 would replicate under unblocked presentation of conditions. A total of 24 participants each performed immediate serial recall of auditorily presented lists under the same four conditions. The materials, design, and procedure were exactly as in Experiment 3, except that presentation of lists was not blocked by condition.

Results and Discussion

All results discussed below were obtained in both the subject and item analyses except where otherwise noted. For item recall, the ANOVA indicated a significant effect of similarity (Table 1). The Tukey test indicated exactly the same pattern of results as in Experiment 3: a significant difference between every pair of conditions except the alliterative and canonically similar conditions, with item recall being highest in the dissimilar condition, which was higher than the rhyming condition, which was higher than the alliterative condition, which did not differ from the canonically similar condition. A significant effect of similarity on strict serial recall was also obtained. A Tukey test on the subjects analysis indicated that all pairwise differences were significant, and a Tukey test on the items analysis indicated the same result except that the difference between canonically similar and alliterative lists was not significant. The effect of similarity on order accuracy was significant, and the Tukey test indicated that all pairwise differences were significant, except between the rhyming and canonically similar lists. Finally, for list recall, a planned comparison indicated that the classic PSE was significant and an ANOVA indicated an overall effect of similarity.

In summary, and of primary interest for present purposes, the item recall results exactly replicated those of Experiment 3. In addition, the strict serial recall results and order accuracy results were also closely similar to those obtained in Experiment 3. Overall, we can draw two conclusions from Experiment 4. First, the results confirm the critical result from Experiment 3, namely, a reduction of item recall advantage when using closed sets: as in Experiment 3, the item recall advantage was eliminated for alliterative over canonically similar lists, and also for rhyming lists over dissimilar lists. Second, as with Experiment 2, the present results suggest that blocking is not a critical determinant of the effects of phonological similarity on immediate serial recall.

Experiment 5

As noted in discussion of Experiment 1, it appears that overall performance in immediate serial recall, as measured by strict serial recall, is a function of how well the identity of the items is recalled, and how well the order of the items is recalled. That is, strict serial recall will be determined by the tradeoff between the possibly facilitatory effects of any category cues on item recall, and the detrimental effects on order recall of any factor such as phonological similarity that increases within-list confusability. This implies that phonological similarity may simultaneously play a detrimental role in retention of order information (as a result of similarity-based interference), and a possibly beneficial role in retention of item identity information (if it forms an effective category cue). This raises the possibility of the facilitatory effect of phonological similarity on item recall being strong enough to outweigh any detrimental effect of phonological similarity on order accuracy, thereby leading to *better strict serial recall* for phonologically similar than for dissimilar lists.

What evidence speaks to this prediction? Previous research has demonstrated beneficial effects of phonological similarity on item recall (as discussed extensively in this article, and as also demonstrated by the present Experiments 1 and 2). The lack of a detrimental effect of phonological similarity on strict serial recall of word lists has also been demonstrated where similarity was operationalized as rhyme (Fallon et al., 1999, Experiment 1; Fallon, Mak, Tehan, & Daly, in press; the present Experiments 1 and 2), and as overlap of the phonemes surrounding the vowel ("consonant frame overlap"; Lian, Karlsen, & Eriksen, 2004, Experiment 2). A beneficial effect of phonological similarity has been demonstrated in an order reconstruction task following a 24-second delay (Nairne & Kelley, 1999), and in non-immediate serial recall following a delay of 4 seconds (Fallon & Tehan, 1995; Fallon, 1999). Recently, Lambert, Chang, and Lin (2003) reported a beneficial effect of phonological similarity on pharmacists' immediate free recall of drug names. Beneficial effects of similarity have also recently been reported for strict serial recall in immediate serial recall of nonword lists, where similarity was operationalized as rhyme (Service & Maury, 2003; Lian & Karlsen, 2004) or as consonant frame overlap (Lian et al., 2004, Experiment 1). To our knowledge, however, there has been no previous demonstration of a beneficial effect of phonological similarity on strict serial recall in a typical immediate (i.e., non-delayed) serial recall task employing known words. Such a result would constitute the most direct reversal possible of the canonical PSE. Experiment 5 investigated whether such a reversal could be obtained.

The goal was to achieve a beneficial category cuing effect of phonological similarity on item recall that would outweigh any detrimental effect of the phonological similarity on order recall. In determining how to achieve this, we were guided by the results of Experiments 1 and 2, which had indicated stronger category cuing for rhyme than alliteration, and Experiments 3 and 4, which had indicated stronger category cuing with open than with closed sets. For Experiment 5, therefore, phonological similarity was operationalized as rhyme, and we used open sets. In addition, we reasoned that the effect of rhyme as a category cue should be even stronger for lists of two-syllable rhyming words, in which the extent of overlap is even greater than in lists of one-syllable rhyming words, and therefore decided to include such lists as well. Finally, to verify that the item recall advantage obtained in Experiments 1 and 2 for rhyming over dissimilar lists would be maintained in an alternative presentation modality, we decided to employ visual presentation. Experiment 5 therefore examined immediate serial recall of one-syllable rhyming and dissimilar lists as well as two-syllable rhyming and dissimilar lists, using visual presentation of lists drawn from open sets.

Method

A 2-factor within-subjects design was used, with the two factors being phonological similarity (dissimilar/rhyming) and word length (one/two syllables) and thus with four conditions defined by the crossing of the two factors. Lists were created using the rhyming and other dictionaries previously noted, and the definition of each word in the lists was verified from a standard dictionary of American English (Houghton-Mifflin, 1996). Sets of four lists were created, with the four lists in a set belonging to each of the four conditions, and with all lists in a set being matched in frequency. Ten such sets of four lists were created, so that ten four-item lists were created for each of the conditions, with mean frequency of words in the lists controlled across conditions. The lists used are shown in Appendix B. The list shown in position n in each condition was matched in frequency with the lists in position n across all conditions. A total of 24 participants each performed immediate serial recall of visually presented lists under each of the four conditions, with the order of presentation of the 40 lists randomized across the experiment, and not blocked by condition. Participants' spoken responses were audiotaped for subsequent analysis, in which each response was scored for item recall and strict serial recall.

Results and Discussion

Analyses were conducted both by subjects and by items. All results discussed below were obtained in both the subject and item analyses except where otherwise noted. Results from the subjects analysis are summarized in Table 2. Planned pairwise comparisons were conducted between dissimilar and similar one-syllable lists and between dissimilar and similar two-syllable lists. For item recall, these comparisons indicated significant differences. That is, item recall was significantly greater for similar than dissimilar lists at both word lengths. Of less direct interest, an omnibus 2 x 2 (similarity x word length) ANOVA indicated significant main effects of phonological similarity and word length, and a significant interaction, as also shown in Table 2. For strict serial recall, the comparisons revealed that strict serial recall was higher for similar than dissimilar two-syllable lists but did not differ significantly for similar and dissimilar one-syllable lists. In the omnibus ANOVA, both main effects and the interaction were significant. For order accuracy, the comparisons revealed that order accuracy was higher for dissimilar than similar one-syllable lists but did not differ significantly for similar and dissimilar two-syllable lists. In the omnibus ANOVA, the main effect of word length was significant, but not the main effect of similarity or the interaction. The null effect of similarity on order accuracy in the two-syllable lists is to our knowledge a novel result. It should be noted, however, that the PSE has not been extensively investigated in polysyllabic word lists.

[Table 2 about here.]

These results confirm the possibility that Experiment 5 was designed to examine, namely, that strict serial recall can be higher for lists of similar (in this case rhyming) words than for lists of dissimilar words. At the two-syllable word length item recall was significantly higher for similar than dissimilar lists, and this benefit, combined with a non-significant order accuracy difference between dissimilar and similar lists, led to higher strict serial recall for the similar than dissimilar lists. At the one-syllable word length also, item recall was significantly higher for similar than dissimilar lists, but this advantage was not large enough to offset the significantly lower order accuracy for similar than dissimilar lists, and strict serial recall was therefore no better (nor worse) for similar than for dissimilar lists.

The list recall measure, however, indicated that strict serial recall was significantly higher for similar than dissimilar lists at both the one-syllable word length and the two-syllable word length. That is, when using the list-based measure of strict serial recall that has been widely employed in studies of the PSE, recall was significantly higher for similar than dissimilar lists not only at the two-syllable word length, but even for one-syllable lists. (Of lesser interest, the omnibus ANOVA indicated that both main effects were significant in both subject and item analyses, and that the interaction was significant in the subject analysis, and marginally significant in the item analyses).

The results of Experiment 5 thus provide clear evidence that phonological similarity in the form of rhyme can facilitate not merely item recall, but even strict serial recall, even for lists of one-syllable words. This is to our knowledge the first demonstration of a beneficial effect of phonological similarity on strict serial recall in a typical immediate serial recall task employing known words, and constitutes the most direct reversal possible of the canonical PSE.

What are the implications of this finding? In our view, the present result is not inconsistent with the account of phonological similarity given within the working memory model (e.g., Baddeley & Hitch, 1974; Baddeley, 1986; Baddeley et al., 1998). According to the working memory model, phonological similarity affects immediate serial recall, thus providing evidence that the code used in verbal short-term memory is phonological in nature. The present results suggest that the effect of phonological similarity need not always be detrimental, as originally proposed; but they support the conclusion that phonological similarity affects immediate serial recall. The present results are thus best viewed as clarifying and extending the treatment of phonological similarity in the working memory model, without contradicting its central conclusion.

General Discussion

In this article, we presented five experiments examining item recall in phonologically similar lists. Experiments 1 and 2 tested alternative accounts of what kind of phonological similarity leads to a category cuing effect that can facilitate item recall. The alternative accounts were designated as the feature account, the linguistic structure account, and the serial order account. The results of Experiment 1 supported the serial order account over the other two accounts, and Experiment 2 confirmed that this result did not depend on blocked presentation of lists.

The results of Experiments 1 and 2 thus support the serial order account in a replicable fashion. This indicates that within-word-form serial order is an important factor to keep in mind when considering the serial ordering of lists of word forms – that is, when considering typical immediate serial recall tasks. In our view, this is an important elaboration of what a theory of immediate serial recall must take into account. It also provides support for models of immediate serial recall that attempt to incorporate serial ordering not only across word forms (i.e., at the within-list level) but also at the within-word-form level; the only such implemented model of which we are aware is that developed in our own previous work (e.g., Gupta, 1996; Gupta & MacWhinney, 1997).

Experiments 3 and 4 investigated the hypothesis that the effectiveness of a category cue in facilitating item recall of a list is determined by its uniqueness, which in turn is dependent on whether the lists are drawn from an open or closed set of items. In contrast with Experiments 1 and 2, which used open sets, Experiment 3 used a closed set, and found no item recall advantage for alliterative over canonically similar lists. Also in contrast with Experiments 1 and 2, there was no advantage for rhyming lists over dissimilar lists, and in fact item recall was significantly greater for dissimilar than for all other types of lists. Experiment 4 replicated these results with unblocked presentation of lists.

Finally, Experiment 5 investigated the possibility that phonological similarity can enhance item recall sufficiently to overcome any detrimental effect on order accuracy, thus leading to better *strict serial recall* for similar than for dissimilar lists. This result was indeed obtained for lists of two-syllable words as well as for lists of one-syllable words, providing a clear reversal of the canonical PSE. However, as we noted in discussion of Experiment 5, this result in our view does not invalidate the working memory model's account of the PSE, but rather, serves to clarify and extend that account.

It could be argued that list items may be more perceptually confusable when lists are presented auditorily (as in Experiments 1 through 4) than visually. This raises the possibility that perceptual errors might have played a role in the results obtained in Experiments 1 through 4. One way to gauge perceptual errors is to examine items produced during recall that are perceptually similar to one or more items from the target list, but that did not in fact appear in either the target list or in the set of words from which the lists were drawn in that experiment ("extra-set intrusions"). If perceptual confusability played a role in the auditory experiments, there should be a greater proportion of perceptual errors in those experiments than in the visual experiment. We therefore compared perceptually similar extra-set intrusions in the auditory experiments (Experiments 1 through 4 combined) with those in the visual experiment (Experiment 5), for the two conditions that were common to all the experiments (rhyming and dissimilar lists of one-syllable items). An ANOVA indicated that there was no significant difference between the proportions of such intrusions for the auditory experiments versus the visual experiment $F(1, 118) = .10, p > .7, MSE = 34.03$. It therefore appears unlikely that the effects in Experiments 1 through 4 were driven by auditory perceptual confusability.

Another issue is whether the differences in the patterns of results obtained in Experiments 3 and 4 as compared with Experiments 1 and 2 are truly due to the use of closed versus open sets, or might simply be due to the use of *different* stimulus sets. This is, of course, a question that also applies to other studies that have employed non-identical open versus closed stimulus sets (e.g., Coltheart, 1993; Fallon et al., 1999; Nairne & Kelley, 1999). In the present study, the fact that the item analyses corroborated the subject analyses makes it unlikely that the results were an artifact of the particular stimuli used. However, to further examine this question, we further analyzed error types. Intrusions that are from outside the target list but not from outside the set of words used in the experiment (“extra-list intrusions”) provide a measure of memory errors (the intrusion is a word that was in another list). If the difference in item recall accuracies in the open set experiments (Experiments 1 and 2) and the closed set experiments (Experiment 3 and 4) was in fact due to the openness of sets, then we would expect a greater proportion of extra-list intrusions in the closed set experiments, because of the greater confusability of lists. If, on the other hand, the difference in item recall accuracies for the closed and open set experiments was due simply to the use of different stimuli, there would be no particular reason to expect this difference in extra-list instrusions. A comparison revealed that the proportion of extra-list intrusions in the two closed set experiments combined was indeed significantly higher than in the two open set experiments combined $F(1, 94) = 53.19, p < .0005, MSE = 16.29$.

As a further test, we examined the *patterns* of extra-list intrusions for the open set and closed set experiments. As discussed in the individual presentation of Experiments 3 and 4, the key item recall differences for the open versus closed set experiments were that there was an item recall advantage for alliterative over canonically similar lists in the open set experiments but not in the closed set experiments; and that an advantage for rhyming over dissimilar lists in the open set experiments was replaced by an advantage for dissimilar over all other types of lists in the closed set experiments. If these item recall differences had been driven by the openness of the sets, then the pattern of differences in extra-list intrusions between the open and closed set experiments should be the same as the pattern of item recall differences. If, on the other hand, the item recall differences between the open and closed set experiments had been driven simply by stimulus differences, there would be no reason to expect the pattern of extra-list intrusions to correspond in this manner. Tukey tests examining pairwise differences in extra-list intrusions for various similarity types for the open set experiments combined and for the closed set experiments combined yielded exactly the same pattern of differences as noted above for item recall. (ANOVAs revealed a significant overall effect of similarity on extra-list intrusions for the open set experiments combined $F(3, 141) = 141.33, p < .0005, MSE = 12.58$ and for the closed set experiments combined $F(3, 141) = 63.85, p < .0005, MSE = 25.67$). This close correspondence provides evidence that the item recall differences between the open and closed set experiments were due to set openness, as gauged by the extra-list intrusions, rather than stimulus differences.

These various analyses thus suggest that the results of Experiments 1 through 4 are robust, and not due simply to differences in the stimulus sets or to the auditory modality of presentation. But what is the bigger picture that can be drawn from these results? As we noted in the introduction, the literature on the PSE presents a conflicting body of results, and the present experiments add to this diverse set of findings. However, we suggest that the results of these various studies can be related to each other within a single theoretical framework. The framework we propose consists of seven points, drawing importantly on ideas that have been put forth in various forms by a number of investigators (e.g., Drewnowski, 1980; Fallon et al., 1999; Nairne & Neumann, 1993; Nairne & Kelley, 1999; Poirier & Saint-Aubin, 1996; Wickelgren, 1965), but that have not to our knowledge previously been brought toghether.

(I) Overall performance in immediate serial recall (i.e., strict serial recall) is a function of how well the *identity of the items* themselves is remembered (i.e., item recall), and of how well the *serial order of the items* is remembered (i.e., order accuracy).

(II) Two factors (among others) that affect immediate serial recall are within-list phonological similarity, and category cues. Category cues are commonalities between the items in a list that can be extracted and used as a cue (e.g., “all the numbers are multiples of 11”, for the list *{11, 55, 77, 22, 88}*).

(III) Phonological similarity has an effect on serial order for items (i.e., order accuracy), and this is a detrimental effect; phonological similarity therefore contributes negatively to overall serial recall performance.

(IV) A category cue has its effect on retention of item identity information (i.e., item recall), and this can be facilitatory, and can therefore contribute positively to overall serial recall performance.

(V) Phonological similarity and category cues thus impact overall immediate serial recall performance in different ways. Strict serial recall in a given situation will therefore be affected by the tradeoff between the effects of category cues and phonological similarity, which have their primary impact respectively on item recall and order accuracy.

(VI) In addition to its detrimental effect on order accuracy, phonological similarity in the form of rhyme can also play a role as a category cue (e.g., “all the words rhyme with *bat*”, for the list *{mat, fat, sat, rat, hat}*). Phonological similarity may therefore at once play a detrimental role in retention of order information (as a result of similarity-based interference), and a possibly beneficial role in retention of item identity information (if it forms an effective category cue).

(VII) The effectiveness of a potential category cue is affected by how many lists it characterizes in the stimulus set. For instance, the “multiples of 11” cue would be maximally useful as a cue to retaining item identity information if it characterized the items in only one list, and less useful if it characterized all lists; and similarly for a cue such as “all the words rhyme with *bat*”. Thus a category cue is generally more effective if the list items are drawn from an open set (where it is possible to have different cues characterizing different lists) than if they are drawn from a closed set (where the cue will necessarily characterize many lists).

How do the present experiments relate to this framework? The three alternative accounts we examined in Experiments 1 and 2 can be seen as proposing hypotheses about *what kind* of phonological similarity leads to a category cuing effect that can facilitate item recall, and thus as fleshing out the sixth point of the framework. Experiments 3 and 4 can be seen as having addressed the seventh point of our framework regarding the greater effectiveness of a category cue for item recall when using an open rather than a closed set. Experiment 5 can be seen as having addressed an implication of the first and sixth points of the framework: the possibility that the facilitatory effect of phonological similarity on item recall may be strong enough to outweigh any detrimental effect of phonological similarity on order accuracy, thereby leading to *better strict serial recall* for phonologically similar than for dissimilar lists. Overall, the results of the present experiments support our proposed framework.

If this framework is generally useful, however, it should also be possible to use it to provide a coherent interpretation of the diverse results that have been reported regarding the PSE. Table 3 summarizes the results of all previously published studies of which we are aware that have examined the effect of phonological similarity in immediate serial recall of real words at both the strict serial recall and item recall levels. Let us interpret each of these in terms of our framework.

[Table 3 about here.]

(1) Wickelgren (1965, Experiment 2) examined immediate serial recall of lists of seven CV items. In five of the six conditions in the experiment, the vowel was the same for all list items; these were referred to as “pure” list conditions. In the sixth “mixed” condition, vowels in each list were mixed. Wickelgren (1965) found that mean recall was greater in the mixed condition than the mean of recall for all the pure lists, when using strict serial scoring. However, item recall was greater for the pure lists than for the mixed lists (Wickelgren, 1965, Experiment 2). In terms of our framework, Wickelgren’s (1965, Experiment 2) pure lists were rhyming lists, while his mixed lists were dissimilar lists. The lists in each condition were drawn from a closed set of items. However, the rhyming lists did not all rhyme in the same way; rather, there were five different kinds of rhyme. Thus, importantly, across the entire corpus of rhyming lists, the set was not completely closed. Wickelgren’s (1965) results are explicable in our framework in terms of rhyme having facilitated item recall by providing a category cue but this facilitation not having been strong enough to overcome the detrimental effect of rhyme on order information, thus yielding a classic PSE in terms of strict serial recall. The item recall result indicates that rhyming similarity can lead to an item recall advantage over dissimilar lists even when using a partially closed set; importantly, however, the set was not completely closed.

(2) In another classic study, Watkins et al. (1974) compared immediate serial recall of phonologically dissimilar and phonologically similar lists. The similar lists were drawn from an open set of items and the dissimilar lists were drawn from a rearrangement of the same set. In terms of our framework, the similar lists were canonically similar, i.e., the items in a list shared

overlap but were neither completely rhyming nor completely alliterative, and therefore did not provide a category cuing advantage for item recall. Consequently, there was no item recall advantage to offset the detrimental effect of canonical similarity on order information, thus yielding a classic PSE in terms of strict serial recall.

(3) In another study, Drewnowski (1980, Experiment 3) compared immediate serial recall of phonologically similar and dissimilar lists. In our terms, the similar lists were rhyming lists; the dissimilar and rhyming lists were each drawn from closed sets. The finding was that both strict serial recall and item recall was better for the dissimilar lists than for the rhyming lists. This is explicable in our framework in terms of rhyming similarity not having provided an item recall cuing advantage because of the use of a closed set; because there was no offset to the detrimental effect of rhyming similarity on order retention, there was a classic PSE in terms of strict serial recall.

(4) Gathercole et al. (1982) compared immediate serial recall of lists that were either phonologically similar or dissimilar, and that were presented either visually or auditorily. The phonologically similar lists consisted mostly of rhyming items (Gathercole et al., 1982, p.177). The dissimilar lists were created by recombining the words in the similar lists. Collapsed across visual and auditory presentation, the results were that strict serial recall was better for the dissimilar than the similar lists, while item recall was better for the similar than for the dissmilar lists (Gathercole et al., 1982, p.180). In our terms, the similar lists were rhyming lists; the dissimilar and similar lists were each drawn from open sets. Rhyming similarity together with the use of an open set provided sufficiently strong category cues that item recall was better than for dissimilar lists, as in Wickelgren’s (1965) study. However, also as in Wickelgren’s (1965) study, this facilitation was not strong enough to overcome the detrimental effect of rhyme on order information, thus yielding a classic PSE in terms of strict serial recall.

(5) Coltheart (1993) compared immediate serial recall of visually presented lists that were either phonologically similar or phonologically dissimilar, and that were drawn from either an open or a closed set. The closed set lists were each drawn from the eight-word pools used by

Baddeley (1966) for his similar and dissimilar lists. The finding was that recall was better for the dissimilar than for the similar lists in terms of both the strict serial and item recall measures, for lists drawn from both open and closed sets. In terms of our framework, the similar lists were canonically similar. Item recall was therefore greater for the dissimilar than canonically similar lists for both the open and closed sets, because canonical similarity did not provide a category cuing advantage for item recall. As there was no item recall advantage to offset the detrimental effect of canonical similarity on order information, there was overall a classic PSE in terms of strict serial recall.

(6) A study by Poirier and Saint-Aubin (1996) examined serial recall of lists of 2-syllable words that were either phonologically similar or dissimilar. Words in the phonologically similar lists rhymed in the second syllable; the specific rhyme changed from list to list. In our terms, the similar lists were rhyming lists; the dissimilar and similar lists were each drawn from open sets. Rhyming similarity did not provide a category cue strong enough to produce an item recall advantage for rhyming over dissimilar lists, despite use of an open set. This contrasts with the results of Wickelgren (1965) and Gathercole et al. (1982) using open sets. It is explicable in that the proportion of overlap, and hence the degree of category cuing, is lower for two-syllable words when only the second syllable rhymes, as in Poirier and Saint-Aubin's (1996) study, than it is when one-syllable words rhyme, as in the studies by Wickelgren (1965) and Gathercole et al. (1982). As there was no item recall advantage in Poirier and Saint-Aubin's (1996) study to offset the detrimental effect of rhyming similarity on order information, there was overall a classic PSE in terms of strict serial recall.

(7) Fallon et al.'s (1999) experiments have already been described in detail. We therefore merely summarize the results in Table 3. These results are completely consistent with our framework.

(8) In the experiments reported in the present article, we see essentially the same effects as in Fallon et al. (1999), as summarized in Table 3. These results are also consistent with our framework.

(9) A study by Lian et al. (2004, Experiment 2) compared immediate serial recall of visually presented lists of Norwegian words that were drawn from open sets and incorporated three types of similarity. Words were either CVCC or CCVCC monosyllables. In rhyming lists, all items shared the middle VC and frequently, additional consonants; in “consonant frame” lists all items shared the consonants surrounding the vowel, and frequently, additional consonants; in dissimilar lists, some overlap was still permitted between list items, such that items within a list frequently shared final consonants (Lian et al., 2004, p. 337). The finding was that item recall was better for consonant frame lists than for rhyming lists, which did not differ from dissimilar lists. In terms of our framework, the lack of item recall advantage for rhyming over dissimilar lists is explicable in that the rhyming lists incorporated less consistent overlap than is typical in other studies discussed so far, in which it has been operationalized as complete _VC overlap in CVC words, and the dissimilar lists shared more overlap than typical in other studies. Further, informal calculation of overlap for the stimuli listed in Lian et al. (2004, p. 337) indicates that words within consonant frame lists shared greater proportional overlap than words within rhyming lists, and this provides a possible explanation of the item recall advantage for the consonant frame lists over the rhyming lists. The strict serial recall finding was that consonant frame lists were recalled as well as dissimilar lists, which were both better than rhyming lists. In terms of our framework, this is because item recall for rhyming lists was no better than for dissimilar lists, and hence did not offset the the detrimental effect of rhyming similarity on order information, leading to worse strict serial recall. The item recall advantage for consonant frame over dissimilar lists, on the other hand, did offset the detrimental effect of consonant frame similarity on order information, leading to equivalent strict serial recall for dissimilar and consonant frame lists, i.e., no classic PSE for consonant frame similarity.

(10) Nimmo and Roodenrys (2004) compared immediate serial recall for phonologically dissimilar and phonologically similar lists with either rhyming (Experiment 1), alliterative (Experiment 2), or consonant frame (Experiment 3) lists. All list items were CVC words. In rhyming lists, all items within a list shared _VC segments, in alliterative lists all list items shared

CV_ segments, and in consonant frame lists, all list items shared C_C segments. The similar lists were in our terms canonically similar, in that they shared some overlap but less so than completely rhyming or completely alliterative or completely consonant frame lists. Lists were drawn from a mostly open set in that list items did not repeat within a condition, but the same list items were used in all three within-subjects conditions. The item recall results as summarized in Table 3 were generally as would be predicted by our framework: item recall was greater for rhyming than for canonically similar lists (Experiment 1), and greater for consonant frame lists than for canonically similar lists (Experiment 3), both of which are explicable in terms of greater overlap for the rhyming and consonant frame lists as compared with the canonically similar lists. However, the lack of item recall advantage for alliterative over canonically similar lists (Experiment 2) differs from the finding in the present Experiments 1 and 2. This is likely because items in the canonically similar lists in Nimmo and Roodenrys (2004, Experiment 2) all shared the final consonant, in addition to sharing some vowels. Thus overlap within these lists was somewhat greater than in our Experiments 1 and 2, and likely created a slight category cuing effect, thus reducing the strength of an item recall advantage for alliterative lists over these lists. Additionally, the fact that lists were not drawn from completely open sets may have played a role in reducing the item recall advantage for alliterative lists.

Turning to an examination of global effects across studies, a glance at the results for strict serial recall in the rightmost column of Table 3 shows that the classic PSE (i.e., better strict serial recall for dissimilar lists) is indeed a robust effect. Only five studies have failed to find it for lists of words – Fallon et al.'s (1999) Experiment 1, our Experiments 1, 2, and 5, and Lian et al.'s (2004) Experiment 2. These were all studies that employed rhyming or consonant frame similarity and open sets, and are interpretable within the framework we have offered. We can thus see that the detrimental effect of phonological similarity on order information is robust, and it is rare for the item recall advantage that may be obtained with phonologically similar stimuli to offset it; hence the classic PSE. Nevertheless, although the classic PSE is a robust effect, it is not exceptionless; and our framework provides a means of understanding when it does and does not arise.

In our view, the present studies provide some important clarifications of the phonological similarity effect, a central phenomenon of verbal working memory. First, in supporting the serial order account (Gupta & Dell, 1999), they emphasize the importance of within-word serial order in theories of immediate serial recall (e.g., Gupta, 1996; Gupta & MacWhinney, 1997). Second, they clarify the basis of the important results obtained by Fallon et al. (1999). Third, they demonstrate the most direct reversal possible of the classic PSE, but also explain this reversal. Fourth, the present studies and our articulation of a theoretical framework offer an integration of the standard working memory model, feature-based accounts, and the importance of within-word serial order in theories of immediate serial recall.

Author Notes

We thank Bruce Lambert for providing interesting discussion of and original motivation for exploring these issues, James Nairne for helpful discussion and clarification of his feature based account of phonological similarity, and Gary Dell for helpful comments. We also thank Ryan Bankson, Sarah Eisenberg, Lindsay Jones, Po-Han Lin, Rebecca Reese, and Jane Wu for assistance in conducting experiments; Rebecca Reese and Tony Buhr for assistance in stimulus creation; Stuart Urban for assistance in creation of experiment-running scripts; and Brandon Abbs, Kate Bullen, Naveen Khetarpal, Ellen Samuel, Kathleen Schnitker, and Jamie Tisdale for assistance in transcription and analysis of participants' responses. This research was supported in part by a University of Iowa CIFRE grant and by grant NIH R01 DC006499 to the first author.

Footnotes

[1]Technically, the rime itself has two constituents: the *nucleus*, which consists of the vowel; and the *coda*, which contains any consonants that follow the vowel.

Appendix A: Lists Used in Experiments 1 and 2

Dissimilar	Canonically Similar
mop zip jaw don bar	cab gab fad gag nan
sup low fed web nut	rat sap pat sax rap
day bus rib log rum	man cat ham cap gap
hot pit cox vat bid	van mad lad had yam
dim nod tic beg jug	tax tag pad can hat
wan win tub few cow	wax lap pan tab mat
cur vet jig rob mud	dab lab dad cad fan
hem fir hub gel his	jam bag nap ram nab
mix dux god rag hap	tap map bat lax sat
joy pen fin sew gal	ban dam jab sad lag

Alliterative	Rhyming
bib big bill bin bit	bale male pale kale sale
did dig dill din dip	lame same game came tame
lead leak lean leap leave	back hack lack pack sack
buck bud budge buff bug	camp lamp ramp samp tamp
nub nudge null nun numb	fain gain main pain rain
pick pig pill pin pip	band hand land rand sand
sake sale same sane safe	bare care dare fare rare
rid rig rill rim rip	face lace mace pace race
tide tight tile time tire	cast fast last past vast
bead beak beam bean beat	date fate hate gate mate

Appendix B: Lists Used in Experiment 5

1-syllable dissimilar	2-syllable dissimilar
bard cube fold leak	basin camel motto zebra
flaw worm sage bead	galley pastel spider helmet
claw dire balk kilt	condor dagger lament parcel
beck fume meed soar	parley stamen cohort foment
clap loft fuse hike	bridle gambit corral deluge
pax tow cob wig	defence mentor pomade bobbin
bast faze leam moat	benign convex levant maggot
jot cub den hap	malice rhesus torpid bantam
cage germ chow grub	domain garment lozenge sabbath
fade clan burn tuck	meadow palette saccade termite

1-syllable rhyming	2-syllable rhyming
hack mack sack tack	cable fable gable sable
lug mug bug hug	jumble mumble rumble tumble
bash mash pash rash	fallow hallow mallow sallow
bock hock nock pock	bangle dangle jangle wangle
kale hale bale vale	barrow harrow marrow yarrow
caw daw haw yaw	giggle jiggle niggle wiggle
gamp samp tamp vamp	fickle pickle sickle tickle
bun hun mun dun	hinder pinder cinder tinder
bin din fin kin	billow killow pillow willow
lawn fawn pawn yawn	basket gasket casket lasket

References

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,* 362–365.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A. D. (1992). Is working memory working? The Fifteenth Bartlett Lecture. *Quarterly Journal of Experimental Psychology, 44A*, 1–31.

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

Baddeley, A. D., & Hitch, G. (1974). Working memory. *The Psychology of Learning and Motivation*, *8*, 47–89.

Baddeley, A. D., & Hitch, G. (1994). Developments in the concept of working memory. *Neuropsychology*, *8*, 485–493.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, *36A*, 233–252.

Bogus, R. (1991). *The complete rhyming dictionary*. New York: Doubleday, revised edition.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*, 257–271.

Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term memory recall of repeated and novel word lists. *Memory and Cognition*, *21*, 539–545.

Conrad, R. (1964). Acoustic confusion and immediate memory. *British Journal of Psychology*, *55*, 75–84.

Drewnowski, A. (1980). Attributes and priorities in short-term recall: A new model of memory span. *Journal of Experimental Psychology: General*, *109*, 208–250.

Fallon, A. B. (1999). *Phonemic similarity, trace strength, and trace degradation in the immediate serial recall task*. Unpublished doctoral dissertation. University of Southern Queensland, Toowoomba, Australia.

Fallon, A. B., Groves, K., & Tehan, G. (1999). Phonological similarity and trace degradation in the serial recall task: When CAT helps RAT, but not MAN. *International Journal of Psychology*, *34*, 301–307.

Fallon, A. B., Mak, E., Tehan, G., & Daly, C. (in press). Lexicality and phonological similarity: A challenge for the retrieval-based account of serial recall. *Memory*.

Fallon, A. B., & Tehan, G. (1995, September). *The cat among the pigeons: The discovery of a reverse phonemic similarity effect in short-term serial recall*. Paper presented at the 30th Annual Australian Psychological Society Conference, Perth, Australia.

Gathercole, S. E., Gardiner, J. M., & Gregg, V. H. (1982). Modality and phonological similarity effects in serial recall: Does one's own voice play a role? *Memory and Cognition*, *10*, 176–180.

Gupta, P. (1996). Word learning and verbal short-term memory: A computational account. In G. W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Meeting of the Cognitive Science Society* (pp. 189–194). Mahwah, NJ: Lawrence Erlbaum.

Gupta, P., & Dell, G. S. (1999). The emergence of language from serial order and procedural memory. In B. MacWhinney (Ed.), *The emergence of language*, 28th Carnegie Mellon Symposium on Cognition. Mahwah, NJ: Lawrence Erlbaum.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language*, *59*, 267–333.

Houghton-Mifflin (1996). *The american heritage dictionary of the english language*. Boston: Houghton Mifflin, 3rd edition.

Kucera, H., & Francis, W. (1967). *Computational analysis of present-day american english*. Providence, RI: Brown University.

Lambert, B. L., Chang, K.-Y., & Lin, S.-J. (2003). Immediate free recall of drug names: Effects of similarity and availability. *American Journal of Health-System Pharmacy*, *60*, 156–168.

Lian, A., & Karlsen, P. J. (2004). Advantages and disadvantages of phonological similarity in serial recall and serial recognition of nonwords.

Lian, A., Karlsen, P. J., & Eriksen, T. B. (2004). Opposing effects of phonological similarity on item and order memory of words and nonwords in the serial recall task. *Memory*, *12*, 314–337.

Merriam-Webster (1987). *Webster's compact rhyming dictionary*. Springfield, MA: Merriam-Webster.

Merriam-Webster (1995). *Merriam-webster's rhyming dictionary*. Springfield, MA: Merriam-Webster.

Merriam-Webster (2001). *Merriam-webster's online dictionary*. http://www.m-w.com: Merriam-Webster.

Mitchell, K. M. (Ed.). (1996). *Essential songwriter's rhyming dictionary*. Los Angeles: Alfred.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251–269.

Nairne, J. S., & Kelley, M. R. (1999). Reversing the phonological similarity effect. *Memory & Cognition*, *27*, 45–53.

Nairne, J. S., & Neumann, C. (1993). Enhancing effects of similarity on long-term memory for order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 329–337.

Nimmo, L. M., & Roodenrys, S. (2004). Investigating the phonological similarity effect: Syllable structure and the position of common phonemes. *Journal of Memory and Language*, *50*, 245–258.

Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity, and item position. *Canadian Journal of Experimental Psychology*, *50*, 408–412.

Service, E., & Maury, S. (2003). All parts of an item are not equal: Effects of phonological redundancy on immediate recall. *Memory & Cognition*, *31*, 273–284.

Sevald, C. A., & Dell, G. S. (1994). The sequential cueing effect in speech production. *Cognition*, *53,* 91–127.

SoundEdit 16 Users Guide [Computer software], Version 1.0. (1997). San Francisco: Macromedia, Inc.

Watkins, M. J., Watkins, O. C., & Crowder, R. G. (1974). The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning and Verbal Behavior*, *7*, 319–348.

Wickelgren, W. A. (1965). Short-term memory for phonemically similar lists. *American Journal of Psychology*, *78*, 567–574.

	DisSim	CanSim	Allit	Rhym
EXPERIMENT 1				
Item Recall	65.00 (12.37)	56.75 (11.00)	68.83 (11.00)	78.75 (10.45)
Similarity:	$F(3, 69) = 38.50, p < .0005, MSE = 51.96$			
Strict Serial Recall	57.33 (15.01)	46.42 (11.13)	49.92 (14.60)	60.92 (13.41)
Similarity:	$F(3, 69) = 14.00, p < .0005, MSE = 75.77$			
Order Accuracy	87.23 (9.85)	81.43 (7.82)	72.02 (14.52)	77.17 (11.39)
Similarity:	$F(3, 69) = 14.83, p < .0005, MSE = 67.36$			
List Recall	15.42 (15.87)	4.17 (7.76)	11.67 (14.65)	17.08 (22.93)
DisSim-CanSim:	$F(1, 23) = 10.18, p < .005, MSE = 149.19$			
Similarity:	$F(3, 69) = 4.82, p < .005, MSE = 164.13$			
EXPERIMENT 2				
Item Recall	69.67 (11.11)	58.83 (12.03)	72.00 (9.96)	82.00 (9.00)
Similarity:	$F(3, 69) = 42.56, p < .0005, MSE = 50.99$			
Strict Serial Recall	62.58 (13.23)	48.92 (11.87)	52.42 (15.89)	64.58 (13.99)
Similarity:	$F(3, 69) = 18.68, p < .0005, MSE = 74.94$			
Order Accuracy	89.24 (8.45)	83.17 (10.67)	71.98 (16.72)	78.30 (12.08)
Similarity:	$F(3, 69) = 13.22, p < .0005, MSE = 97.40$			
List Recall	17.92 (15.87)	6.67 (9.63)	14.17 (14.72)	24.17 (18.86)
DisSim-CanSim:	$F(1, 23) = 26.24, p < .0005, MSE = 57.88$			
Similarity:	$F(3, 69) = 10.16, p < .0005, MSE = 126.40$			
EXPERIMENT 3				
Item Recall	90.56 (8.84)	68.54 (8.99)	69.86 (13.02)	79.17 (16.10)
Similarity:	$F(3, 69) = 38.44, p < .0005, MSE = 64.72$			
Strict Serial Recall	81.04 (15.35)	47.50 (11.23)	46.46 (16.44)	59.58 (17.37)
Similarity:	$F(3, 69) = 70.84, p < .0005, MSE = 87.55$			
Order Accuracy	88.89 (10.67)	68.86 (10.95)	65.28 (16.31)	74.15 (12.09)
Similarity:	$F(3, 69) = 33.45, p < .0005, MSE = 77.45$			
List Recall	54.86 (31.84)	7.99 (9.98)	6.95 (11.44)	19.79 (18.85)
DisSim-CanSim:	$F(1, 23) = 76.07, p < .0005, MSE = 346.65$			
Similarity:	$F(3, 69) = 49.89, p < .0005, MSE = 241.67$			
EXPERIMENT 4				
Item Recall	89.65 (6.95)	67.15 (8.10)	67.15 (10.51)	82.92 (9.43)
Similarity:	$F(3, 69) = 81.50, p < .0005, MSE = 38.16$			
Strict Serial Recall	81.11 (11.58)	52.43 (8.63)	45.42 (8.73)	62.43 (13.13)
Similarity:	$F(3, 69) = 78.82, p < .0005, MSE = 73.19$			
Order Accuracy	90.15 (8.08)	78.17 (9.08)	68.05 (10.64)	74.85 (9.80)
Similarity:	$F(3, 69) = 26.67, p < .0005, MSE = 76.97$			
List Recall	51.04 (25.58)	5.21 (6.87)	2.08 (3.69)	21.18 (23.44)
DisSim-CanSim:	$F(1, 23) = 86.97, p < .0005, MSE = 289.84$			
Similarity:	$F(3, 69) = 51.73, p < .0005, MSE = 232.71$			

Table 1: Means (with standard deviations), and statistical analyses for Experiments 1, 2, 3, and 4. DisSim = Dissimilar, CanSim = Canonically Similar, Allit = Alliterative, Rhym = Rhyming.

ITEM RECALL			
1syl-Rhym	1syl-DisSim	2syl-Rhym	2syl-DisSim
88.33 (8.46)	77.08 (12.74)	86.35 (10.00)	52.29 (15.81)
1syl Rhym-DissSim: $F(1, 23) = 20.25, p < .0005, MSE = 75.00$			
2syl Rhym-DissSim: $F(1, 23) = 171.47, p < .0005, MSE = 81.20$			
Similarity: $F(1, 23) = 118.94, p < .0005, MSE = 103.58$			
Length: $F(1, 23) = 46.10, p < .0005, MSE = 93.27$			
Interaction: $F(1, 23) = 59.34, p < .0005, MSE = 52.62$			
STRICT SERIAL RECALL			
1syl-Rhym	1syl-DisSim	2syl-Rhym	2syl-DisSim
76.35 (13.13)	71.98 (15.45)	69.06 (16.37)	41.88 (18.54)
1syl Rhym-DissSim: $F(1, 23) = 1.89, p > .15, MSE = 121.81$			
2syl Rhym-DissSim: $F(1, 23) = 57.23, p < .0005, MSE = 154.98$			
Similarity: $F(1, 23) = 34.08, p < .0005, MSE = 175.38$			
Length: $F(1, 23) = 96.42, p < .0005, MSE = 87.02$			
Interaction: $F(1, 23) = 30.79, p < .0005, MSE = 101.40$			
ORDER ACCURACY			
1syl-Rhym	1syl-DisSim	2syl-Rhym	2syl-DisSim
86.00 (8.43)	92.96 (9.52)	79.31 (13.90)	78.48 (18.94)
1syl DissSim-Rhym: $F(1, 23) = 9.60, p < .01, MSE = 60.50$			
2syl Rhym-DissSim: $F(1, 23) = .05, p > .8, MSE = 182.73$			
Similarity: $F(1, 23) = 1.56, p > .2, MSE = 144.23$			
Length: $F(1, 23) = 22.81, p < .0005, MSE = 117.97$			
Interaction: $F(1, 23) = 3.68, p = .07, MSE = 99.01$			
LIST RECALL			
1syl-Rhym	1syl-DisSim	2syl-Rhym	2syl-DisSim
57.92 (21.06)	45.000 (22.65)	45.83 (19.76)	15.83 (19.54)
1syl DissSim-Rhym: $F(1, 23) = 7.74, p < .05, MSE = 285.61$			
2syl Rhym-DissSim: $F(1, 23) = 57.23, p < .0005, MSE = 154.98$			
Similarity: $F(1, 23) = 38.96, p < .0005, MSE = 283.65$			
Length: $F(1, 23) = 65.86, p < .0005, MSE = 155.03$			
Interaction: $F(1, 23) = 10.26, p < .005, MSE = 170.61$			

Table 2: Means (with standard deviations), and statistical analyses for Experiment 5. DisSim = Dissimilar, Rhym = Rhyming, 1syl = 1-syllable, 2syl = 2-syllable.

	Study	Presentation Modality	Stimulus Set Type	Item accuracy scoring	Strict serial scoring
(1)	Wickelgren (1965)	Auditory	Closed	$Dissimilar < Rhyming$	$Dissimilar > Rhyming$
(2)	Watkins et al. (1974)	Visual, Auditory	Open	$Dissimilar = Canonical$	$Dissimilar > Canonical$
(3)	Drewnowski (1980)	Visual	Closed	$Dissimilar > Rhyming$	$Dissimilar > Rhyming$
(4)	Gathercole et al. (1982)	Visual, Auditory	Open	$Dissimilar < Rhyming$	$Dissimilar > Rhyming$
(5)	Coltheart (1993)	Visual	Closed	$Dissimilar > Canonical$	$Dissimilar > Canonical$
			Open	$Dissimilar > Canonical$	$Dissimilar > Canonical$
(6)	Poirier and Saint-Aubin (1996)	Visual	Open	$Dissimilar = Rhyming$	$Dissimilar > Rhyming$
(7)	Fallon et al. (1999)	Visual	Open	$Rhyming > Dissimilar > Canonical$	$Dissimilar = Rhyming > Canonical$
			Closed	$Dissimilar = Rhyming > Canonical$	$Dissimilar > Rhyming > Canonical$
(8)	Present Experiments 1, 2	Auditory	Open	$Rhyming > Alliterative = Dissimilar > Canonical$	$Rhyming = Dissimilar > Alliterative = Canonical$
	Present Experiment 3	Auditory	Closed	$Dissimilar > Rhyming > Alliterative = Canonical$	$Dissimilar > Rhyming > Alliterative = Canonical$
	Present Experiment 4	Auditory	Closed	$Dissimilar > Rhyming > Alliterative = Canonical$	$Dissimilar > Rhyming > Canonical > Alliterative$
	Present Experiment 5	Visual	Open	$Rhyming > Dissimilar$	$Rhyming > Dissimilar$
(9)	Lian et al. (2004, Experiment 2)	Visual	Open	$ConsonantFrame > Rhyming = Dissimilar$	$ConsonantFrame = Dissimilar > Rhyming$
(10)	Nimmo and Roodenrys (2004, Experiment 1)	Auditory	∼Open	$Rhyming > Dissimilar = Canonical$	$Dissimilar > Rhyming = Canonical$
	Experiment 2			$Alliterative = Dissimilar = Canonical$	$Dissimilar > Canonical > Alliterative$
	Experiment 3			$ConsonantFrame = Dissimilar > Canonical$	$Dissimilar > ConsonantFrame > Canonical$

Table 3: Meta-analysis of item recall and strict serial recall results from previous studies