Applied Psycholinguistics **29** (2008), 627–667
Printed in the United States of America
doi:10.1017/S0142716408080272

Is overt repetition critical to expressive word learning? The role of overt repetition in word learning with and without semantics

BRANDON ABBS, PRAHLAD GUPTA, and NAVEEN KHETARPAL
University of Iowa

Received: May 28, 2007 Accepted for publication: May 10, 2008

ADDRESS FOR CORRESPONDENCE
Prahlad Gupta, Department of Psychology, University of Iowa, E11 Seashore Hall, Iowa City, IA 52242. E-mail: prahlad-gupta@uiowa.edu

ABSTRACT
Five experiments examined whether overt repetition (i.e., saying a word aloud) during exposure is critical to the expressive learning of new words. When participants did not engage in overt repetition during exposure, they nevertheless exhibited clear expressive learning, both with and without an accompanying semantics, indicating that overt repetition is not critical to expressive word learning. In addition, learning without overt repetition did not differ from learning with overt repetition, suggesting that overt repetition confers no benefit for learning in this situation. These results are discussed in relation to previous studies, and it is suggested that benefits of repetition may accrue primarily in second language rather than in first language word learning.

Is the spoken production of new words critical to their expressive learning? That is, can a learner acquire the spoken forms of new words without overtly speaking them? The relevance of this question lies in its implications for understanding of vocabulary acquisition in general, as well as in its implications for methods of vocabulary instruction, which differ in their emphasis on the overt production of new vocabulary material. The answer to the question is not, however, obvious. Indeed, surprisingly little is known definitively about this issue.

Suppose, for instance, that a human infant was exposed to two languages that were equally represented in its linguistic environment (say, English and French). In one of the languages (say, French), the infant developed in typical ways, including engaging in linguistic communication receptively (showing comprehension) and expressively (making linguistic utterances). In English, however, the infant engaged only in receptive communication, never speaking. At age 3, would this child have a typical expressive English vocabulary if tested through picture naming (e.g., via the Peabody Picture Vocabulary Test; Dunn & Dunn, 1997)? Note that the child would lack neither articulatory skill nor "having language," because of his/her normal development in French. Demonstration of *any* expressive vocabulary would indicate that overt production is not an absolute necessity for some expressive vocabulary growth to occur. Any difference in the size of his/her expressive English vocabulary compared with that of peers would presumably be attributable to a lack of overt production of words in English, and would indicate that overt production of words does aid in development of expressive vocabulary. Would the child demonstrate any expressive vocabulary knowledge? Would the size of his/her expressive vocabulary differ from that of peers? The fact that we cannot be sure of the answer to these questions indicates that we do not know what role overt production plays in the expressive learning of spoken words.

At least two kinds of overt production can be distinguished, based on the context of production. One type of overt production occurs at the time of exposure to the sound pattern of the word (i.e., to the *word form*), in the form of immediate repetition. We will term this *overt repetition.* A second type of overt production is that occurring outside the immediate context of exposure to the word form, as in a spontaneous naming event or an utterance in which the word form is sought out to be used, or in a picture-naming event. We term this *overt naming*. Here we are concerned with the role in expressive word learning (WL) of the former type of overt production: overt (spoken) repetition.

Even with respect to this more limited sense of overt production, it is unclear how crucial it is to WL. If the hypothetical child had never engaged in overt spoken repetition, would s/he be able to engage in overt naming at all? Would the size of his/her expressive vocabulary differ from that of peers? Again, the answer to these questions is not definitively known.

Before turning to an examination of theoretical considerations as well as empirical evidence that bear on these questions, let us clarify our usage of the term "word learning." We adopt an operational definition of expressive WL: we treat it as being measured by the increase in accuracy (over some number of exposures) of overt spoken production of the correct name in a novel name-referent pairing, when cued with the referent, that is, by the increase in accuracy of a particular type of overt naming. We assume that such learning requires, at minimum (a) the establishment of some internal representation of the referent (a semantic representation, which may or may not be very well specified), (b) the establishment of some internal representation of the name (a phonological representation, which must eventually be specified well enough to support spoken production of the name), and (c) the establishment of a link from the former to the latter. In addition, what is manifested as expressive WL very likely involves some facilitation of perceptual processing of the auditory novel name, as well as some facilitation of articulatory processing in production of the novel name. These assumptions are little more than a functional description of what must occur for expressive WL to be manifested; we therefore take these assumptions to be uncontroversial (see also Gupta, 2005).

In expressive WL of this type, performance is gauged by production of the phonological word form. Accordingly, the aspect of expressive WL as described above that was of primary interest in the present work was the learning of the internal phonological representation. Furthermore, it is particularly in the formation of this phonological representation that the role of repetition has been thought to apply (e.g., Desrochers & Begg, 1987; Ellis & Beaton, 1993; see Gupta, 2005, for additional discussion of expressive WL). The aim of the present work was thus to further clarify the role of repetition, by examining the effect of overt spoken repetition on establishment of the internal phonological representation in expressive WL.

Let us further consider the critical independent variable of interest in our examination of overt spoken repetition. To discuss this, it is necessary to provide some information about the experimental design adopted. As just indicated, the dependent measure was performance in an expressive WL task as described above, in which participants received multiple exposures to pairings of novel auditory names with novel visual referents, and were tested on production of the names when cued with the referents. That is, the dependent variable was correct naming performance when cued with the referent. The manipulated variable was whether or not participants were required to overtly repeat the names during exposure trials, that is, whether or not the participants engaged in overt spoken repetition of the names during exposure trials.

What are the functional components of performance of overt spoken repetition of a word form? At least two components are clearly identifiable: first, articulatory–phonological *planning* of the word form must occur. Second, the articulatory–phonological plan must be *executed.* A critical difference between the experimental condition in which participants perform overt spoken repetition compared with the experimental condition in which participants do not perform overt spoken repetition is this *execution* of the articulatory–phonological plan.

It is important to note here that we expected that participants would engage in internal (silent) repetition, that is, internal rehearsal, in either or both experimental conditions. Evidence for this comes from the work of Papagno, Valentine, and Baddeley (1991), who found that concurrent articulation interfered with learning the pairings of known word forms with novel phonological word forms, but not with other known phonological word forms. This indicated, first, that the locus of the effect of articulatory interference was at phonology, rather than in learning the associative link, and second, that rehearsal, that is, silent internal repetition, was playing a role in the formation of the internal phonological representation. Thus, we expected participants to engage in internal rehearsal in the experimental condition that did not require overt spoken repetition.

What is this internal rehearsal? In our view, it can be considered to be articulatory–phonological planning (see, e.g., Gupta & MacWhinney, 1995). But this means that articulatory–phonological planning may be expected to occur in both conditions of our experimental design: both in the condition in which participants perform overt spoken repetition and in the condition in which participants do not perform overt spoken repetition. The critical difference between these two conditions of the experimental design is then the execution of the articulatory–phonological plan. This differentiated the two conditions, and constituted the key independent variable of interest. A finding of significant WL in the no-overt repetition condition would indicate that execution of the articulatory–phonological plan is not *necessary* for learning of the internal phonological representations of word forms. The finding of a difference between WL performance in the two conditions would suggest that execution of the articulatory–phonological plan can play a *beneficial* role in learning of the internal phonological representation, whether or not it is necessary.

Thus, the questions we ask in the present work are (a) whether overt spoken repetition (and in particular, overt execution of the articulatory–phonological plan) is *necessary* for establishment of the internal phonological representation in expressive WL, and (b) whether overt spoken repetition (overt execution of the articulatory–phonological plan) *benefits* establishment of the internal phonological representation in expressive WL. In the remainder of this article, we will use the term *overt spoken repetition* as a shorthand for this critical variable we have discussed, namely, overt execution of the articulatory–phonological plan.

In terms of theoretical or a priori considerations, there are reasons to posit as well as deny a role for overt spoken repetition in expressive WL. In contrast, intuition and observation suggest that children do not consistently engage in overt repetition, making it appear unlikely that it could be crucial to expressive WL.

The learning of word forms can be conceived of as very similar to the so-called Hebb effect. This effect refers to the finding that when subjects are presented with lists of digits for immediate serial recall, their recall improves for digit lists that, without the participant's awareness, occur repeatedly, that is, *learning* of the lists occurs (Hebb, 1961). However, when learning is measured through overt recall of a list, that is an oral response, the effect is dependent on participants' overt repetition of the lists during presentation (i.e., the opportunity to repeat the entire sequence immediately following presentation), and is not found without this overt repetition (Cunningham, Healy, & Williams, 1984). It has been argued that such learning is analogous to that underlying expressive word form learning: in either case, what is learned expressively is a *novel sequence of familiar items*: a novel ordering of the phonemes of a language, in the case of word forms, and a novel ordering of digits, in the case of lists (Burgess & Hitch, 2005; Cumming, Page, & Norris, 2003; see also Gupta, Lipinski, Abbs, & Lin, 2005). This would suggest that overt repetition may be critical to word form learning, and hence, to expressive WL.

In the language learning literature, evidence regarding the role of repetition is mixed. It has often been suggested that for learning of language as a whole (including not only vocabulary, but also pragmatics, grammar, etc.), rote repetition and memorization is a less effective strategy than more engaging methods, such as spaced recall (Atkinson, 1972; Royer, 1973). Further, the benefit of learning a language in context, such as through immersion, suggests that in comparison, rote memorization is not an effective language-learning tool. However, in second-language learners, evidence suggests that learners who generally dismiss rote memorization strategies as an effective way to learn the overall language still endorse oral, or overt, repetition of a novel phonological form as an effective strategy for learning new *words* in a language, suggesting a specific role for overt repetition in WL (Gu & Johnson, 1996). In this work, which studied adult Chinese natives who were learning English, Gu and Johnson (1996) found that, in general, the most successful learners eschewed rote repetition in favor of more contextualized approaches for learning a language, but still endorsed overt repetition as an effective rehearsal strategy for learning words. Further, this endorsement correlated significantly with a quantitative measure of English proficiency indicating that participants who believed that overt repetition of novel word forms improved their subsequent memory for the word (and were presumably more likely to engage in overt repetition) showed greater linguistic proficiency. This correlational study thus suggests a benefit for overt repetition. Others have also suggested a role for overt repetition in vocabulary acquisition (e.g., Kuhl, 2000; Studdert-Kennedy, 1986). For example, Kuhl and Meltzoff (1982, 1988) suggested that young children resort to a strategy of overt repetition of a heard sequence of phonemes during the period of language acquisition, with the possibility that such repetition may be important to WL.

In the few experimental studies that have examined the effectiveness of various rehearsal strategies in WL, overt repetition has been found to facilitate WL, but has not been found to be critical (Duyck, Szmalec, Kemps, & Vandierendonck, 2003; Ellis & Beaton, 1993; Papagno, Valentine, & Baddeley, 1991; Seibert, 1927). These studies found that overtly repeating words during learning is a more effective method for vocabulary acquisition than either silent or visual methods of rehearsal, but the silent and visual methods still showed significant learning. However, these studies focused more on associative learning between foreign language words and native language translation equivalents, and examined what role Baddeley and Hitch's (1974) phonological loop has in creating this link (but see Duyck et al., 2003, Experiment 2). Further, these studies primarily examined expressive WL in terms of written, rather than spoken response, thus limiting their implications for early WL. Thus, these studies do not provide the strongest test of a critical role for overt repetition in expressive WL. Taken together, the correlational and experimental data presented above provide some evidence for a role of overt repetition in WL. They do not, however, directly examine whether overt repetition is critical to expressive WL and have not always examined auditory presentation of word forms with performance measured through overt naming.

The approach in the present investigation is to directly examine the role of overt repetition employing a paradigm in which participants were presented with auditory novel names and visually presented novel referents (rather than definitions or translation equivalents), and asked to learn the names expressively, so that they could subsequently produce the name of each novel referent, when cued with its visual depiction. We focused directly on two questions: (a) Can participants demonstrate expressive WL, in the absence of overt repetition? (b) Does any such demonstrated learning differ from that achieved in the presence of overt repetition?

EXPERIMENT 1

The task used to measure production is what we call an expressive recall (ER) task. In this task, participants go through a number of learning phases during which they are simultaneously exposed to novel auditory stimuli (hereafter, "nonwords") and novel visual stimuli (line drawings of "space aliens"), and are asked to learn an association between the two, such that they can produce the nonword when presented with the space alien. In all of the WL experiments reported here, participants learned to produce three-syllable nonwords in response to the presentation of its associated space alien by either overtly repeating the nonword during presentation or simply being exposed to the nonword. Again, the hypothesis is that only participants who are given the opportunity to overtly repeat the nonword will demonstrate the ability to produce the nonword upon presentation of the alien.

Method

Participants. Thirty-six undergraduates at the University of Iowa participated in this experiment and received course credit in exchange for their participation.

Stimuli. All auditory and visual stimuli used in this experiment were randomly sampled from a larger corpus of experimental stimuli that have been systematically constructed in accordance with English language phonology, but are completely novel (Gupta et al., 2005). The auditory stimuli were three-syllable nonwords that had a consonant–vowel (CV–CV–CVC) phonological structure (see Appendix A for a full list of stimuli). The visual stimuli were line drawings of space aliens that were drawn by a commissioned artist (see Appendix B for a sample of these stimuli).

Design. The ER task can be broken down into two phases: an exposure phase and a test phase, which are repeated multiple times and occur at regular intervals.

EXPOSURE PHASE. In the exposure phase, a participant is seated in front of a computer screen and microphone while wearing a set of headphones. A series of five nonwords are presented over the headphones. On three of these exposure trials, a foil nonword is presented and the participant sees only a black fixation cross on the computer screen in front of him. On the other two trials, a target nonword is presented to the participant and is paired with a visual presentation of four aliens on the computer screen. One of these four space aliens is highlighted by the presence of a black box surrounding it; this alien is the referent for the target nonword. The participant is instructed that on these target exposure trials, the nonword presented is the name for the space alien that is surrounded by the box and that the task is to associate this nonword with the alien (i.e., the referent). Further, the participant is instructed to learn a target–referent pair so that he can produce the target when given the referent during the test phase. The other three aliens in the array can be thought of as visual distracters for this trial. For any given exposure phase, the foil and target presentations alternate. Thus, an exposure phase has the trial structure: Foil_1, Target_1, Foil_2, Target_2, Foil_3. On Target_2, a new target nonword is presented along with the same aliens seen on Target_1. The quadrant of the screen in which a space alien appears is randomly determined. Along with the new nonword, a different alien is highlighted on Target_2 than was highlighted on Target_1, but the instructions for the participant remain the same: to learn a target–referent pair so that he can produce the target when given the referent during the test phase. Thus, each unique exposure phase (another unique exposure phase will be introduced later) results in two target–referent pairs that the participant must learn. Whether a given pair occurs on Target_1 or Target_2 is randomly determined.

TEST PHASE. Following the exposure phase, a test phase occurs. In the test phase, the participant is again presented with the four alien array and the alien that is the referent for one of the target nonwords is surrounded by a box, just as it was during exposure (although its location on the screen may have changed). Upon presentation of this array, the participant is instructed to say, out loud, the nonword that has previously been presented as the name for the alien during the exposure phase or "I don't know" if he cannot remember the nonword. The order of the target pair testing is randomly determined.

SETS. An exposure-test phase pairing (as was just described) can be thought of as a self-contained "set" involving the presentation of two target–referent pairs and an immediate test of how well this pair was learned. Once a participant goes through one set (Set 1) of target–referent pairs, the process is repeated with another exposure–test phase pair (Set 2) that introduces two new target–referent pairs for the participant to learn and immediately tests learning for just these pairs. At this point, the participant has gone through Set 1 and Set 2, during which he has been exposed to and tested on four target–referent pairs. Together, these two sets can be thought of as comprising an epoch (Epoch 1). Of importance, the auditory and visual stimuli of Set 1 are completely independent of the stimuli of Set 2.

Once the participant has finished an epoch, there is a test phase that is the exact same structure as described above, but which tests learning for all four of the target–referent pairs, rather than the two most recently viewed pairs. This test is called a recap (Recap 1), and follows each epoch of learning. The order of target–referent pair testing is randomly determined.

Following completion of Recap 1 the participant begins Epoch 2, starting with the exposure phase of Set 1. For each new exposure phase, unique foil nonwords are presented to the participant, but the target–referent pairs remain the same. Thus, the target–referent pairs associated with Set 1 will always be associated with Set 1, although their presentation order within Set 1 will vary randomly (as previously discussed). Further, for each set, the visual array presented during each epoch will contain the same stimuli as those presented in Epoch 1. One of these distracters was highlighted as the referent during Target$_2$, but the other two remain insignificant to the participant. Thus, the participant will be hearing different foil nonwords, the same target nonwords, and will see the same visual referents for each set during each subsequent Epoch.

After Epoch 6, the participant engages in a paper and pencil distracter task (see below) for 6 min and then comes back to the computer screen for a final recap. This recap is the same as an epoch recap, but because it occurs following a distracter-filled time delay, it is designed to measure retention of the learning that took place during the experiment.

DISTRACTER TASK. The distracter task is a semantic fluency task wherein participants are given the name of a category and asked to write down as many members of that category as they can think of in a 3-min time span. After the first 3 min passed, participants were given a second category and listed members of that category for 3 min, for a total time of 6 min. The categories used in this experiment were occupations, animals, US states, and college majors. Occupations and animals were always given together, as were US states and college majors, but any given participant only listed items from one of these two pairings during the 6 min between Recap 6 and Recap 7.

Procedure. All stimuli were presented on a Macintosh G3 PowerPC using the experimental development program, PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993). Each experimental session was recorded on audiotape for later review and scoring. Participants were randomly assigned to one of two conditions: overt ($N = 18$), or silent ($N = 18$). In the overt condition, participants were asked to overtly repeat each of the nonwords they heard during the exposure phase immediately after they heard it. In the silent condition, participants were not asked to overtly repeat each of the nonwords; thus, they remained silent during the exposure phase and merely listened to each of the words that were presented to them. Participants were also randomly assigned to the stimulus set that they would have to learn. For this task, there were three sets of target–referent pairings that a participant could be asked to learn. Overall, 12 different target–referent pairings were used in the ER task (see bold items in Appendix A). Any given participant was trying to learn four pairs per task. One full epoch of practice, with separate practice nonwords and space aliens used for presentation, was completed before beginning the experimental task.

Scoring.

ER. The main measure of interest is a participant's recall performance at each of the recaps following the six epochs (subsequently referred to as Recap 1, Recap 2, etc.), and Recap 7. At each of these points, the participant response was examined using the audio recording of the experimental session. Each response was phonetically transcribed and compared to a phonetic transcription of the target nonword as it was presented during the experiment. The transcriptions used a "Klattese" mapping of the International Phonetic Alphabet to standard computer keyboard characters (see table IV in Klatt, 1987). From this transcription, we obtained the percentage of whole names, syllables, and phonemes correctly recalled at each recap of interest. Statistical analyses were done on all three levels of scoring to gain an increased level of sensitivity in the measure.

Analysis. For each measure of the task a 2 (Condition) $\times$ 6 (Recap) mixed-design analysis of variance (ANOVA) was conducted to examine performance at Recaps 1–6 and determine how much information was learned by each group during the course of the task and the trajectory of that learning. Then a separate 2 (Condition) $\times$ 2 (Recap) mixed-design ANOVA was conducted to see if the there was a significant change in performance from Recap 6 to Recap 7 and if the change differed between groups. Recap 7 was not included in the initial ANOVA because the length and nature of the delay between Recap 6 and Recap 7 make any comparison qualitatively different than a comparison between, for example, Recap 6 and Recap 5. Unless otherwise specified, an α level of .05 was adopted for each statistical test performed.

For the word and syllable measures, the number of possible values in each cell is small (for words: either zero, one, two, three, or four words correct), raising the possibility of nonnormality of distributions, and the possibility that parametric tests might be inappropriate. Examination of the error residuals of the ANOVAs for words, syllables, and phonemes correct did, in fact, reveal those of words and syllables to be nonnormally distributed. We therefore conducted a 2×6 and 2×2 ANOVA on the rank transformation (equivalent to the Friedman two-way nonparametric ANOVA; Conover & Iman, 1981) of words and syllables correct for this and the two word-learning experiments to follow. These ANOVAs are *in addition* to the ANOVAs run with proportion correct as the dependent measure. The nonparametric results are presented in parentheses next to the parametric results below.

Results

Words. For the 2×6 ANOVA, the main effect of recap was significant, F (5, 170) = 21.17, $p < .01$, $\eta^2 = .38$ (F [5, 170] = 20.71, $p < .01$, $\eta^2 = .38$), but there was no main effect of condition, F (1, 34) < 1, $\eta^2 = .002$ (F [1, 34] < 1, $\eta^2 = .003$), and no significant interaction, F (5, 170) < 1, $\eta^2 = .01$ (F [5, 170] < 1, $\eta^2 = .02$). At Recap 6, the overt condition recalled 44.4% (*SD* = 25.1) of the alien names and the silent condition recalled an average of 43.1% (*SD* = 29.5). The learning trajectories for each group across the course of the experiment are depicted in Figure 1.

The 2×2 ANOVA revealed a main effect of recap, F (1, 34) = 11.51, $p < .01$, $\eta^2 = .25$, (F [1, 34] = 13.22, $p < .01$, $\eta^2 = .28$) but no main effect of condition, F (1, 34) < 1, $\eta^2 = .02$ (F [1, 34] < 1, $\eta^2 = .02$), and no interaction effect, F (1, 34) = 1.56, $p = .22$, $\eta^2 = .04$ (F [1, 34] = 1.14, $p = .29$, $\eta^2 = .03$). At Recap 7, the overt condition recalled an average of 36.1% (*SD* = 30.0) and the silent condition recalled an average of 25.0% (*SD* = 27.1).

Syllables. The 2×6 ANOVA found a main effect for recap, F (5, 170) = 44.45, $p < .01$, $\eta^2 = .57$ (F [5, 170] = 44.57, $p < .01$, $\eta^2 = .57$), but no main effect of condition, F (1, 34) < 1, $\eta^2 = .0003$ (F [1, 34] = <1, $\eta^2 = .002$), and no interaction effect, F (5, 170) < 1, $\eta^2 = .01$ (F [5, 170] < 1, $\eta^2 = .02$). At Recap 6, the overt condition could recall 64.4% (*SD* = 20.4) of the syllables and the silent condition could recall 63.4% (*SD* = 26.4).

The 2×2 ANOVA showed a main effect of recap, F (1, 34) = 17.14, $p < .01$, $\eta^2 = .34$ (F [1, 34] = 5.24, $p < .05$, $\eta^2 = .13$), but no main effect of condition, F (1, 34) < 1, $\eta^2 = .01$ (F [1, 34] < 1, $\eta^2 = .02$), and no interaction effect, F (1, 34) = 1.28, $p = .27$, $\eta^2 = .04$ (F [1, 34] < 1, $\eta^2 = .002$). At Recap 7, the overt condition recalled 55.1% of the syllables (*SD* = 20.4) and the silent condition recalled 47.2% (*SD* = 28.6).

Phonemes. The 2×6 ANOVA revealed a main effect of recap, F (5, 170) = 60.91, $p < .01$, $\eta^2 = .64$, with no main effect of condition, F (1, 34) < 1, $\eta^2 = .002$, and no interaction, F (5, 170) < 1, $\eta^2 = .02$. At Recap 6, the overt condition

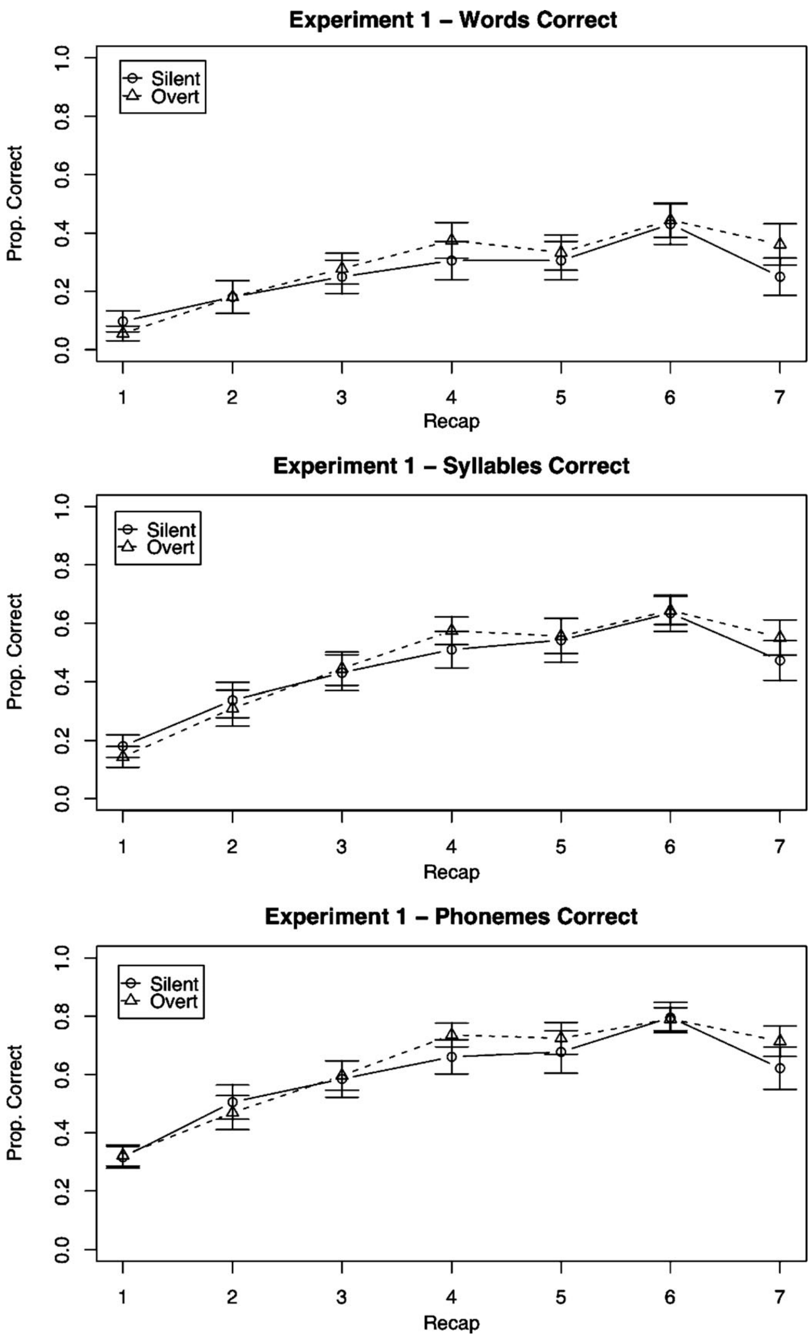

Figure 1. The performance in Experiment 1 for participants in the silent and overt conditions across all seven recaps. Performance is measured as the proportion correct and is plotted separately in terms of (top) words, (middle) syllables, and (bottom) phonemes correct. Error bars are standard errors.

recalled 79.0% of the phonemes ($SD = 16.6$) and the silent condition recalled 79.6% ($SD = 22.0$).

The 2×2 ANOVA revealed that there was a main effect of recap, $F(1, 34) = 25.84$, $p < .01$, $\eta^2 = .43$, no main effect of condition, $F(1, 34) < 1$, $\eta^2 = .01$, and a borderline interaction effect, $F(1, 34) = 3.95$, $p = .055$, $\eta^2 = .10$. At Recap 7, the overt condition recalled an average of 71.4% ($SD = 22.2$) of the phonemes and the silent condition recalled an average of 62.2% ($SD = 30.8$).

Power analysis. We examined the sensitivity of the experiment, for each of the foregoing analyses, by assessing the size of the effect our tests can detect with power of .80. This sensitivity was computed with the assistance of G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007), which implements the procedures described by Cohen (1988).

In the 2×6 ANOVAs, for the between-groups main effect of repetition, this experiment was capable of detecting a medium or large effect size of $f \geq .29$ (corresponding to $\eta^2 \geq .08$) with a power of .80. Similarly, for the within-group main effect of recap and the within-between interaction, a medium or large effect size of $f \geq .23$ (corresponding to $\eta^2 \geq .05$) would be detected. For smaller effect sizes, the execution of "retrospective power analyses" based on the "observed power" implied by our data is highly criticized by statisticians as an invalid way of assessing the validity of null effects (for a discussion, see Faul, Erdfelder, Lang, & Buchner, 2007; for criticism, see Gerard, Smith, & Weerakkody, 1998; Hoenig & Heisey, 2001; Kromrey & Hogarty, 2000; Lenth, 2001; Steidl, Hayes, & Schauber, 1997). One accepted way of assessing effects retrospectively is through the comparison of effect sizes (Zumbo & Hubley, 1998), which we have done in our discussion. Another alternative to a retrospective approach is to execute a "criterion analyses" that tells us how much we need to adjust our alpha level to safely achieve a particular level of power. Such an analysis has its own critics (e.g., Gigerenzer, Krauss, & Vitouch, 2004) will yield α levels far removed from the traditional .05 level; however, its use in conjunction with an analysis of observed effect sizes lessens these criticisms some. Thus, we executed a criterion analysis using G*Power 3 and found that to detect a small effect ($f = .10$) with a power of .80, the α value of this experiment would need to be adjusted to .69 for the between-group effect and .60 for the within–between interaction. Such a change would not change the significance for the main effects of learning reported here for words, syllables, or phonemes. Further, it would not change the significance for the interaction effects for words, or syllables, but would make the phoneme level measure statistically significant. However, with an η^2 value of .02 for this effect, the functional significance of such an effect in explaining the variance between the overt repetition and silent group would be in doubt even if one accepted the statistical significance following alpha level adjustment.

For the measures of retention (i.e., in the 2×2 ANOVAs), this experiment was capable of detecting a large between-group effect ($f \geq .42$, corresponding to $\eta^2 \geq .15$), and a medium or large within-group and interaction effect ($f \geq .24$, corresponding to $\eta^2 \geq .054$) with a power of .80. The difficulty with detecting between-group effects lies in the strong correlation between the within-group measures. To detect a small ($f = .10$, which is closer to what is being observed in these experiments) between-group effect with power of .80, the α level would need to be adjusted to .75. Such an adjustment would not make the between-groups effects significant at either the word, or syllable level, but it would make an effect at the phoneme level significant. However, the functional significance of such an effect would still be in question, as the maximum η^2 for these effects is .01. For the interaction, an adjustment of α to .61 would need to be made to detect a small effect. Such an adjustment would make the interaction significant at the word level, but not the syllable level. Once again, though, such an effect would account for just .04 of the variance in this measure.

It should be noted here that these sensitivity analyses will also apply to Experiments 2 and 3, as the paradigm (WL Task), measures (word, syllable, and phoneme accuracy), experimental design (2×6 and 2×2 mixed design), and sample size ($N = 36$) of these experiments will not change.

Discussion

For the above analyses, one can think of the first analysis of performance, which compared performance at Recaps 1–6 as examining differences in learning across multiple repetitions and retaining linguistic information across short time spans. The time between measurement points is small, less than 1 min, and is filled by exposures to the target stimuli as well as tests of even shorter term retention. The second analysis, comparing Recap 6 with Recap 7, is one of long-term retention without further exposure to the experimental stimuli. Here, the delay between tests is 6 min, and is filled by an irrelevant distracter task.

Performance at test was fairly low (fewer than two out of four of words correctly recalled even after six recaps). This might appear to contrast with real-world performance, in that children are thought to learn new words in as little as one exposure ("fast mapping"; Carey, 1978). It may therefore be worth emphasizing the fact that the phonological targets in the present experiment were three syllables long, in contrast with the one-syllable words typically studied in fast mapping situations. Learning trisyllabic phonological forms can, of course, be expected to be considerably slower than learning monosyllabic forms. In addition, recent evidence (Horst & Samuelson, in press) indicates that real-world "fast mapping" is actually much less fast than previously believed. Learning in the present experiments therefore does not, in fact, appear to be particularly slower than real-world learning.

It is also worth reiterating that we expected participants in the silent condition to be rehearsing silently, that is, engaging in what we previously termed internal repetition. As noted in the Introduction, the literature indicates that internal repetition/rehearsal may be important for expressive WL (e.g., Duyck et al., 2003; Papagno et al., 1991), although its role may depend on aspects of the experimental paradigm, a point we return to in the General Discussion. Further investigation of the role of internal repetition, especially in a paradigm of the present kind, is therefore an important question for future research. However, the focus of the present investigation is on what role *overt (spoken) repetition* plays (and in particular, what role execution of the articulatory–phonological plan plays), which is a question that has hitherto remained unanswered, and whose importance is independent of whether or not internal repetition plays a role.

With regard to overt repetition, the two questions of interest were (a) whether expressive learning would occur without overt repetition, and (b) whether overt repetition would be beneficial to WL. The robust learning observed in the silent condition with all three types of scoring (word level, syllable level, and phoneme level) indicates clearly that learning can occur in the absence of overt repetition, which is therefore not critical to expressive learning. Experiment 1 thus provides strong evidence regarding our first question of interest.

Turning to the second question, there was no evidence for a benefit of overt repetition in the 2×6 ANOVAs. Although there was a significant effect of recap, establishing that participants in this experiment were learning, there was no Recap $\times$ Condition interaction that would suggest differential learning based on whether or not a participant was engaging in overt repetition. Learning clearly occurs in the silent group and the acquisition trajectory is no different for either group. These results hold no matter the level of analysis for accuracy (i.e., words, syllables, or phonemes correct).

The effect size and sensitivity analyses are relevant to interpreting these null effects. For the 2×6 ANOVAs, the main effect of condition and the Recap $\times$ Condition interaction accounted for at most 2% of the variance in the data ($\eta^2 =$ .02), in the various analyses. This corresponds to an effect size index of $f = .143$, classified by Cohen as a small effect size (Cohen, 1988). Our sensitivity analyses confirm that the experiment was capable of detecting moderate to large effects with a power of 0.80, and therefore would (with 80% probability) have yielded a significant main effect of Condition or Recap $\times$ Condition interaction had any such effect been of even moderate size. The effect size and sensitivity analyses thus provide evidence that overt repetition does not contribute in any substantial way to expressive WL. However, it remains possible that there is a small beneficial effect of overt repetition on expressive WL, accounting for on the order of 1–2% of variance in the experimental data.

For the longer term retention examined by the 2×2 ANOVAs, the effect size for the interaction was larger than for any other nonsignificant result ($\eta^2 = .10$, which corresponds to an effect size index of $f = .333$, classified by Cohen as a medium effect size). Our sensitivity analyses indicate that the experiment was capable of detecting a medium or large effect size with a power of 0.80. This is consistent with the borderline significant interaction in the long-term retention of phonetic information, wherein participants in the overt group recalled more phonemes of the words that they were trying to learn following a 6-min task-filled delay. This borderline effect suggests that although overt repetition is not critical to acquisition, it may be critical to retention, a qualitative distinction that has also been made in other literatures (cf. Schmidt & Bjork, 1992). Participants appear to be affected by the repetition manipulation in such a way that overt repetition aids the establishment of longer term expressive representations of word forms. Having determined that participants can acquire new words in an expressive learning task without overt repetition and received some information that participants may differ on their retention of these words as a function of repetition, we turn now to the question of whether participants can *retain* new words over long delays without overt repetition. To do this, we extend the delay imposed in Experiment 1 to a delay of more than 23 hr.

EXPERIMENT 2

Experiment 2 is designed to extend the delay between the final exposure and final test to determine if overt repetition is critical to retention with an approximately 23-hr delay. If any benefit of overt repetition is related to retention, and thus time dependent, then extending the delay could lead to a greater benefit for the overt repetition condition over the silent condition. In addition, if the benefit is time dependent, then such a finding would provide information about the functional significance of the delay effect seen in Experiment 1, because one might expect that a "testing" situation (i.e., a situation in which one can demonstrate expressive knowledge about a word) during natural language acquisition does not occur within minutes of hearing a new word; rather, it occurs after longer periods of time (i.e., hours or days). A corollary benefit to Experiment 2 is a complete replication of Experiment 1 with a new set of stimuli, because the first session of Experiment 2 will be identical to Experiment 1.

Method

Participants. Thirty-six participants were involved in this experiment. Five participants were recruited through a posting on the University of Iowa's JobNet website and received $16 for participating. The remaining participants were undergraduates at the University of Iowa and received course credit for participating.

Stimuli. All auditory and visual stimuli used in this experiment were randomly sampled from the same corpus as Experiment 1. None of the target–pair stimuli overlapped with the stimuli in Experiment 1 (see Appendix C).

Procedure. The procedure for Experiment 2, Session 1, was identical to that of Experiment 1. Participants engaged in the ER task, including the distracter portion of this task. Stimuli were presented through the same apparatus and counterbalancing as well as the use of multiple stimulus sets was undertaken in the same manner. Upon completion of the ER task, participants were instructed to return at the same appointment time the next day. Participants were encouraged to avoid alcohol and recreational drugs and to get a full night's sleep during their time away from the lab. They were not told the nature of the tasks that they would be asked to complete during the next day's session, Session 2.

During Session 2, participants first completed a recap test phase (delayed recap) that tested all four of the target–referent pairs that had been studied and tested during Session 1. This test phase was conducted in the same manner as the recap test phases completed the previous day. No further instructions or prompting were provided beyond the referent for the target nonword that the participant was to produce.

Scoring and analysis. Responses were scored using the same criteria and method as Experiment 1. An additional 2 × 2 mixed-design ANOVA was conducted to assess the difference in accuracy at the end of Session 1 (final recap) and the beginning of Session 2 (delayed recap).

Results

Words. The 2 × 6 ANOVA revealed a main effect of recap, F (5, 170) = 19.66, $p < .01$, $\eta^2 = .37$ (F [5, 170] = 19.54, $p < .01$, $\eta^2 = .37$), no main effect of condition, F (1, 34) < 1, $\eta^2 = .005$ (F [1, 34] < 1, $\eta^2 = .009$), and no interaction, F (5, 170) < 1, $\eta^2 = .01$ (F [5, 170] < 1, $\eta^2 = .02$). At Recap 6, the overt and silent condition both recalled 41.7% of the space alien names (see Figure 2).

The 2 × 2 ANOVA assessing the change in learning from Recap 6 to Recap 7 showed a significant effect of Recap for the parametric test, but not for the nonparametric, F (1, 34) = 5.29, $p < .05$, $\eta^2 = .13$, (F [1, 34] = 2.37, $p = .13$, $\eta^2 = .07$), no main effect of condition, F (1, 34) < 1, $\eta^2 = .0001$ (F [1, 34] < 1, $\eta^2 = .0003$), and no interaction, F (1, 34) < 1, $\eta^2 = .001$ (F [1, 34] < 1, $\eta^2 = .001$). At Recap 7, the silent condition recalled 34.7% of the aliens ($SD = 33.4$) and the overt condition recalled 33.3% ($SD = 32.1$).

The 2×2 ANOVA assessing the change in learning from Recap 7 to the delayed recap, showed a main effect for recap, F (1, 34) = 5.64, $p < .05$, $\eta^2 = .14$ (F [1, 34] = 7.15, $p = .01$, $\eta^2 = .18$), no main effect for condition, F (1, 34) < 1, $\eta^2 = .0002$ (F [1, 34] < 1, $\eta^2 = .0001$), and no interaction, F (1, 34) < 1, $\eta^2 = .001$ (F [1, 34] < 1, $\eta^2 = .008$). Upon return to the lab, participants in both the silent and overt condition recalled 25.0% of the space alien names, which was significantly above 0 for both groups combined, t (35) = 4.66, $p < .01$, $d = .78$.

Syllables. The 2 × 6 ANOVA found a main effect for recap, F (5, 170) = 44.66, $p < .01$, $\eta^2 = .57$ (F [5, 170] = 44.67, $p < .01$, $\eta^2 = .57$), no main effect of condition, F (1, 34) < 1, $\eta^2 = .01$ (F [1, 34] < 1, $\eta^2 = .01$), and no interaction effect, F (5, 170) < 1, $\eta^2 = .03$ (F [5, 170] < 1, $\eta^2 = .03$). At Recap 6, the overt condition could recall 62.0% ($SD = 30.3$) of the syllables and the silent condition could recall 63.9% ($SD = 28.9$).

The 2 × 2 ANOVA (Recap 6–final recap) for showed a main effect of recap, F (1, 34) = 6.62, $p < .01$, $\eta^2 = .17$ (F [1, 34] = 5.41, $p < .05$, $\eta^2 = .14$), no effect of condition, F (1, 34) < 1, $\eta^2 = .0001$ (F [1, 34] < 1, $\eta^2 = .0005$), and no interaction effect, F (1, 34) < 1, $\eta^2 = .02$ (F [1, 34] < 1, $\eta^2 = .02$). At Recap 7, the overt condition recalled 56.0% of the syllables ($SD = 28.8$) and the silent condition recalled 53.7% ($SD = 30.7$).

The additional 2 × 2 mixed-design ANOVA (final recap–next day) showed a main effect for recap, F (1, 34) = 18.11, $\eta^2 = .35$ (F [1, 34] = 21.38, $p < .05$, $\eta^2 = .39$), no main effect for condition, F (1, 34) < 1, $\eta^2 = .003$ (F [1, 34] < 1, $\eta^2 = .008$), and no interaction, F (1, 34) < 1, $\eta^2 = .001$ (F [1, 34] < 1, $\eta^2 = .004$). Participants in the silent recalled 37.5% ($SD = 35.2$) of the syllables and participants in the overt condition recalled 41.2% ($SD = 32.6$) of the syllables when tested at the delay recap. The combined performance of both groups (39.4%), was significantly above 0, t (35) = 7.05, $p < .001$, $d = 1.18$.

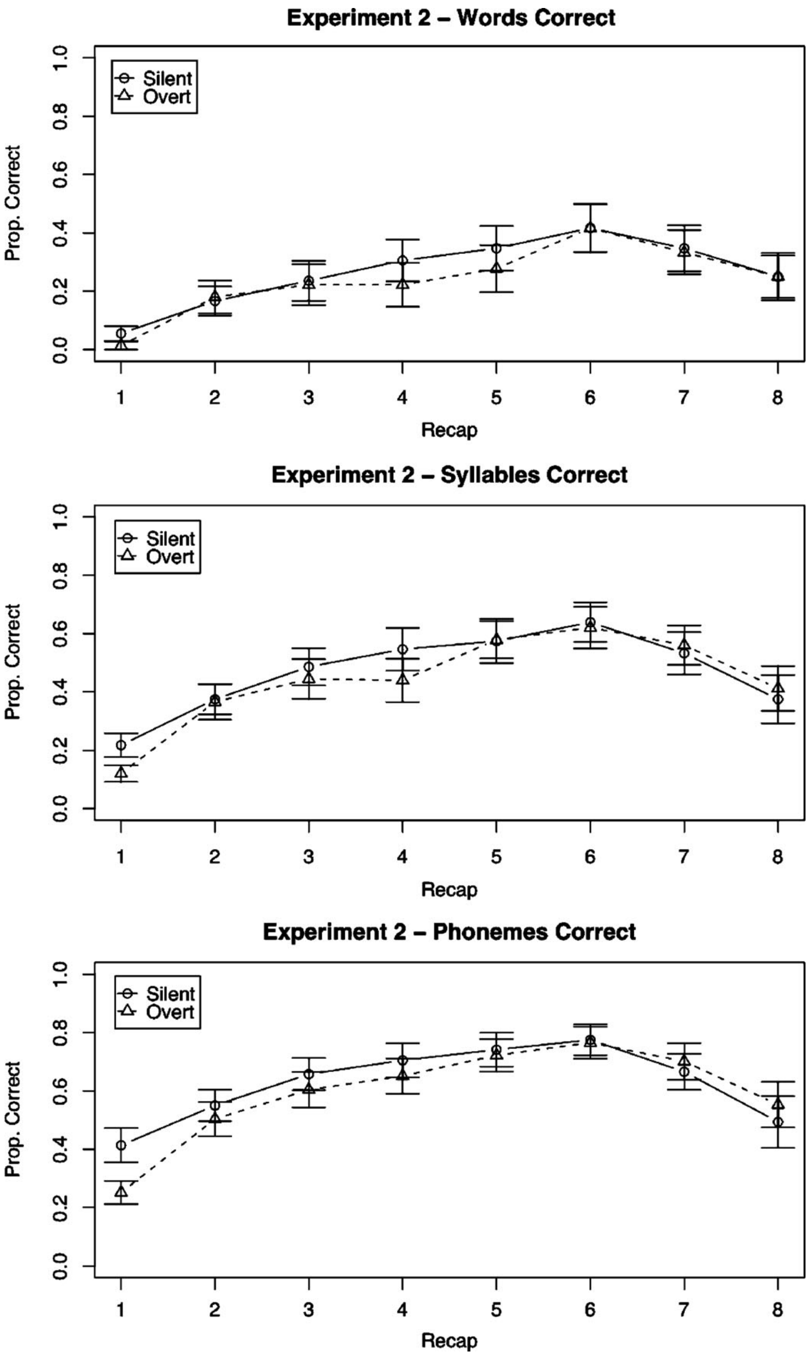

Figure 2. The performance in Experiment 2 for participants in the silent and overt conditions across the entire first session (Recaps 1–7) and the first test of the second session (Recap 8). Performance is plotted separately in terms of (top) words, (middle) syllables, and (bottom) phonemes correct. Error bars are standard errors.

Phonemes. A 2×6 mixed-design ANOVA found a main effect of recap, F (5, 170) = 59.01, $p < .01$, $\eta^2 = .63$, with no main effect of condition F (1, 34) = < 1, $\eta^2 = .02$, and no interaction, F (5, 170) = 1.68, $p = .14$, $\eta^2 = .05$. At Recap 6, the overt condition recalled 76.5% of the phonemes (SD = 23.2) and the silent condition recalled 77.5% (SD = 22.3).

A 2×2 mixed-design ANOVA (Recap 6–final recap) revealed that there was a main effect of recap, F (1, 34) = 9.92, $p < .01$, $\eta^2 = .25$, no main effect of condition, F (1, 34) < 1, $\eta^2 = .001$, and no interaction effect, F (1, 34) < 1, $\eta^2 = .02$. At Recap 7, the overt condition recalled an average of 70.1% (SD = 26.5) of the phonemes and the silent condition recalled an average of 67.8% (SD = 26.2).

Again, the additional of 2×2 ANOVA (final recap–next day) showed a main effect for recap, F (1, 34) = 20.50, $\eta^2 = .23$, no main effect for condition, F (1, 34) < 1, $\eta^2 = .0003$, and no interaction, F (1, 34) < 1, $\eta^2 = .01$. Participants in the silent condition recalled 49.4% (SD = 37.8) of the phonemes in the space alien names at the delay recap, whereas the overt condition recalled 55.3% (SD = 33.1). The combined performance of both groups (52.3%) was significantly above 0, t (35) = 8.93, $p < .01$, $d = 1.49$.

Criterion analysis. The power analysis of Experiment 1 also applies to Experiment 2 because it involves the same paradigm, measures, experimental design, and sample size. For learning, we previously determined that in order to detect a small effect ($f = .10$) with power of .80, the α of the experiment would need to be adjusted to .69 for the main effect and .60 for the interaction. Such a change would change the significance for the main effects of learning reported for Experiment 2 for words, syllables, or phonemes. Further, it would change the significance for the interaction effects for syllables, and phonemes, but it would not affect the significance of the word measure. However, the η^2 values for these effects range from just .005 to .05, and they pale in comparison to the .37–.63 values found for the recap effect. Thus, although an effect of overt repetition cannot be unequivocally rejected on statistical grounds, its effect on learning is trivial compared to the robust learning that silent participants demonstrate both here and in Experiment 1.

To detect a small ($f = .10$) between-group effect with power of .80, the α level would need to be adjusted to .75. Such an adjustment would not make the between-groups effects of the initial 6-min delay (i.e., Day 1) or across the longer 23-hr delay (i.e., Day 2) significant at either the word, syllable, and phoneme level. For the interaction effect, an α level of .61 would be needed to detect a small effect. This adjustment would not change the significance of the effect at the word or syllable level, but would make the phoneme-level significant. However, we stress again that such an effect accounts for little of the variance ($\eta^2 = .01$) compared to the extremely robust learning ($d = 1.49$) that is maintained at the phoneme level across both conditions.

Discussion

The result of Experiment 1 that indicated that overt repetition may benefit the retention of new words did not hold up in the face of replication or a more powerful test of retention following an extended delay. Participants in the both conditions showed equivalent retention on Day 2 as evidenced by a lack of an interaction between conditions in the decrease from Day 1 to Day 2 and performance for both groups combined that was significantly above 0 on Day 2 for all measures.

The sensitivity analyses of Experiment 1 are also applicable to Experiment 2. These, together with the effect sizes reported above for Experiment 2, indicate once again that had there been moderate to large effects of condition, or of the Condition × Recap interaction, the experiment would have detected them with 80% probability, that is, with a power of .80 (except in the case of the retention main effect of condition, where only a large effect would be detected). In fact, however, these effects are small in all analyses. Thus, in Experiment 2, as in Experiment 1, the sensitivity analyses support the conclusion that overt repetition does not contribute in any substantial way to expressive WL, although it remains a possibility that there is a small beneficial effect of overt repetition on expressive WL, accounting for on the order of 1–2% of variance in the experimental data.

Thus far, learning has been shown for participants who do not overtly repeat a stimulus that they are later asked to expressively recall. Further, this learning has been shown to be nearly identical to the performance of participants who are overtly repeating the words given to them. However, the participants who are repeating are not only repeating stimuli, but also receiving an additional stimulus presentation provided by their vocalization. Given this confound, there remains the possibility that participants who do not overtly repeat may actually perform *better* than participants who are not repeating if they are given the same number of exposures during test. In the current design, overt participants are able to hear the stimulus spoken by both a standardized voice and their own voice, whereas silent participants hear only the standardized voice. In Experiment 3, we attempt to completely equalize the experience of participants in both conditions to isolate the effect of overt repetition and investigate whether there is a situation in which participants who do not overtly repeat stimuli may exceed the performance level of participants who do.

EXPERIMENT 3

The goal of Experiment 3 was to address the confound wherein participants in the overt condition enjoy the benefit of an additional stimulus presentation because of their own vocalizations. Given that the overt condition has thus far shown nearly identical learning to the silent condition, there is the possibility participants in the silent condition might surpass the accuracy of participants in the overt condition if they are given the same number of exposures to the to be learned word.

Method

Participants. Thirty-six undergraduates at the University of Iowa participated in this experiment. They received course credit for their participation.

Stimuli. All auditory and visual stimuli used in this experiment were randomly sampled from the same corpus as Experiment 1. The nonwords presented to participants were three syllables in length and had the same CV–CV–CVC phonetic structure. However, the target–referent pairs were changed where necessary to ensure that there was no overlap (in either target or referent) with the stimuli presented as targets and referents in Experiments 1 and 2 (see Appendix D).

Nonwords presented to participants in the silent condition were created using the editing software Sound Edit 16, Version 2. The original sound file was simply copied and pasted next to itself in the editing palette, creating two copies of the nonword, side by side. The effect of this editing was to create a single sound file that had two presentations of the same nonword, without a predetermined interstimulus interval. The only break between presentations was the poststimulus silence of the first presentation, which was minimal (<100 ms).

Procedure. Other than the change in stimuli for the silent condition, the procedure for Experiment 3 was identical to that of Experiment 1. Participants engaged in an ER of identical structure and the same distracter task as Experiment 1. Stimuli were presented through the same apparatus and counterbalancing and the use of multiple stimulus sets was undertaken in the same manner.

Scoring and analysis. Experiment 3 was scored and analyzed using the same criteria and tests as Experiment 1.

Results

Words. The 2×6 ANOVA revealed a main effect of recap, F (5, 170) = 44.76, $p < .01$, $\eta^2 = .57$ (F [5, 170] = 45.47, $p < .01$, $\eta^2 = .57$), but no main effect for condition, F (1, 34) < 1, $\eta^2 = .01$ (F [1, 34] < 1, $\eta^2 = .003$), and no interaction effect, F (5, 170) = 1.28, $p = .27$, $\eta^2 = .04$ (F [5, 170] = <1, $\eta^2 = .02$). At Recap 6, the overt condition averaged 56.0% names correctly recalled (SD = 39.8) whereas the silent condition averaged 47.0% (SD = 27.0). Thus, despite the overt condition achieving higher levels of performance the magnitude of learning did not significantly differ between groups (see Figure 3).

The 2×2 ANOVA revealed a statistical trend for a main effect of recap, F (1, 34) = 3.86, $p = .06$, $\eta^2 = .10$ (F [1, 34] = 2.79, $p = .10$, $\eta^2 = .08$), no main effect of condition, F (1, 34) < 1, $\eta^2 = .02$ (F [1, 34] < 1, $\eta^2 = .01$), and no interaction effect, F (1, 34) < 1, $\eta^2 = .001$ (F [1, 34] < 1, $\eta^2 = .02$). At Recap 7, participants in the overt condition correctly recalled 48.6% of the words (SD = 33.7), whereas participants in the silent condition correctly recalled 38.9% (SD = 28.7).

Syllables. The 2×6 ANOVA revealed a main effect of recap, F (5, 170) = 64.68 $p < .01$, $\eta^2 = .66$ (F [5, 170] = 63.68 $p < .01$, $\eta^2 = .65$), but no main effect for condition, F (1, 34) < 1, $\eta^2 = .01$ (F [1, 34] < 1, $\eta^2 = .01$), and no interaction effect, F (5, 170) < 1, $\eta^2 = .005$ (F [5, 170] < 1, $\eta^2 = .007$). At Recap 6, the overt condition averaged 70.8% syllables correctly recalled (SD = 31.1), whereas the silent condition averaged 67.6% (SD = 27.2).

The 2×2 ANOVA revealed a statistical trend for a main effect of recap, F (1, 34) = 3.87, $p = .06$, $\eta^2 = .10$ (F [1, 34] = 4.65, $p < .05$, $\eta^2 = .12$), no main effect of condition, F (1, 34) < 1, $\eta^2 = .01$ (F [1, 34] < 1, $\eta^2 = .01$), and no

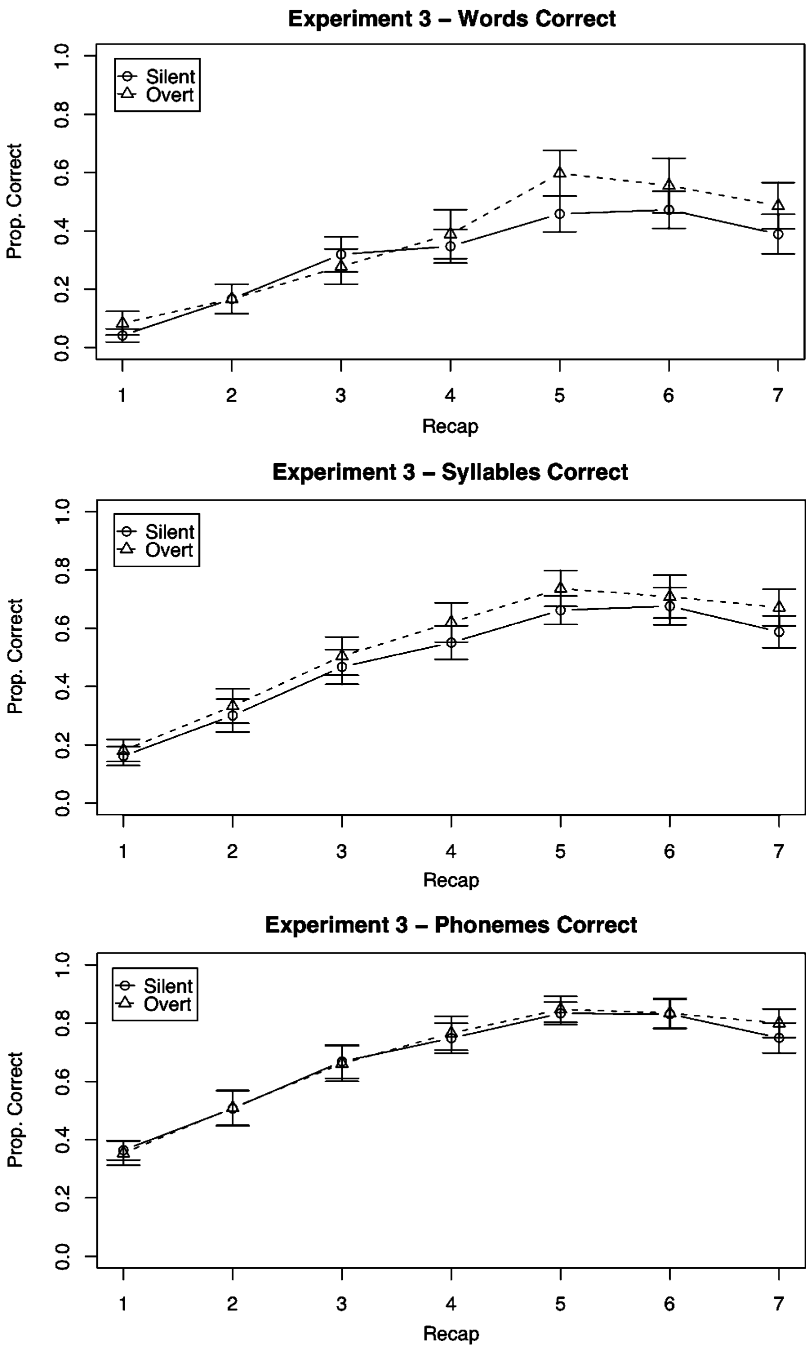

Figure 3. The performance in Experiment 3 for participants in the silent and overt conditions across all seven recaps. Participants in the repeat condition of this experiment received two presentations of all nonwords heard during the experiment. Performance is plotted separately in terms of (top) words, (middle) syllables, and (bottom) phonemes correct. Error bars are standard errors.

interaction effect, F (1, 34) < 1, $\eta^2 = .02$ (F [1, 34] < 1, $\eta^2 = .03$). At Recap 7, participants in the overt condition correctly recalled 67.1% of the syllables ($SD =$ 26.6), whereas participants in the silent condition correctly recalled 58.8% ($SD =$ 23.1).

Phonemes. The 2×6 mixed-design ANOVA revealed a main effect of recap, F (5, 170) $= 73.92$ $p < .01$, $\eta^2 = .68$, but no main effect for condition, F (1, 34) < 1, $\eta^2 = .0002$, and no interaction effect, F (5, 170) < 1, $\eta^2 = .002$. At Recap 6, the overt condition averaged 83.4% phonemes correctly recalled ($SD = 21.7$), whereas the silent condition averaged 83.1% ($SD = 21.1$).

The 2×2 mixed-design ANOVA revealed a main effect of recap, F (1, 34) $=$ 5.99, $p = .02$, $\eta^2 = .15$, no main effect of condition, F (1, 34) < 1, $\eta^2 = .005$, and no interaction effect, F (1, 34) < 1, $\eta^2 = .03$. At Recap 7, participants in the overt condition correctly recalled 83.1% of the phonemes ($SD = 20.8$), whereas participants in the silent condition correctly recalled 74.9% ($SD = 21.9$).

Criterion analysis. At an adjusted α level of .69, the between-group effect on learning would still be insignificant at the phoneme level, but not the syllable or word level. For the interaction ($\alpha = .60$), differences between accuracy in terms of phonemes and syllables would still be insignificant, but accuracy in terms of words would be significant.

For adjusted alpha levels on the retention measure, the between-group effect would be significant ($\alpha = .75$) at the word, syllable, and phoneme level. However, the interaction would be significant only at the syllable and word level.

Discussion

The power analysis here confirms that this experiment comes closest to demonstrating an effect of overt repetition on WL, perhaps because of the additional manipulation that conferred an extra presentation on participants in the overt repetition condition. Participants in this experiment showed the most robust learning and the least forgetting of any of the three experiments reported here. Under very relaxed (to the point of being unreasonable) alpha levels there are effects of overt repetition, but we again stress that these effects are incredibly small ($\eta^2 = .0002$–.04) compared with the extremely robust learning ($\eta^2 = .57$–.68) shown across all conditions and measures of learning.

Thus, there again was clear evidence of learning in the silent condition but no evidence for a benefit of overt over silent repetition nor for a benefit of silent over overt repetition, during a task involving expressive WL. As in Experiments 1 and 2, the present experiment would in most cases have detected moderate or large effects of condition, or of the Condition $\times$ Recap interaction with power of .80. The effects, however, are of small magnitude. The results provide further evidence that overt repetition does not substantially benefit expressive WL, although once again, the possibility remains of a small beneficial effect accounting for about 1–2% of the variance.

However, there is one effect, the between-group effect for retention, for which we only have the sensitivity to detect a large effect ($f \geq .42$). It is also clear that there is a large amount of variance in the overall performance of both groups in these three experiments, which reduces the experimental sensitivity of each experiment. However, we now have data from three experiments with identical structure and nearly identical procedure up until Recap 7. To increase the sample size, and hence, the sensitivity of the analysis, therefore, we undertook a meta-analysis of data from all three experiments before drawing our final conclusions about the benefit of overt repetition.

Meta-analysis

To achieve the power of .80 that we were previously achieving through adjustments to the alpha level of our statistical tests, an impractical number of subjects would need to be run in this paradigm. For the between-subjects effect of our main manipulation, overt repetition versus silent, a total sample size of 298 participants would allow us sufficient power (.80) to detect a small ($f = .10$) effect on each measure of learning. For the interaction effect, an N value of 188 is needed to detect a small effect. Medium effect sizes can be detected with much more modest gains in sample size ($N = 50$ for the between-subjects effect; $N = 32$ for the interaction effect, which we have already achieved).

For the small effects of our manipulation on retention, even larger sample sizes would be needed. The between-subjects effect on retention requires 592 participants, and the interaction requires 200. For medium effects, 98 participants is still needed for the between subjects effect and 34 is needed for the interaction.

For the meta-analysis, the data from all three experiments was combined resulting in 108 participants with 54 participants who had been in the overt condition and 54 participants who had been in the silent condition. The same type of analysis as Experiments 1, 2, and 3 was conducted, except experiment was added as a factor to ensure our assumption of no difference between experiments is valid. Thus, we conducted a 3 (Experiment) × 2 (Condition) × 6 (Recap) ANOVA and a 3 (Experiment) × 2 (Condition) × 2 (Recap) ANOVA for words, syllables, and phonemes recalled.

The increased sample size resulted in greater sensitivity compared with Experiments 1–3 individually, while maintaining power at .80. For effects across Recaps 1–6, sensitivity increased from being able to detect effects of $f \geq .29$ to detecting effects of $f \geq .17$ ($\eta^2 \geq .03$) for the main effect of condition and from $f \geq .23$ to $\geq .12$ ($\eta^2 \geq .015$) for the Condition × Recap interaction. For effects across Recaps 6–7, sensitivity increased from $f \geq .42$ to $\geq .24$ ($\eta^2 \geq .05$) for the main effect of condition, and from $f \geq .24$ to $\geq .14$ ($\eta^2 \geq .02$) for the Condition × Recap interaction.

Results

Experiment factor.

SHORT-TERM LEARNING. No main effect for experiment was found in the 3 × 2 × 6 ANOVAs conducted on the accuracy scores for words, $F(2, 102) = 1.45$, $p = .24$, $\eta^2 = .006$ ($F[2, 102] = 1.64$, $p = .20$, $\eta^2 = .03$), syllables, $F(2, 102) < 1$,

$\eta^2 = .01$ (F [2, 102] < 1, $\eta^2 = .01$), and phonemes, F (2, 102) $= 1.03$, $p = .36$, $\eta^2 = .02$. There was an interaction between experiment and recap for words recalled, F (10, 510) $= 2.82$, $p < .05$, $\eta^2 = .05$ (F [10, 510] $= 2.39$, $p < .01$, $\eta^2 = .05$), syllables recalled, F (10, 510) $= 2.08$, $p < .05$, $\eta^2 = .04$ (F [10, 510] $= 2.05$, $p < .05$, $\eta^2 = .04$), and a borderline interaction for phonemes recalled, F (10, 510) $= 1.61$, $p = .10$, $\eta^2 = .03$. Post hoc ANOVAs using a Bonferroni corrected α of .017 at each level of the dependent measure that displayed a significant interaction (i.e., word and syllable) revealed that the effect was driven by a steeper learning curve in Experiment 1 compared to Experiment 3 and Experiment 3 compared with Experiment 2 for words, Fs (5, 350) > 3.20, ps $< .01$, η^2s $> .04$ (Fs [5, 350] > 3.23, ps $< .01$, η^2s $> .04$) and a marginal effect for Experiment 1 compared to Experiment 3 for syllables, F (5, 350) $= 2.22$, $p = .05$, $\eta^2 = .03$ (F [5, 350] $= 2.13$, $p = .06$, $\eta^2 = .03$). Experiments 1 and 2 did not differ significantly for either words or syllables recalled, Fs (5, 350) < 1, η^2s $< .01$ (Fs [5, 350] < 1, η^2s $< .009$).

It is critical that the effect sizes of the experiment factor are small and we found no interaction between experiment and condition, Fs (2, 102) < 1, η^2s $< .008$ (Fs [2, 102] < 1, η^2s $< .005$), and no three-way interaction between experiment, condition, and recap, Fs (10, 510) < 1, η^2s $< .02$ (Fs [10, 510] < 1, η^2s $< .02$).

LONG-TERM LEARNING. No main effect for the experiment was found in the $3 \times 2 \times 2$ ANOVAs conducted on the accuracy scores for words, F (2, 102) $= 1.40$, $p = .25$, $\eta^2 = .03$ (F [2, 102] $= 1.31$, $p = .27$, $\eta^2 = .03$), syllables, F (2, 102) $= 1.14$, $p = .32$, $\eta^2 = .02$ (F [2, 102] $= 1.13$, $p = .33$, $\eta^2 = .02$), and phonemes, F (2, 102) $= 1.33$, $p = .27$, $\eta^2 = .03$. In addition, there were no interactions between experiment and recap for words, F (2, 102) $< .1$, $\eta^2 = .002$ (F [2, 102] $= 1.53$, $p = .22$, $\eta^2 = .03$), syllables, F (2, 102) $= 1.11$, $p = .33$, $\eta^2 = .02$ (F [2, 102] $= 1.27$, $p = .29$, $\eta^2 = .02$), or phonemes recalled, F (2, 102) $= 1.90$, $p = .16$, $\eta^2 = .04$. Again, we found no interaction between experiment and condition, Fs (2, 102) < 1, η^2s $< .006$ (Fs [2, 102] $= < 1$, η^2s $< .004$), and no three-way interaction between experiment, condition, and recap, Fs (2, 102) < 1, η^2s $< .009$ (Fs [2, 102] < 1, η^2s $< .007$). These results, in conjunction with the short-term learning results, suggest that collapsing across experiments is warranted and that no problematic differences between experiments exist. The meta-analysis may be introducing some further error variance into the ANOVAs that we are reporting, but they do not involve the between-subjects manipulation and we can still be confident that we are achieving gains in power by collapsing across these three experiments. The result of this averaging is depicted in Figure 4 for each of our three measures.

Words. The $3 \times 2 \times 6$ ANOVA found a main effect for recap, F (5, 510) $= 79.17$, $p < .01$, $\eta^2 = .42$ (F [5, 510] $= 78.18$, $p < .01$, $\eta^2 = .43$), no main effect of condition, F (1, 102) < 1, $\eta^2 = .0005$ (F [5, 510] < 1, $\eta^2 = .00001$), and no interaction effect between recap and condition, F (5, 510) < 1, $\eta^2 = .0034$ (F [5, 510] < 1, $\eta^2 = .002$). Averaging across all three experiments, the overt condition

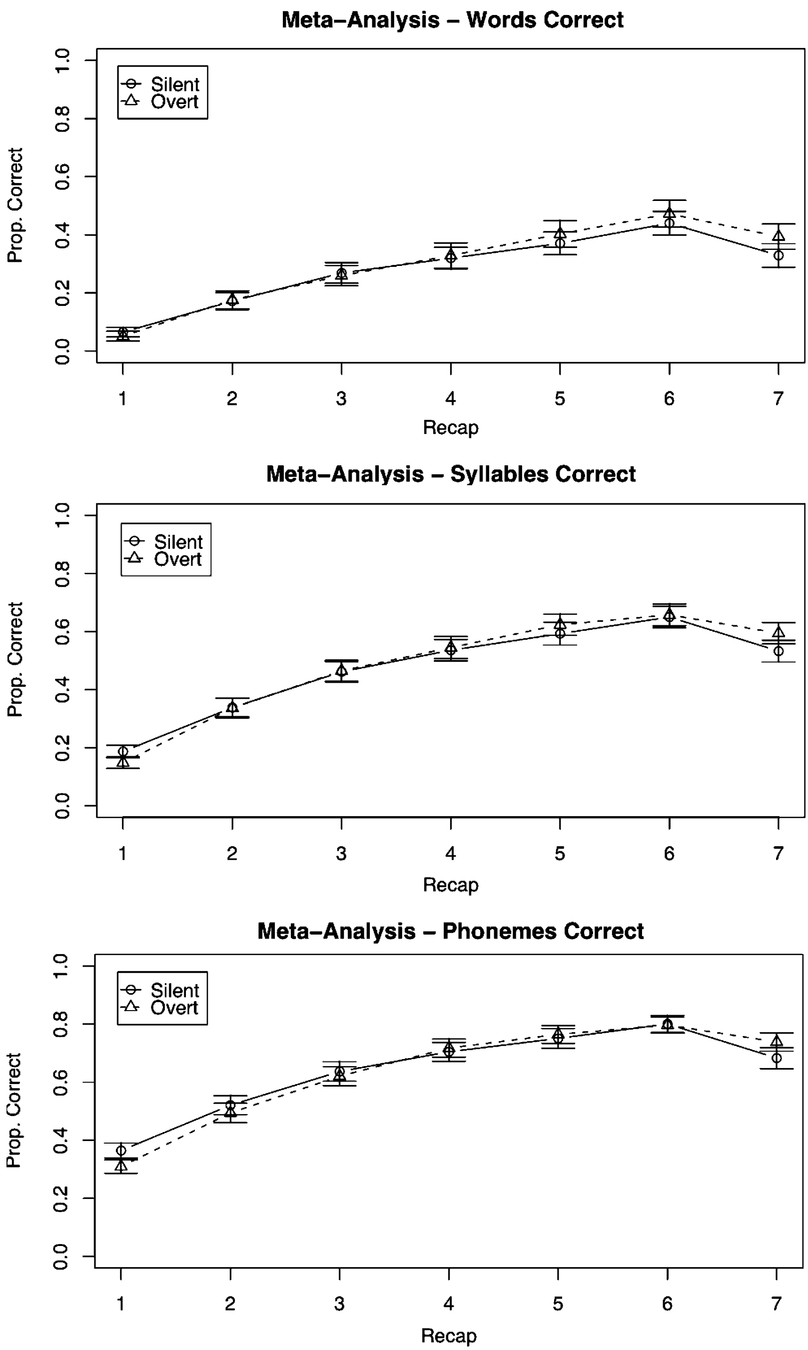

Figure 4. The overall performance for participants averaged across all three experiments ($N = 108$). Performance is plotted separately in terms of (top) words, (middle) syllables, and (bottom) phonemes correct. Error bars are standard errors.

recalled 47.2% ($SD = 33.9$) of the words at Recap 6, whereas the silent condition recalled 44.0% ($SD = 29.9$).

The 3 × 2 × 2 ANOVA showed a main effect of recap, F (1, 102) = 19.64, $p < .01$, $\eta^2 = .16$ (F [1, 102] = 16.23, $p < .01$, $\eta^2 = .14$), no effect of condition, F (1, 102) < 1, $\eta^2 = .01$ (F [1, 102] < 1, $\eta^2 = .004$), and no interaction effect, F (1, 102) < 1, $\eta^2 = .01$ (F [1, 102] < 1, $\eta^2 = .01$). At Recap 7, the overt condition recalled 39.4% of the words ($SD = 32.1$) and the silent condition recalled 32.9% ($SD = 29.9$).

Syllables. The 3 × 2 × 6 ANOVA found a main effect for recap, F (5, 510) = 150.62, $p < .01$, $\eta^2 = .58$ (F [5, 510] = 149.82, $p < .01$, $\eta^2 = .60$), no main effect of condition, F (1, 102) < 1, $\eta^2 = .0001$ (F [1, 102] < 1, $\eta^2 = .58$), and no interaction effect between recap and condition, F (5, 510) < 1, $\eta^2 = .005$ (F [5, 510] < 1, $\eta^2 = .0000007$). At Recap 6, the overt condition recalled an average of 65.7% ($SD = 27.4$) of the syllables and the silent condition recalled 65.0% ($SD = 27.1$).

The 3 × 2 × 2 ANOVA showed a main effect of recap, F (1, 102) = 24.45, $p < .01$, $\eta^2 = .19$ (F [1, 102] = 23.89, $p < .01$, $\eta^2 = .19$), no effect of condition, F (1, 102) < 1, $\eta^2 = .005$ (F [1, 102] < 1, $\eta^2 = .004$), and no interaction effect, F (1, 102) = 2.19, $p = .14$, $\eta^2 = .02$ (F [1, 102] = 2.54, $p = .11$, $\eta^2 = .02$). At Recap 7, the overt condition recalled 59.4% of the syllables ($SD = 27.1$) and the silent condition recalled 53.2% ($SD = 27.5$).

Phonemes. The 3 × 2 × 6 ANOVA found a main effect for recap, F (5, 510) = 191.48, $p < .01$, $\eta^2 = .64$, no main effect of condition, F (1, 102) < 1, $\eta^2 = .001$, and no interaction effect between recap and condition, F (5, 510) = 1.09, $p = .36$, $\eta^2 = .01$. Averaging across all three experiments, the overt condition recalled 79.7% ($SD = 20.5$) of the phonemes at Recap 6, whereas the silent condition recalled 80.1% ($SD = 21.5$).

The 3 × 2 × 2 ANOVA showed a main effect of recap, F (1, 102) = 38.13, $p < .01$, $\eta^2 = .26$, no effect of condition, F (1, 102) < 1, $\eta^2 = .004$, and an interaction effect, F (1, 102) = 4.34, $p < .05$, $\eta^2 = .04$. At Recap 7, the overt condition recalled 73.8% of the phonemes ($SD = 32.1$) and the silent condition recalled 68.3% ($SD = 26.6$).

Criterion analysis. Given the power of this meta-analysis, more reasonable adjustments to alpha level can be made; however, such adjustments change the results reported above in only one case. The adjusted α level for the main effect of the manipulation on learning is .38 and the interaction between condition and learning is .22. Such adjustments would not change the results in terms of words, syllables, or phonemes correct. Thus, we have evidence for accepting the null hypothesis that overt repetition has no more than a small effect on learning in this paradigm. In terms of retention, the adjusted α level for the main effect of the manipulation is .61, and the adjusted level for the interaction is .22. Such adjustments would not change the results of these tests in the case of the main effect. It would change the results of the tests of the interaction using syllable and phoneme levels of accuracy, but not whole words. Thus, we can accept the null hypothesis that overt repetition has no more than a small effect on the retention of words, but may have a small (between .10 and .14) effect on the retention of syllables and phonemes.

Discussion

The meta-analysis reveals that across all three experiments, the silent condition is clearly demonstrating learning for each of our three measures. However, across all three experiments, the overt condition is nearly indistinguishable from the silent condition. Furthermore, there is clearly enough learning during the exposure portion of the experiments and forgetting during the task filled delay to reveal differences between repetition conditions in acquisition or retention, if such differences existed. This is evidenced by the strong main effects of recap in the $3 \times 2 \times 6$ and $3 \times 2 \times 2$ ANOVAs. What is not revealed is a significant difference between the overt and silent conditions in either the overall learning or retention, or the rate of learning or forgetting. The one difference in the rate of forgetting that does emerge is in terms of phonemes recalled, with little indication that this difference affects the total number of whole syllables or whole words recalled. Further, this effect was not exacerbated by the longer retention interval imposed in Experiment 2. Both of these facts suggest that the functional significance of this effect is limited.

Thus, we conclude that expressive WL can occur without overt repetition, and that overt repetition is therefore not critical to expressive WL as it is operationalized in this task. Experiments 1–3 and the subsequent meta-analysis indicate that overt repetition is not critical. In addition, all three experiments and the meta-analysis indicate that overt repetition is not beneficial either. The last question that we ask is whether or not a purer measure of expressive word form learning, one that does not include any semantic information, might reveal a critical role for overt repetition. It may be that the additional semantic component (i.e., mapping the phonological form to the alien) of our expressive word-learning task provides supplementary, or alternative, means to learning in the silent condition.

Accordingly, we conducted two further experiments employing a purer form of phonological learning. It may be the case that the components of the WL task that had participants learn the meanings that are given to the nonwords (i.e., the aliens) is masking an effect of overt repetition in the WL task. A task whose variance would be attributable to only phonological word form learning and that focuses solely on learning the sequence of phonemes associated with a new word is an nonword repetition (NWR)-priming task.

This method is discussed and offered as a paradigm for studying word form learning by Gupta and colleagues (Gupta & Cohen, 2002; Gupta & Dell, 1999). In this paradigm, the participant is exposed to a set of novel word forms multiple times, repeating each one immediately after its presentation, which is the same as in the expressive WL paradigm of Experiments 1 to 3. Unlike the paradigm used in Experiments 1 to 3, however, there is no referent to which the novel word is to be attached. Learning in this task thus incorporates a subset of the processing requirements of the paradigm of the previous Experiments: in terms of the discussion in the introduction, the requirements to learn a semantic representation, and to learn the link between the phonological and semantic representations are here eliminated. The learning observed in this paradigm thus constitutes expressive WL that is purely phonological, and without any semantics, and the rationale for use of this task was to examine purely phonological expressive word form learning might reveal an effect of overt repetition.

An NWR-priming task involves the presentation of novel auditory stimuli to participants. With each presentation, participants traditionally overtly repeat the stimulus just as they heard it. Multiple exposures are given by designating a set of the nonwords as "repeat" stimuli and a set as "unique" stimuli. Repeat stimuli occur in every block of presentation, whereas unique stimuli occur only once per experiment. Thus, during the final block of the experiment priming can be measured as the difference in repetition ability (in terms of accuracy) for the repeat stimuli versus the unique stimuli for that block. This learning is referred to as repetition priming. In the present study, we investigate repetition priming for six-syllable nonwords and the effect of overt repetition on such learning.

EXPERIMENT 4

In Experiment 4, we simply aim to establish that priming (measured as a greater increase in overt repetition accuracy immediately following stimulus presentation for repeat stimuli than for unique stimuli) is evident at this length before turning, in Experiment 5, to the question of whether the act of overt repetition is the source of this priming.

Method

Participants. Twenty-four participants were involved in this experiment. Sixteen participants were recruited through a posting on the University of Iowa's JobNet Web site and received $8 for participating. The remaining participants were undergraduates at the University of Iowa and received course credit for participating.

Materials. Stimuli were six-syllable nonwords drawn from the same corpus as the first three experiments (Gupta et al., 2004). Stimuli were again presented through headphones via a Macintosh PowerPC G3 computer using PsyScope (Cohen et al., 1993).

Design and procedure. Each participant was presented with 120 randomly selected six-syllable nonwords. Random selection of stimuli was done independently for each of the 24 participants. Stimuli were grouped into seven blocks of 30 nonwords each. Fifteen of the 120 six-syllable nonwords selected for presentation were randomly designated as the "repeat" stimuli and appeared in each of the seven blocks (see Appendix E for nonwords used as "repeat" stimuli). The remaining 105 nonwords designated as the "unique" stimuli. The unique nonwords were randomly distributed across the seven presentation blocks and were presented only once to each subject.

On each trial, a nonword was presented and a fixation cross appeared on the computer display immediately following the offset of the nonword. This cross remained on the display for 500 ms. Participants were instructed to repeat each nonword as soon as the fixation cross appeared. Both the stimulus presentation and participant responses were recorded on audiotape for subsequent offline scoring of repetition accuracy. The next nonword was presented 6,000 ms after the onset of the previous nonword.

Participants were allowed a brief break and were provided with water to drink between blocks.

Scoring. For repetition accuracy, a strict scoring criterion was used. A syllable was scored as correct if and only if it was repeated at the correct serial position and all phonemes were correct, unreduced, and in the correct serial order. A nonword was scored as correct if and only if each syllable in the response was correct and no additional syllables were appended. We did not analyze accuracy in terms of syllables correct or phonemes correct, as these analyses yielded little increase in sensitivity in Experiments 1 through 3.

Results

A 2 (Stimulus Type) $\times$ 7 (Block) repeated-measures ANOVA indicates a significant main effect of both type and block, F (1, 23) $= 8.74$, $p < .01$, $\eta^2 = .29$ and F (6, 138) $= 3.32$, $p < .01$, $\eta^2 = .13$, respectively. A significant interaction, F (6, 138) $= 2.47$, $p < .05$, $\eta^2 = .10$, further indicates that the change in accuracy across blocks is greater for the repeat stimuli than for the unique stimuli, the latter of which shows little to no improvement across blocks.

Discussion

Experiment 4 demonstrates significant priming can be achieved in an NWR-priming task using six-syllable nonwords as the stimuli to be learned. We turn now to the role of overt repetition in this learning.

EXPERIMENT 5

The logic for Experiment 5 is the same as in the WL tasks, but here the learning is focused solely on the phonological form of the presented stimuli and learning the link between this form and the motor processes needed to execute expression of this form (Figure 5). There are no target–referent mappings to learn; instead, participants must simply learn the sequences of phonemes that make up each of the repeating nonwords presented in this modified NWR-priming task. If overt repetition is either critical or beneficial in learning the phonological representations of new words, then participants who do not experience this repetition should show either no priming or less priming than participants who do overtly repeat following stimulus presentation.

Method

Participants. Thirty-six participants were involved in this experiment. All were undergraduates at the University of Iowa and received course credit in exchange for participation.

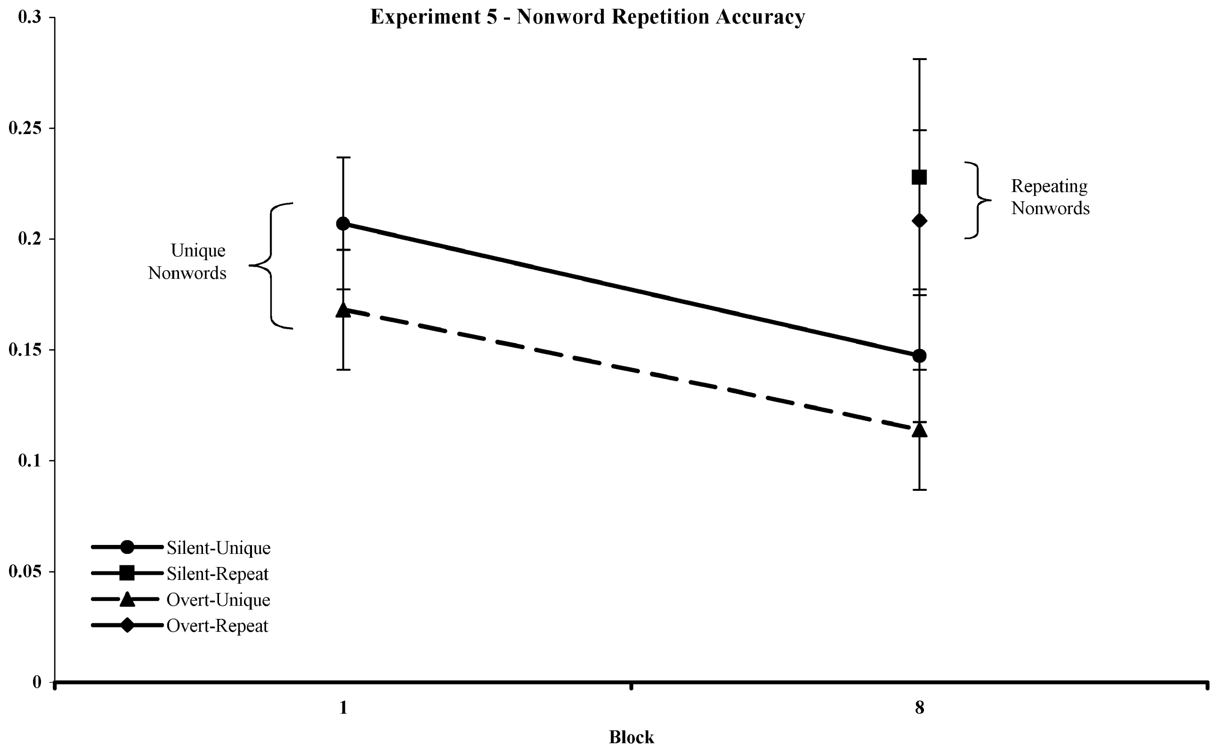

Figure 5. The nonword repetition accuracy as a function of condition (silent vs. overt) and word type (unique vs. repeat). Because of the nature of the experimental paradigm, unique nonwords have data points for Blocks 1 and 8, whereas repeating nonwords only have data points at Block 8. The graph depicts a significant difference between unique and repeat nonwords at Block 8 but no differences between the silent and overt conditions. Error bars are standard errors.

Materials. The stimuli were of the same type as those presented in Experiment 4 and were presented using the same experimental apparatus (see Appendix F for nonwords used as "repeat" stimuli).

Design and procedure. The design and procedure were the same as Experiment 4, except that participants completed eight blocks of NWR, with Blocks 2–8 utilizing an NWR-priming design. In Block 1, all participants were presented with 40 nonwords, all of which were unique stimuli as previously defined. Participants in both conditions completed an NWR task with these stimuli, overtly repeating the nonword as quickly and as accurately as possible following the onset of a fixation cross that appeared at stimulus offset. This provided a baseline NWR measure for participants in both the overt and silent conditions.

Following this first block, participants went through six blocks of an NWR-priming paradigm; however, only half ($N = 18$) of the participants, those who were randomly assigned to the overt condition, actually performed an overt repetition in response to a stimulus presentation. The other half of the participants, those who were randomly assigned to the silent condition, were merely exposed to a stimulus presentation and were instructed to say the word to themselves in their mind without actually producing a sound or moving their mouth. In other words, they were *instructed* to engage in internal repetition or rehearsal. This instruction was given to reduce ambiguity as to the strategy engaged by participants in the silent condition. Thus, in both conditions, participants engaged in articulatory–phonological planning. However, in the overt condition, participants in addition executed the articulatory–phonological plan. Again, there were 40 stimuli within a block, and 29 of these were repeat stimuli and 20 were unique stimuli. The unique stimuli occurred only once, whereas the repeat stimuli appeared once in every block from Blocks 2–8.

In the final block, participants in both the overt and silent condition were instructed to overtly repeat the nonword stimuli upon presentation. This procedure provided a measure of accuracy on the repeating stimuli for the silent condition that was based only on the number of exposures participants in this condition received, not on the ability to overtly repeat the stimulus.

Scoring. For repetition accuracy, a strict scoring criterion was again used. For the overt condition, performance for the unique stimuli in Block 1 as well as the unique and repeat stimuli for Blocks 2–8 was measured. For the silent condition, performance for unique stimuli in Block 1 as well as the unique and repeat stimuli for Block 8 was measured.

Results and discussion

An analysis of the available data for the overt condition's repetition of repeat stimuli (Blocks 2–8) shows improvement on these stimuli across blocks. A 2 (Stimulus) × 7 (Block) repeated-measures ANOVA revealed a main effect of stimulus, F (1, 17) = 22.44, $\eta^2 = .57$, $p < .01$, no main effect of block, F (6, 102) < 1, $\eta^2 = .05$, but an interaction, F (6, 102) = 3.57, $p < .01$, $\eta^2 = .17$. Thus, participants in this condition show gains in accuracy on the repeat stimuli relative to the unique stimuli that lead to the significant interaction. Post hoc contrasts motivated by the significant interaction reveal no difference between Block 2 and Block 8 on the repeat stimuli, $F(1, 102) = 1.43, p = .23, \eta^2 = .10$, but a difference between unique stimuli in the same Blocks 2–8 comparison, $F(1, 102) = 8.95$, $p < .01$, $\eta^2 = .29$. When performance on Blocks 2 and 3 are pooled together as a measure of initial performance and Blocks 7 and 8 are pooled together as a measure of final performance then a significant effect is uncovered for the repeat stimuli, $F(1, 17) = 4.18, p < .05$. The same pooling also strengthens the effect for the unique stimuli, $F(1, 17) = 10.52, p < .01$. Thus, priming was observed in the overt condition as in Experiment 4.

The question of primary interest, however, related to learning in the silent condition relative to the overt condition. To address this question, we examined Block 8 NWR accuracy in a 2 (Condition) × 2 (Stimulus) repeated-measures ANOVA, which revealed no main effect of condition, $F(1, 34) < 1, \eta^2 = .007$, a main effect of stimulus, $F(1, 1) = 21.05, p < .01, \eta^2 = .38$, and no interaction, $F(1, 34) < 1, \eta^2 = .004$. Thus, in Block 8, repeating stimuli were repeated more accurately than unique stimuli, this difference did not differ for the silent and overt conditions, and there was no main effect of silent versus overt. That is, overt repetition did not confer any repetition accuracy benefit compared with silent repetition.

GENERAL DISCUSSION

On the question of whether expressive WL can occur without overt repetition, Experiments 1–3 and a meta-analysis of these experiments indicate that it can, whereas Experiment 5 extends this finding to phonological learning where no target–referent pairing is necessary. Thus, overt repetition (specifically, execution of the articulatory–phonological plan of a word form) does not appear to be critical to expressive WL (specifically, to the learning of the internal phonological representation). The results of Experiment 1 and the meta-analysis further indicate that any benefit for overt repetition would be in the retention of information about the individual phonemes of a new word, but not in any larger conglomeration based on phonetic information (i.e., syllables or words). Experiment 2 demonstrates that any differences in the retention of phonemes disappear over longer periods of time that may be more important time scales for assessing language acquisition. This finding, plus the isolation of any effect to the phonetic level, questions the *functional* significance of any observed differences between the overt and silent conditions of these studies. Experiment 3 neutralizes the role of immediate repetitions of a stimulus in terms of aural benefits by showing that participants in the silent condition show no learning benefit to two exposures of a nonword stimulus within a very short time period. Last, Experiment 5 validates the conclusions from the WL paradigm and extends them to a purer case of phonological learning in an NWR-priming study, which does not include any information about the meanings of nonwords, only their phonological forms. Experiment 4 simply established the paradigm of Experiment 5 as one that has a learning component, and can thus be used to evaluate differences in learning between groups.

How do these results compare with those of previous studies? In Siebert's (1927) investigation of French and English vocabulary pairings, he reported no benefit for repeating a list of paired associates aloud during learning at the first test of production accuracy (measured by written, not spoken, response). The present findings are consistent with this, showing learning without overt repetition, and no benefit of overt repetition, but extending the finding to spoken responses. Siebert also found a benefit of overt repetition at delayed tests of accuracy and savings for later relearning. Although a similar result was obtained in the present Experiment 1, this did not replicate in Experiment 2. What leads to the difference? There are at least three differences between the learning situations. First, the studies differed in novel word presentation modality (visual in Seibert, auditory in the present study). Second, they also differed in novel referent presentation modality: Siebert's study asked participants to associate the novel word from to its native-language translation equivalent, rather than to a visual referent, as in the present experiments. Third, the studies differed in response modality (written vs. spoken).

We suggest that the key difference may be the use of verbal versus visual referents. This is suggested by the findings of Duyck et al. (2003; Experiment 2), who showed that articulatory suppression had an effect on the short-term acquisition of vocabulary (suggesting that articulatory–phonological rehearsal was likely to play a beneficial role) when the task was to learn word–nonword pairs (i.e., verbal referents, as in Seibert's 1927 study), but had no effect when the nonwords themselves, had an association with a visual referent (a situation more akin to the present experiments). This finding suggests that articulatory–phonological planning, and therefore also overt repetition, may be beneficial when verbal referents are used, but not when visual referents are used. Furthermore, Duyck et al. (2003, Experiment 3) obtained the same result in adolescents employing auditory presentation of the novel words. This suggests that presentation modality is not the key difference differentiating Seibert's results from ours. It is, in addition, possible that the findings may differ as a result of the difference in response modalities: with written responses, overt repetition may add a delayed cross-modal benefit, but no such cross-modal benefit accrues for spoken responses. These hypotheses are consistent with the benefits of articulatory–phonological planning reported by Ellis and Beaton (1993) and Papagno et al. (1991) in studies that employed written or typed responses rather than spoken responses and used verbal rather than visual referents, again suggesting that any benefit of overt repetition may accrue only with written responses and/or without a visual referent.

Participants in all of the studies discussed thus far (including the present one) are normal, college-age adults with roughly 18 or more years of experience producing words in their native language. The present stimuli maintain phonological and phonotactic regularities within the participant's native language and this concordance could significantly reduce the learning problem faced by the language processing system and weaken the role of overt repetition as a mechanism for gaining general experience with one's language and the articulatory processes associated with expressive language. Thus, although many studies have used college-age adults to investigate the issues addressed by these experiments, it may be that the issue is a more prominent one in younger children who are actively engaged in the language acquisition process or older children acquiring a language that uses an unfamiliar phonology. Although the previously cited studies did examine the acquisition of foreign languages, most had relatively similar phonologies to participant's native language and/or have the difficulty that no spoken responses were ever required of participants.

Overall, our data suggest that overt repetition (and in particular, execution of articulatory–phonological plans) is not critical to establishment of phonological representations in expressive WL; that is, that the acquisition of new words to the point that they can be recalled and produced when cued with semantic information, or immediately following their presentation, can be achieved without such overt repetition. This conclusion holds true whether both the phonological form and the meaning of a word are to be acquired or just the phonological form is being acquired. Moreover, overt repetition does not appear to confer any benefit. This latter conclusion, however, may apply primarily to the situation where expressive learning is gauged through spoken rather than written or typed responses, and where the referent of the novel word is visual rather than verbal. This situation is reasonably similar to that of much preliterate WL. In situations where the expressive learning involves referents that are translation equivalents without visual support and/or where the response is written rather than spoken (as is typical in second language instruction), overt repetition may, indeed, confer a benefit, as has been widely supposed to be the case for second language learning.

APPENDIX A

STIMULI FROM EXPERIMENT 1

The stimuli used in "target" sets are in bold.

babitade
balimot
balugake
banareke
basodote
batoken
bebinege
bedesor
begalif
bemasute
besunete
bimopun
binasig
birasin
bisirel
bodedite
bofogad
bonikak
bonitet
borokite
dafetak
dagadabe
daketane
dakotul
daselipe
dasotut
daterene
davekik
dedalete
degarite
dekerume
delanose
dipanen
disasus
disokebe
ditomis
donakebe
donaner
doninene
dukalede
dulenan
dutenet
gakilire
gedeneg
gefirase
gemelole
genovake
gesinefe
getedeb
ginatip
gisabif
gisunin
gitotis
kadoneke
kafasene
kafiron
kanesav
kanesuve
kekabage
kekarane
kepibop
keratot
kerilos
kevarate
kikirek
kiledege
kilorir
kinatet
kirigede
kirikod
kisumepe
kitetul
kiturun
kodigil
kogolil

komasol	porodeg	teguvok
kufesade	potapofe	telunar
pasakese	potedele	tereben
patalit	pumadose	tesokane
pelugeke	**putonite**	tesuruk
penedeke	tadasibe	tetitile
pesadeb	tademal	tilisek
pesidide	taganofe	tirenele
pilesit	**tagukit**	tisenen
pinopes	takudek	toginon
pirogege	tapudan	toveruk
pisisope	tavasad	tovupus
pokibote	tebidov	tovutoke
pomikog	**tedanale**	tudemar
ponetase	**tefelose**	tugisuke

APPENDIX B

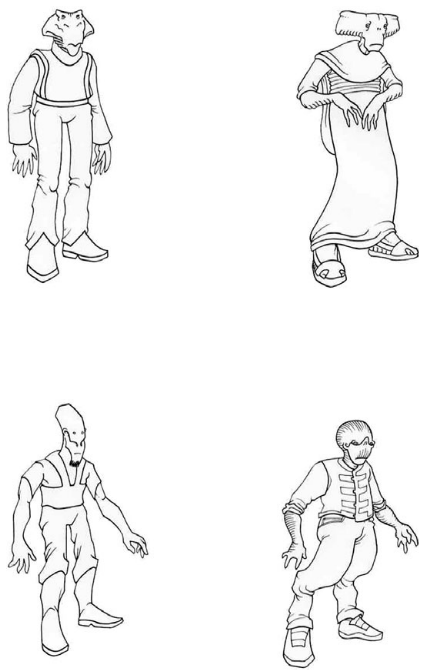

Figure B.1. Four "target" alien images from Experiment 1.

APPENDIX C

STIMULI FROM EXPERIMENT 2 DAY 1

The stimuli used in “target” sets are in bold.

balevel	ditivese	kukirete
balugake	divigure	kukurene
banareke	doganov	kuresen
batikude	doninene	pamutet
bebinege	**dorageg**	papages
bekinore	dosetote	pekigok
bemasute	**dosisis**	pelugeke
benerad	dotebat	penedeke
benutite	duniseb	petafud
besunete	dupodofe	pevanaf
binenule	dusenane	pevotode
bipetog	gataton	pifekete
bisateke	gatofob	pirogege
bisefem	gedeneg	pisitel
bodenose	gemelole	ponulip
bofenene	genetik	popison
bofogad	genonuf	potonon
bokonip	gepagos	putonite
bolutol	gepiroke	tadasibe
bomisul	gibanoke	tademal
bonitet	gitotis	tafadus
bosapen	gosefer	**takudek**
bosurike	**kakikot**	tedanale
botunoge	kalemese	telunar
bukesede	kamutin	tenosan
busagite	kariride	tereben
dalovet	kedosil	tesokane
darusete	kegulol	**tesuruk**
daselipe	**kekesuke**	**tetevase**
dedalete	kenatap	tilevone
dedilus	kepefal	**tinosoke**
degoruge	kerilos	tiremun
dekedol	kesatep	togogeme
dekerume	kesonet	tokarene
deketig	kiledege	**toreret**
dekutan	kinatet	totifite
desafene	komasol	totikore
desudore	konanope	tudemar
detasene	konatot	tukogem
dimaset	korumene	tutefele

APPENDIX D

STIMULI FROM EXPERIMENT 3

The stimuli used in “target” sets are in bold.

bafoted	dofenis	**kosirime**
bakunel	doganov	kosotile
baliseb	donakebe	kukirete
banisite	donaner	kurotit
batitene	**doninene**	**papages**
batoken	dosekase	pebasap
bebinege	dosetote	pevanaf
bekinore	dosisis	pevotode
benisure	**dukalede**	pifekete
besunete	dupodofe	pisamov
betenege	dusimase	**pomikog**
bifetet	dutenet	ponulip
bimopun	galanan	potutom
binelud	gatofob	putonite
binenule	getedeb	tasibore
binonere	**gevutide**	tavakes
bipetog	gitotis	tavasad
birasin	kadoneke	tefelose
bisirel	kafiron	tefilib
bisonir	kagapone	teguvok
bofogad	kalemese	temereme
bonikak	kamutin	tenosan
bosalere	katepite	tesetit
bosurike	katigal	tesokane
bulinep	kedosil	tesuruk
busagite	**kekabage**	tetevase
dabibup	kekarane	tevadeg
dagadabe	kepibop	tibefose
daterene	keratot	tiremun
davatas	kerilos	tirenele
debusiv	kesonet	**tisatef**
dekutan	kiledege	tisenen
delanose	kilorir	tokarene
delesep	kirigede	tokatare
desudore	kisumepe	**tovupus**
dimasat	kitetul	tudemar
disasus	kobepage	tufaned
disokebe	kogolil	tugisuke
ditomis	konikale	tukogem
dodolek	konovose	tutefele

APPENDIX E

STIMULI USED IN "REPEAT" SETS IN EXPERIMENT 4

badetarerikik
badidadodukik
bakakitasemud
bakerematinem
bakidinetorope
balavodisabor
balidarutifad
bamekikotepav
baminenanonad
banatelobunid
banosomasalem
banumerenesane
bapetigatinat
baripenalisute
batisiritafabe
befonemovibaf
begaditosemade
begotepetesas
begotetonafuk
bekevesenebeke
bekotivepasen
belukelefises
benitipanaret
bepisafosorif
beraronefarake
beselatagorar
besitanedatem
besotokamipufe
betebetudenofe
betitalenotine
bevesesekamute
bevotopulimep
bidovekakasek
bikagometesive
bikamumorolese
bikativinisime
bilegepiteseg
bilotatimeted
bimekanigarote
binesetesodes
binikemulosod
bipibunosenese
biregonategife
biretukekekote
bisekevedikak
bisetegekirite
bisorelaniboke
bitanelirolak
bivekotatotos
bofomamivilade
bogobavikitin
bokedalamipev
boketesikesale
bokikoletinef
bokogitenuvet
bolesitenerir
bolimebagotem
bolipifafodore
bonananimileme
bonosogananed
boralenerisan
bosanutigigese
bosedesisinipe
bosedigamodim
bovenotadinil
budisetelebone
budupisedukos
bunenedesivise
bupotosanenet
dabemuvulanet
dadasetelelole
dadikakitarase
dafonipisomot
dagelobetabet
dagetekununate
dakesesotosek
dakinisedolete
dakotetatasip
dakupadivekek
danirepulurat
daresefaliren
dasalofelatik
dasegorenedale
dasetolakaken
dasivunitudase
datarikitenur
datovudatitek
davokisesoguge
dedasesosotit
dedesonenenet
dedulenelumug
defusebonitag
degasunetekake
degogerabutet
dekanomegotupe
dekedidegagape
dekediderakete
delonimasakid
delumenafegos
denetivoketure
denitekinonofe
denodanukidase
depunudagekik
desasenagibep
desiredoperute
desodenatolade
detelefinisan
detusedulilebe
diditivavemite
didutedepitate
digidetegefase
dikilokelalin
diritonononate
disekotakakare
disesemerinise
disolekimapan
ditalofavoras
ditasenuredos
ditoresogilike
ditotesaneteke
divunadefapane
dodasesikeron
dodesogebonen
dofuvimesisote
doketekosenome
dolarobatelose
domanumonalin
donosenesanut
doretedonisan
dorodanunatase
dosasamodatade
doserarosepuk
dotiditinakake
dovikonetupese
dudatotitigat
dutifotodotute
gabasefepelede
gagisetonanir
gamikenefapeke
ganitosavekos
gasitenilisere
gategalosesas
gatodinunidove
gavebogavanek
gedarudugetap
gekipireninat
geletekeselate
gemokasatatan
gemuvetinonoke
genefisodifone
gepelatovabek
gepolomabilute
gerotefuvulene
gesanotisesip
gevakutetonofe
gifenokuvidive
gikakemekopek
giletimosemate
gimekofenamod
gisanebililut
gisetalakenes
gitonetitikes
godapibaroteb
godosirometire
gogirivaresose
gomimomariran
gonedemokubev
gopatanitafave

gosetidikodele
gusutiporoneke
gutanolunetise
gutesofegipoke
kadafofusafeg
kakedatabesif
kalasatatasik
kaleropevasete
kamefekekaton
kanesekunedas
kanokegarasole
karotovadikene
kasalusesetil
kasasasesudag
katutosedetoke
kebetulesisate
kegitedonatot
kekofemifurame
kelegelefemef
kemonumerunoge
kenefafenidele
kepanasekelule
kepegemadipel
kepotekisodin
kerefisonuruse
kerositumidise
kesidubemikose
kesinififadede
kesubitoporase
ketesoseporole
ketibetetasit
ketileponoguge
keturipavedaf
kevanolevevike
kiderikidokon
kikegelegidev
kikilesevenem
kikosikosulig
kikosotedunune
kilelakomomige
kilidunudoves
kilitidonegus
kimatitegutate
kinapogedasaf
kinelasesilav
kinonoburoseme
kipoderutavib
kiridevuvatepe
kiriditeteteb
kirilekadasede
kirimutirelole
kisatinotanis
kisinekepoden
kisononigeset
kitidokenubet
kitolupugeses
kobetevafefat
kobomatakelot
kofukogikukafe
kokinadakineme
kolesasosatab
konerekipetak
kopireribetim
korogodesikede
kosaketakutog
kotasogusadame
kotidomoverade
kotumutafapak
kotusedakelif
kovotunerukere
kovuredemedose
kudunokakefete
kunekotanisale
kuneragikitap
kunoratinepede
kusulevogunir
padegeredinis
padevenorakag
padilenenunole
panemalakilek
parevadikokese
patebenisatode
pedoleluraregе
pegebulamekap
pegegelutesege
pekikipebarele
pekusadavogik
pelisibevemog
pelivitidasop
peloninemiker
pesatatevogupe
pesosanuresepe
pesugamokoneke
petobigididas
pidatatakutase
pidikudovenus
pidinabatitob
pigadonatonule
pigidosenekese
pimetonekedud
piridaledokok
pirisepivovade
pitamesefarik
pivakesusaseme
pogigusarorube
polasudasonin
pomanikorenule
porekitalevuge
posobasogufile
pugetisitibuke
purinakusaviv
puronemifisas
tabokadakitel
tabokedobemame
tabomitafosone
tadukenokapon
tagiditerenun
takikelivenese
takilununedif
tamenonesoleg
tanobefototine
tapekepodatune
tapenodutifame
tasanituregus
tataketusinote
tatanesurunike
tatavepikanuk
tatukisinisige
tebovukokipil
tededuvupetere
tedoserutokar
tedotolekusane
tefevoferibete
tekuselotaton
telititagarife
tenakotetanoke
tenananebasis
tenivorilekeg
tenufinudinube
terenatomimede
tereseriferese
terokesisisor
tesadonekadate
tesakalolelone
tesesumetanet
tetadutotunom
tetegatudodir
tetekadikepit
tetitatanufune
tetoditinipis
tetomelanenuse
tevigenemedem
tidegatikegote
tidosokamepen
tinaderomubove
tinadokefelome
tiradigenitote
tiredagaselipe
tisodakekotes
titekopimireve
titenekekimefe
titosagesagote
tobolululolot
todepimositek
tofasanonegure
tokenigosarik
tokigokaveten
tonebunanamer
tonegudetirol
tonemiloresum
tonimadunesad
toredekisakeve
tosanarudunam
tosonenukidote
totakiseriken
totalosanudibe
tudikoseseden
tukibepikasel
tunamudabenore
tunokitapalet
tusapedonilele
tusotanopisole
tusubipanasen

APPENDIX F

STIMULI USED IN "REPEAT" SETS IN EXPERIMENT 5

badidadodukik	kadafofusafeg
banumerenesane	kepotekisodin
bekevesenebeke	kesidubemikose
bekotivepasen	ketesoseporole
besitanedatem	kiderikidokon
betebetudenofe	kikosikosulig
bikamumorolese	kinelasesilav
bitanelirolak	kinonoburoseme
bivekotatotos	kirimutirelole
bolimebagotem	kofukogikukafe
bolipifafodore	konerekipetak
bosanutigigese	kovuredemedose
dafonipisomot	panemalakilek
dagelobetabet	pekikipebarele
dakupadivekek	petobigidadas
dasalofelatik	pidikudovenus
datovudatitek	pirisepivovade
davokisesoguge	polasudasonin
desasenagibep	tabokadakitel
ditotesaneteke	tagiditerenun
dodesogebonen	tapekepodatune
domanumonalin	tatanesurunike
doserarosepuk	tatavepikanuk
dovikonetupese	tebovukokipil
gasitenilisere	tefevoferibete
gategalosesas	tekuselotaton
gemokasatatan	tenakotetanoke
gemuvetinonoke	tetitatanufune
gonedemokubev	tokigokaveten
gopatanitafave	tonegudeti

ACKNOWLEDGMENTS

The authors thank Sara Evan, Charles Geilenfeld, Jon Hintz, Melissa Hodapp, John Lipinski, Byron Murphy, Anna Napawan, Sierra Spies, Liz Tener, and Matthew Woodin for their assistance in setting up and running the experiments presented here and members of the Language Discussion Group at the University of Iowa for helpful discussion of the results. Experiments 1 and 3 were previously presented at the 46th Annual Meeting of the Psychonomic Society, Minneapolis, MN. This research was partially supported by Grant NIH R01 DC006499 (to P.G.).

REFERENCES

Atkinson, R. C. (1975). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology, 96*, 124–129.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press.

Baddeley, A. D., & Wilson, B. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language, 24*, 240–502.

Burgess, N., & Hitch, G. (2005). Computational models of working memory: Putting long-term memory into context. *Trends in Cognitive Sciences, 9*, 535–541.

Carey, S. (1978). The child as word learner. In J. Bresnan, G. Miller, & M. Halle (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

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.

Conover, W. J., & Iman, R. L. (1981). Rank transformation as a bridge between parametric and nonparametric statistics. *The American Statistician, 35*, 124–129.

Cumming, N., Page, M., & Norris, D. (2003). Testing a positional model of the Hebb effect. *Memory, 11*, 43–63.

Cunningham, T. F., Healy, A. F., & Williams, D. M. (1984). Effects of repetition on short-term retention of order information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 575–597.

Desrochers, A., & Begg, I. (1987). A theoretical account of encoding and retrieval processes in the use of imagery-based mnemonic techniques: The special case of the keyword method. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processes*. New York: Springer–Verlag.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test III*. Circle Pines, MN: American Guidance Service.

Duyck, W., Szmalec, A., Kemps, E., & Vandierendonck, A. (2003). Verbal working memory is involved in associative word learning unless visual codes are available. *Journal of Memory and Language, 48*, 527–541.

Ellis, N. C., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology, 46A*, 553–558.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (in press). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavioral Research Methods, Instruments, and Computers*.

Gerard, P. D., Smith, D. R., & Weerakkody, G. (1998). Limits of retrospective power analysis. *Journal of Wildlife Management, 62*, 801–807.

Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual: What you always wanted to know about significance testing but were afraid to ask. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 391–408). Thousand Oaks, CA: Sage.

Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning, 46*, 643–679.

Gupta, P. (2005). What's in a word? A functional analysis of word learning. *Perspectives on Language Learning and Education, 12*, 4–8.

Gupta, P., & Cohen, N. J. (2002). Theoretical and computational analysis of skill learning, repetition priming, and procedural memory. *Psychological Review, 109*, 401–448.

Gupta, P., & Dell, G. S. (1999). The emergence of language from serial order and procedural memory. In B. MacWhinney (Ed.), *Emergence of language*. Hillsdale, NJ: Erlbaum.

Gupta, P., Lipinksi, J., Abbs, B., & Lin, P.-H. (2005). Serial position effects in nonword repetition. *Journal of Memory and Language, 53*, 141–162.

Gupta, P., Lipinski, J., Abbs, B., Lin, P.-H., Aktunc, M. E., Ludden, D., et al. (2004). Space aliens and nonwords: Stimuli for investigating the learning of novel word-meaning pairs. *Behavioral Research Methods, Instruments, and Computers, 36*, 599–603.

Gupta, P., & MacWhinney, B. (1995). Is the articulatory loop articulatory or auditory? Re-examining the effects of concurrent articulation on immediate serial recall. *Journal of Memory and Language, 34*, 63–88.

Hoenig, J. N., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician, 55*, 19–24.

Horst, J. S., & Samuelson, L. K (in press). Fast mapping but poor retention by 24-month-old infants. *Infancy*.

Klatt, D. (1987). Review of text-to-speech conversion in English. *Journal of the Acoustical Society of America, 82*, 737–793.

Kromrey, J., & Hogarty, K. Y. (2000). Problems with probabilistic hindsight: A comparison of methods for retrospective statistical power analysis. *Multiple Linear Regression Viewpoints, 26*, 7–14.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America, 97*, 11850–11857.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138–1141.

Kuhl, P. K., & Meltzoff, A. N. (1988). Speech is an intermodal object of perception. In A. Yonas (Ed.), *Perceptual development in infancy: Minnesota symposia on child psychology* (pp. 235–260). Hillsdale, NJ: Erlbaum.

Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician, 55*, 187–193.

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language, 30*, 331–347.

Royer, J. M. (1973). Memory effects for test-like-events during acquisition of foreign language vocabulary. *Psychological Reports, 32*, 195–198.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualization of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207–217.

Seibert, L. C. (1927). An experiment in learning French vocabulary. *Journal of Educational Psychology, 18*, 294–309.

Steidl, R. J., Hayes, J. P., & Schauber, E. (1997). Statistical power analysis in wildlife research. *Journal of Wildlife Management, 61*, 270–279.

Studdert-Kennedy, M. (1986). Sources of variability in early speech development. In J. S. Perkell, & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 58–84). Hillsdale, NJ: Erlbaum.

Zumbo, B. D., & Hubley, A. M. (1998). A note on misconceptions concerning prospective and retrospective power. *The Statistician, 47*, 385–388.