American Psychological Association

Journal of Experimental Psychology: General

ISSN: 0096-3445

2020, Vol. 149, No. 8, 1477–1508
http://dx.doi.org/10.1037/xge0000732

Temporal Dynamics of Free Recall: The Role of Rehearsal Efficiency in Word Frequency and Bilingual Language Proficiency Effects

Wendy S. Francis, Marcela M. Arteaga, Mary K. Liaño, and Randolph S. Taylor
University of Texas at El Paso

Under a recall model in which presentations and rehearsals are treated as equivalent encoding events, we investigated whether rehearsal efficiency differences explain the effects of word frequency and bilingual proficiency on the temporal dynamics of rehearsal and free recall. Experiments 1 and 3 were conducted with monolingual English speakers, and Experiments 2 and 4 were conducted with Spanish-English bilinguals with matched age, education, and socioeconomic status. In Experiments 1 and 2, lower word frequency, lower proficiency, and bilingualism were associated with less accurate free recall of items from early serial positions, beginning recall with items from later serial positions, and making fewer transitions to items from later or adjacent serial positions. These effects were replicated and rehearsal-based explanations were validated in Experiments 3 and 4 using a rehearse-aloud protocol. With lower frequency words or lower language proficiency, rehearsal was less efficient with fewer rehearsals between item presentations. As a result, items from early serial positions had fewer rehearsals that stopped earlier in the study sequence, less spacing between repeated rehearsals, and fewer transitions to items from later or adjacent serial positions. Rehearsal-contingent analyses revealed that these rehearsal patterns were associated with less accurate recall, beginning recall with items from later serial positions, and consistent transition patterns from rehearsal to recall. These patterns support a model in which presentations and rehearsals are treated as equivalent encoding events and the effects of word frequency and language proficiency on recall accuracy are mediated by less efficient rehearsal.

Keywords: free recall, associative memory, bilingualism, word frequency

Supplemental materials: http://dx.doi.org/10.1037/xge0000732.supp

Free recall test performance depends on encoding conditions, characteristics of the materials, and characteristics of the learners as well as the strengths of semantic, temporal, and contextual associations formed either intentionally or incidentally during encoding. Free recall is consistently less accurate for low-frequency than for high-frequency word lists, with rehearsal differences long cited as a major contributing factor (e.g., Brodie & Murdock, 1977; Richardson & Baddeley, 1975). Documenting the impact of word frequency on the temporal dynamics of rehearsal and retrieval would enhance our understanding of how item and temporal information are encoded and retrieved, but these phenomena are relatively unexplored (but see Ward, Woodward, Stevens, & Stinson, 2003). Similarly, language proficiency and bilingualism are often associated with differences in recall accuracy (e.g., Fernandes, Craik, Bialystok, & Kreuger, 2007; Harris, Cullum, & Puente, 1995), but little is known about the temporal dynamics of retrieval, and there is no previous documentation of rehearsal patterns in bilinguals. It is therefore unknown whether these patterns would parallel those observed with word frequency manipulations or whether similar mechanisms are involved.

The present study investigates how word frequency, language proficiency, and bilingual status impact the temporal dynamics of rehearsal and free recall, with a focus on the incidental learning of new temporal associations. We show that rehearsal efficiency varies with these factors and contend that differences in rehearsal patterns contribute substantially to the effects of word frequency, language proficiency, and bilingualism on a range of recall phenomena. We also make the case that treating presentations and rehearsals as equivalent encoding events improves the explanatory power of existing variable-context and retrieved-context approaches to free recall dynamics.

Temporal Dynamics of Free Recall

Our theoretical reasoning is based primarily on the variable context model and the retrieved context framework with the modification that rehearsals are treated as episodes equivalent to externally generated presentations. The present study will probe the utility and added explanatory value of this modified model (see Figure 1). In the following paragraphs, we explain the derivation

This article was published Online First January 9, 2020.

Wendy S. Francis, Marcela M. Arteaga, Mary K. Liaño, and Randolph S. Taylor, Department of Psychology, University of Texas at El Paso.

This research was supported by NIH Grant R15HD078921 to Wendy S. Francis. Results were presented at the 56th and 58th Annual Meetings of the Psychonomic Society. We thank Amaris Soltero and Juan Carlos Etienne for their assistance with data collection and transcription and Erika Guedea for assistance in data processing. We thank Veronica Whitford for her input on the statistical analyses.

Correspondence concerning this article should be addressed to Wendy S. Francis, Department of Psychology, University of Texas at El Paso, 500 West University Avenue, El Paso, TX 79968. E-mail: wfrancis@utep.edu

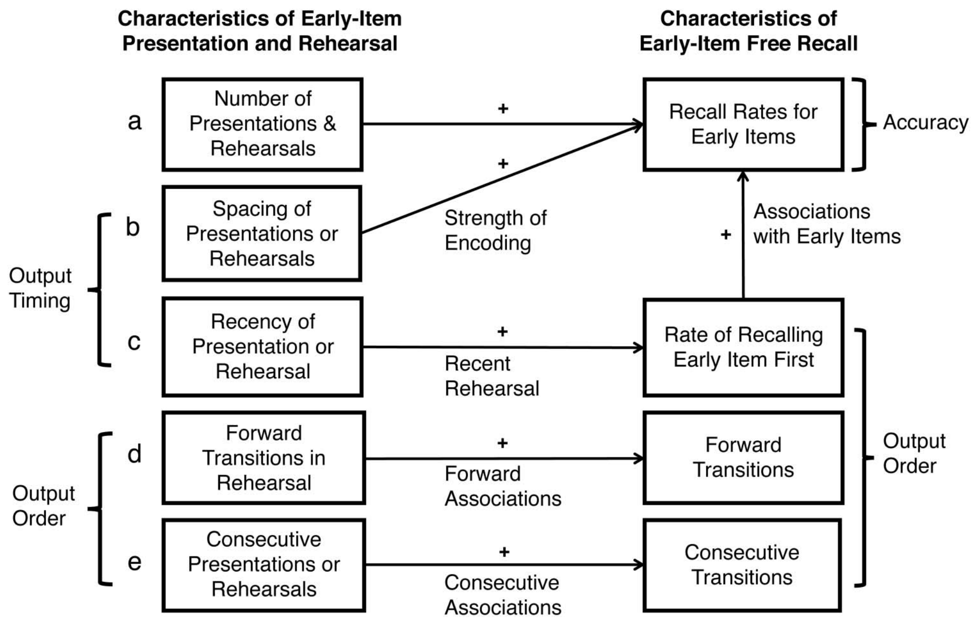

Figure 1. A modified variable context model of free recall. The number (a) and spacing (b) of rehearsals of items from early serial positions are hypothesized to increase accuracy by strengthening encoding. Rehearsing an early item to a later point in the study sequence (c) makes its rehearsal more recent, which is hypothesized to make it more likely to be recalled first, which in turn increases the recall rates of other early items. Rehearsal transitions to later or adjacent items (d, e) are hypothesized to strengthen associations, thus increasing the likelihood that these transitions will be repeated in recall output.

of the model and some related alternatives before turning to the impact of inefficient rehearsal within this model.

Variable Context Models

Documentation of the U-shaped serial position curve goes back to the seminal work of Ebbinghaus (1885). The *primacy effect*, a free recall advantage for items presented at the beginning of a sequence, depends critically on rehearsal (Brodie & Murdock, 1977; Rundus, 1971; Tan & Ward, 2000; Ward et al., 2003). Prevention of rehearsal reduces or eliminates the primacy effect (Bjork & Whitten, 1974; Glenberg, 1984; Howard & Kahana, 1999; Richardson & Baddeley, 1975). The most widely accepted explanations for the *recency effect*, a free recall advantage for items presented at the end of a sequence, come from *variable context* models (Glenberg & Swanson, 1986; Mensink & Raaijmakers, 1988). In these models, contextual features fluctuate gradually over time and therefore have greater overlap for events that occur closer together in time. The recency effect emerges because items in later serial positions are associated with contextual features that overlap more with the retrieval context, making these recent items more likely to come to mind when recall is initiated.

The observation that recall tends to begin with items at or near the end of the list dates back at least to the 1950s (Deese & Kaufman, 1957). Graphs illustrating the relative frequency distribution for the serial position of the first item recalled (PFR) seem to have originated with Hogan (1975). An explanation consistent with the variable context approach is that the initial recall context is used as a cue to retrieve items that were presented in a similar context, which would be items most recently presented (Howard & Kahana, 1999; Lohnas, Polyn, & Kahana, 2015; Polyn, Norman, & Kahana, 2009).

Recent research suggests that a refinement to the variable context model may be warranted, because rehearsals and additional presentations at study appear to be encoded in a similar manner, leading to some recall models that treat presentations and rehearsals as equivalent events (e.g., Laming, 2009; Tan & Ward, 2000). Indeed, examinations of overt rehearsal patterns at study showed that the *recency of the last rehearsal* of an item was a better predictor of its successful recall than its serial position in the presentation sequence (Laming, 2009; Tan & Ward, 2000; Ward et al., 2003). *We propose that the contextual features of both presentations and rehearsals fluctuate over time.* As a consequence, items *presented or rehearsed* near the end of a study sequence would be associated with context that has greater overlap with the test context (see Figure 1c).

The added value of this modification is that the same mechanisms used to explain recency and beginning recall with items from late serial positions can also accommodate the primacy effect and the substantial rate at which recall begins with items from early serial positions. Specifically, items presented at or near the beginning of the study sequence tend to be rehearsed more often and later into the study sequence than items from the middle of the list (Laming, 2008, 2009; Tan & Ward, 2000; Ward et al., 2003), which has two important consequences. First, the primacy effect can be explained as a recency of rehearsal effect (as in Laming, 2009; Tan & Ward, 2000; Ward et al., 2003). Second, the first item presented often makes it all the way to the last rehearsal cycle, putting it in contention to be the first item recalled (see Figure 1c).

Indeed, in one study, 93% of free recall trials were initiated with a word that was rehearsed all the way to the last rehearsal set (i.e., the cycle that included the last item; Rundus, 1971).

Retrieved Context Framework

In an effort to focus attention on the dynamics of retrieval and factors that determine output order in free recall, Kahana (1996) developed a new method to measure transitions among items in recall output and plot conditional response probabilities. Once an item was recalled, the next item recalled was most often an item from an adjacent serial position, a pattern attributed to the high degree of contextual overlap between items presented consecutively (e.g., Golomb, Peelle, Addis, Kahana, & Wingfield, 2008; Howard & Kahana, 2002; Kahana, Howard, Zaromb, & Wingfield, 2002). Recall sequences contain more transitions in the forward direction than in the backward direction (e.g., Howard & Kahana, 2002; Laming, 2008, 2009; Ward et al., 2003), meaning that the next word recalled was more often one from a later serial position.

Kahana (1996) later proposed a model in which items in the study sequence form temporal associations that are mediated by their partially shared temporal context (e.g., Howard & Kahana, 1999; Kahana et al., 2002; Lohnas et al., 2015; Polyn et al., 2009). The greater contextual similarity and stronger associations for items presented in closer serial positions was attributed to the gradual context fluctuation cited in the variable context models. According to the *retrieved context framework* (Howard & Kahana, 1999, 2002), retrieving a studied word involves partial reinstatement of its temporal context, which in turn elicits the retrieval of other items studied in similar contexts, which are items from neighboring serial positions. With each successive item recalled, the corresponding context is reinstated, and the retrieval process is repeated. Here again, if we consider rehearsals as additional presentations, when the presentation or rehearsal contexts of recalled items are reinstated, *they will cue the retrieval of other items that were presented or rehearsed close together at study* (Figure 1e).

This approach also has implications for the effects of spacing between repeated presentations or rehearsals. According to variable context models, spacing of repeated presentations improves recall performance because spaced presentations have less overlap in their associated contextual features, making them sensitive to distinct sets of retrieval cues (Ross & Landauer, 1978). According to the retrieved context framework, the item develops high contextual overlap with more of the other studied items, giving it more chances to be cued by their reinstated contexts (Lohnas, Polyn, & Kahana, 2011). We contend that greater spacing of rehearsals will exhibit the same consequences for contextual overlap and context reinstatement, leading to increased recall accuracy with *both spaced presentations and spaced rehearsals* (Figure 1b). Indeed, spaced rehearsals do contribute to higher recall accuracy above and beyond the effects of the number of rehearsals and recency of the last rehearsal (Tan & Ward, 2000).

The Role of Rehearsal Efficiency in Free Recall Dynamics

The central overarching hypothesis in the present study is that the effects of word frequency, language proficiency, and bilingualism on free recall performance arise as a natural consequence of their impact on rehearsal efficiency (see Figure 2). Specifically, we propose that rehearsal efficiency, along with rehearsal patterns driven by rehearsal efficiency, mediate the effects of these three factors on free recall accuracy and output order. In this section, we first provide evidence that links these three factors to rehearsal efficiency (Figure 2a–c). Based on the model described in the previous section, we then summarize the hypothesized consequences of inefficient rehearsal for specific rehearsal patterns (Figure 2d–h) and the hypothesized consequences of these patterns

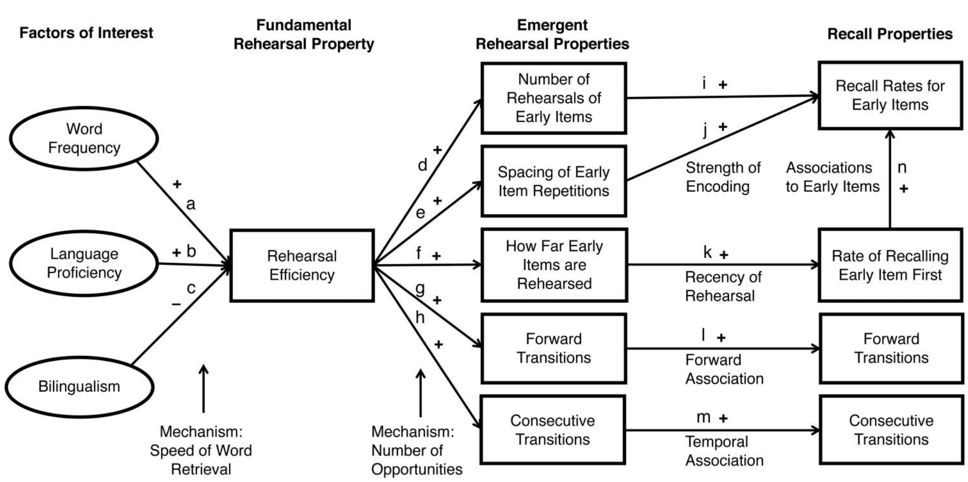

Figure 2. Proposed mechanisms of word frequency, language proficiency, and bilingual status effects on rehearsal and recall dynamics. Word frequency, language proficiency, and bilingual status impact the speed of word retrieval and therefore rehearsal efficiency (a–c). Rehearsal efficiency impacts several measurable properties of emergent rehearsal sequences (d–h). These patterns in turn impact recall accuracy and aspects of recall output order (i–m). When the first item recalled is from an early serial position, other items from early positions are more likely to be recalled (n). Labels on arrows indicate mechanisms, and signs indicate the direction of association.

on recall performance (Figure 2i–n). For each rehearsal or recall property, we make predictions about the effects of the three key factors and review relevant evidence from the limited previous research that addresses these questions.

Factors That Elicit Less Efficient Rehearsal

Rehearsal is more efficient when it is faster and therefore allows more rehearsals to intervene between successive presentations of new items. We hypothesize that word frequency, language proficiency, and bilingualism will affect the efficiency of rehearsal (Figure 2a–c). Rehearse-aloud protocols show that fewer rehearsals intervene between presentations in low-frequency relative to high-frequency word lists (Tan & Ward, 2000; Ward et al., 2003). A plausible explanation is that low-frequency words take longer to retrieve for rehearsal. Indeed, retrieving low-frequency words for production is slower than retrieving high-frequency words (e.g., Gollan, Montoya, Cera, & Sandoval, 2008; Griffin & Bock, 1998; Wheeldon & Monsell, 1992).

By the same logic, we also expect less efficient rehearsal with lower language proficiency and bilingualism, because retrieving words for production in bilinguals is slower in their less proficient language (L2; e.g., Francis, Corral, Jones, & Sáenz, 2008; Potter, So, Von Eckardt, & Feldman, 1984; Sholl, Sankaranarayanan, & Kroll, 1995) and slower in their more proficient language (L1) than in monolinguals (e.g., Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Ivanova & Costa, 2008). (Note that L1 and L2 are used here to denote the more and less proficient languages, respectively, and do not necessarily reflect the order of acquisition).

Emergent Rehearsal Properties

Inefficient rehearsal, which is slower and allows fewer rehearsals between successive item presentations, is hypothesized to constrain several properties of rehearsal patterns, as illustrated in Figure 2d–h. The quality of a rehearsal pattern has been characterized previously by at least two aspects of performance: the number of rehearsals of items from early serial positions and how far into the study sequence these items are rehearsed (Tan & Ward, 2000; Ward et al., 2003). When fewer rehearsals intervene between successive item presentations, the most obvious consequence is that the average number of rehearsals per item will decrease. The decrement is more severe for items from early serial positions, because after the first few items are presented, the incorporation of each newly presented item into its rehearsal set requires not repeating the rehearsal of an earlier item. For the same reasons, these items are not rehearsed as far into the study sequence, making their last rehearsals less recent relative to the test.

As shown in Table 1, rehearse-aloud protocols show that items from early serial positions are indeed rehearsed fewer times and not as far into the study sequence in low-frequency word lists as in high-frequency word lists (Tan & Ward, 2000; Ward et al., 2003). We test for the first time whether the effects of language proficiency and bilingualism on rehearsal patterns will parallel the patterns observed for word frequency, with lower proficiency and bilingualism being associated with fewer rehearsals that do not continue as far into the study sequence.

Less efficient rehearsal is hypothesized to have additional consequences for the rehearsal sequence, because it allows fewer opportunities for spaced repetition of individual items and fewer transitions to items from later or neighboring serial positions. To our knowledge, patterns of spacing and transitions among rehearsals across more and less efficient rehearsal conditions have not been compared in previous research. We predict that the less efficient rehearsal patterns associated with low word frequency, lower language proficiency, and bilingualism will be characterized by less spacing between repetitions, fewer transitions to items from later serial positions, and fewer transitions to items from adjacent serial positions.

Table 1
Summary of Previous Findings

Measure	Word frequency	Language proficiency	Monolingual versus bilingual status
Rehearsal rate	LF < HF	—	—
Rehearsal of items from early serial positions			
Number of rehearsals	LF < HF	—	—
Last rehearsal set	LF < HF	—	—
Spacing of repetition in rehearsal	—	—	—
Forward transitions in rehearsal	—	—	—
Consecutive transitions in rehearsal	—	—	—
Recall of items from early serial positions			
Accuracy	LF < HF	LP < HP	—
Serial position of first item recalled	—	—	—
Functional serial position of first item recalled	—	—	—
Forward transition rate in recall (early)	LF < HF[a]	LP < HP[b]	B2 < M[b]
Consecutive transition rate in recall (early)	LF < HF	—	—
Functionally consecutive transition rate (early)	—	—	—

Note. LF = low frequency; HF = high frequency; LP = low proficiency; HP = high proficiency; B1 = bilingual L1; B2 = bilingual L2; M = monolingual; > and < indicate direction of effect.
[a] This comparison was directly tested only for lists of six to eight items, where participants knew that order could be tested. Ward et al. (2003) used longer lists but did not report this measure; however, this effect can be inferred based on the lag recency functions. [b] Based on evidence from relative order scores in free recall.

Implications for Recall

According to the modified variable context and retrieved context models in which rehearsals and presentations are treated as equivalent encoding events, the rehearsal properties that emerge under conditions with less efficient rehearsal are expected to have consequences for multiple aspects of free recall performance (Figure 2i–n). We hypothesize that these rehearsal properties mediate the relationship between rehearsal efficiency and recall performance.

Implications for recall accuracy. In recall accuracy, at least three emergent properties of less efficient rehearsal are hypothesized to exert a differential negative impact on items from early serial positions, thus decreasing the primacy effect (see Figure 2i–k, n). First, items from early serial positions will be rehearsed fewer times, and because these items are rehearsed more than other items, the reduction is more severe. Second, not rehearsing these items as far into the study sequence decreases the recency of the last rehearsal, or *functional serial position* (Brodie & Murdock, 1977), thereby reducing contextual overlap with the test context. When more presentations and rehearsals intervene between the final rehearsal of an item and the test, the item is less likely to be recalled (Laming, 2009; Tan & Ward, 2000; Ward et al., 2003). Third, with fewer spaced rehearsals, words from early serial positions would exhibit disproportionate decreases in recall accuracy, because items from later positions are not typically rehearsed far beyond their initial presentation even in more efficient rehearsal sequences (Tan & Ward, 2000).

We therefore predict that low word frequency, lower language proficiency, and bilingualism will be associated with lower recall accuracy for items from early serial positions, or weaker primacy effects. Previous research shows that recall of items from early but not late serial positions is indeed less accurate for low-frequency than for high-frequency word lists, showing a weaker primacy effect for low-frequency words (e.g., Raymond, 1969; Sumby, 1963; Tan & Ward, 2000; Ward et al., 2003). The importance of differential rehearsal to this frequency effect is supported by the finding that the high-frequency advantage in recall accuracy was eliminated when rehearsal was prevented (Gregg, Montgomery, & Castano, 1980). Alternative explanations for the high-frequency advantage in free recall have included stronger preexperimental associations (e.g., Raaijmakers & Shiffrin, 1981); more efficient formation of new associations at study (Deese, 1960), or stronger subjective organization (e.g., Postman, 1970). Any of these mechanisms might compound the impact of individual item accessibility on rehearsal efficiency, which would be a more proximal cause for the recall differences observed.

In bilinguals, a number of studies have compared overall free recall accuracy across L1 and L2 or across monolinguals and bilinguals (e.g., Fernandes et al., 2007; Francis & Baca, 2014; Francis et al., 2018; Haritos, 2002; Harris et al., 1995; and several earlier studies) but only one reported serial position effects, with results suggesting more accurate recall in L1 relative to L2 for early but not late serial positions (Yoo & Kaushanskaya, 2016). Although the weight of the evidence across studies suggests better recall in L1 relative to L2 and better recall in monolinguals relative to bilinguals, the results are inconsistent and do not allow definitive conclusions because of confounds and/or limited power. In most studies, proficiency was confounded with language-specific effects or language groups were not adequately matched. Power was limited because of (a) long lists for which the primacy items make up only a small proportion, (b) only one or two study-test cycles per language or group, or (c) small sample sizes. One explanation offered for less accurate recall in L2 or in bilinguals was less effective rehearsal (Yoo & Kaushanskaya, 2016); others included slower decoding (i.e., comprehension; Nott & Lambert, 1968) and more difficult lexical access for retrieval (Fernandes et al., 2007), either of which could contribute to performance through rehearsal efficiency.

Implications for the first item recalled. Less efficient rehearsal is also hypothesized to impact which item is recalled first. Specifically, when items from early serial positions are not rehearsed as far into the sequence, their temporal contexts will have less overlap with the test context, making them less likely to be recalled first. The PFR function would therefore shift toward later serial positions (see Figure 2k). Also, if the recency of both presentations and rehearsals contributes to selection of the first item recalled, selection of the first item recalled should depend to a greater degree on its *functional* serial position (recency of rehearsal) than on its nominal serial position, and this functional measure should be even more sensitive to inefficient rehearsal. We tested for the first time whether low-frequency words, lower language proficiency, and bilingual status would be associated with later PFR, particularly when defined by functional rather than nominal serial position.

Transitions in recall sequences. We hypothesize that less efficient rehearsal will impact the recall sequence by changing relative frequencies with which other items are recalled next. Because transitions in recall output are hypothesized to arise from the transition patterns in rehearsal (see Figure 2l–m), when less efficient rehearsal leads to fewer rehearsal transitions to items from later or adjacent serial positions, there are fewer opportunities to form or strengthen interitem associations. As a result, recall sequences will also have fewer transitions to items from later and adjacent serial positions. Also, relative to items presented in consecutive serial positions, items *rehearsed consecutively* at study, or *functionally consecutive* items, ought to be even more closely associated with the pattern of transitions in free recall and more sensitive to decrements in rehearsal efficiency.

We therefore predicted that low word frequency, lower language proficiency, and bilingual status would be associated with lower rates of forward and consecutive transitions in free recall output, particularly in the case of functionally consecutive transitions. Indeed, low-frequency word lists elicited lower rates of forward and consecutive transitions in free recall output (Ward et al., 2003), but functionally consecutive transitions have not been reported. Patterns of recall transitions in bilingual participants have never been reported. However, one study of bilingual free recall reported lower relative order scores (which are based on forward transitions) in L2 than in L1 (Francis & Baca, 2014), suggesting weaker formation of directional associations with lower proficiency.

The Present Study

Hypotheses and Predictions

Four experiments were designed to test four sets of hypotheses embodied in the proposed model. The first goal was to test hypotheses about the contribution of rehearsal efficiency to recall dynamics. We tested predictions about the effects of word frequency, language proficiency, and bilingualism on four primary recall measures: recall as a function of serial position; serial PFR; forward transitions in recall output (to items from later serial positions); and consecutive transitions in recall output (to items from adjacent serial positions). As explained in the preceding section, the proposed model predicts that with low-frequency, lower proficiency, and bilingualism, recall would be less accurate for items from early serial positions, begin with items from later serial positions, and exhibit fewer forward and consecutive transitions. We also tested whether differences in PFR across the three critical factors could explain differences in their effects on recall accuracy across serial positions (Figure 2n). The effects of frequency were tested in Experiments 1 and 2, the effects of language proficiency were tested in Experiment 2, and the effects of bilingualism were tested by comparing performance across Experiments 1 and 2. Experiments 3 and 4 replicated these comparisons.

The second goal was to test the assumptions and hypotheses about rehearsal that were used to derive the recall predictions. First, we tested the assumption that the three critical variables would impact rehearsal efficiency (Figure 2a–c). Next, we tested hypotheses about how inefficient rehearsal would impact five emergent rehearsal properties, particularly for items from early serial positions (Figure 2d–h): the number of rehearsals (Figure 2d); how far the rehearsals continued into the study sequence (Figure 2f); spacing between repeated rehearsals (Figure 2e); forward rehearsal transitions to items from later serial positions (Figure 2g); and consecutive rehearsal transitions to items from adjacent serial positions (Figure 2h). As explained in the preceding section, the proposed model predicts that for items from early serial positions, low frequency, lower proficiency, and bilingualism would be associated with fewer rehearsals, rehearsals that did not continue as far into the study sequence, less spacing between repeated rehearsals, and fewer forward and consecutive rehearsal transitions. The effects of frequency were tested in Experiments 3 and 4, the effects of language proficiency were tested in Experiment 4, and the effects of bilingualism were tested by comparing performance across Experiments 3 and 4.

The third goal was to better establish the relationships between the rehearsal-based antecedents and the recall outcomes that would be expected if rehearsals are treated as study events equivalent to additional presentations. We tested the hypothesis that emergent rehearsal properties mediate the effects of rehearsal efficiency on recall performance (Figure 2i–n) by examining whether the effects of word frequency and language proficiency on recall were mediated by the emergent rehearsal properties. First, we examined recall accuracy as a function of item rehearsal properties, including the number of rehearsals (Figure 2i), recency of the last rehearsal (Figure 2k and 2n), and maximal repetition spacing (Figure 2j). Second, we examined PFR using the functional serial position of the first item recalled rather than the nominal serial position (Figure 2k). Third, we examined transitions in recall output to items that were rehearsed consecutively. Finally, we examined forward and consecutive transitions in recall output as a function of the numbers of forward and consecutive transitions (Figure 2l and 2m), respectively, in rehearsal of the corresponding items. To further establish the mediation, in each case, we examined whether including rehearsal properties in the regression models substantially reduced or eliminated the effects of word frequency and language proficiency. Hypotheses about mediation were tested in Experiments 3 and 4.

The fourth goal was to test more general hypotheses about the status of rehearsals as encoding events similar to presentations. As explained in the preceding sections, if contextual features of both presentations and rehearsals fluctuate, then both recall accuracy and the probability of an item being recalled first will be monotonic functions of recency of rehearsal (Figure 2k and 2n). When repeated rehearsals are spaced further apart, they have less overlap in their associated contextual features, making them sensitive to distinct sets of retrieval cues from reinstated contexts of recalled items, leading to increased recall accuracy (Figure 2j). Reinstated presentation or rehearsal contexts of recalled items will cue the retrieval of other items that were presented or rehearsed close together at study, resulting in high rates of transitions to consecutively rehearsed items in recall output. Finally, forward and consecutive transitions in rehearsal will be associated with forward and consecutive transitions (Figure 2l and 2m), respectively, for those same items in recall output. These derivations are tested in Experiments 3 and 4.

Experiment Overview

In Experiments 1 and 2, English-speaking monolinguals and Spanish-English bilinguals (matched on age, education, and socioeconomic status) completed 32 study-test cycles in which they intentionally attempted to commit 10 words to memory and completed an immediate free recall test. In Experiments 3 and 4, English-speaking monolinguals and Spanish-English bilinguals (matched on age, education, and socioeconomic status) completed 16 study-test cycles in which they intentionally attempted to commit 20 words to memory while using a rehearse-aloud protocol and then completed an immediate free recall test. In all experiments, half of the lists consisted of high-frequency words and half consisted of low-frequency words. For bilingual participants (Experiments 2 and 4), half of the lists were presented and recalled in English, and half were presented and recalled in Spanish.

Power and Sample Size

Sample size was determined for all experiments at the proposal stage based on a power analysis and counterbalancing requirements. With a completely within-subjects design, a power analysis showed that at least 34 participants would be needed to have 80% power to detect a medium-sized effect in a repeated measures ANOVA. (In the end we conducted logistic mixed-effects regression analyses, in which methods of estimating necessary sample size are not yet well developed, but given that the stimulus pool contained 320 different items, each of which appeared in English and in Spanish across participants, the estimates based on ANOVA should be sufficient.) Counterbalancing put constraints on the sample size such that Experiment 1 required a multiple of 20, Experiment 2 required a multiple of 80, Experiment 3 required a multiple of 40, and Experiment 4 required a multiple of 80. Therefore, we decided to collect data from 80 participants in all experiments to have similar levels of power to detect effects of interest in all experiments. IRB approval was obtained before data collection began.

Experiment 1

Experiment 1 examined the effects of word frequency on the temporal dynamics of free recall. Participants attempted to memorize 16 sets of 10 high-frequency words and 16 sets of 10 low-frequency words and free recall each list aloud. Recall measures included accuracy, the position of the first item recalled, and transitions rates meant to capture temporal associations.

Method

Participants. Participants were 80 monolingual English speakers. The participants were students at the University of Texas at El Paso who were compensated with credit toward a course research requirement or a payment of $10. Because even monolingual English speakers in El Paso have regular exposure to ambient Spanish, both English and Spanish proficiency were assessed using the picture vocabulary subtest of the Woodcock-Muñoz Language Survey-Revised (Woodcock, Muñoz-Sandoval, Ruef, & Alvarado, 2005), a standardized objective language assessment that was developed, normed, and calibrated with English speakers and Spanish speakers in the Americas. To qualify, participants had to indicate that English was their only proficient language and converse with native-like fluency, be unable to converse in Spanish, and have Spanish picture vocabulary scores below an 8-year-old level. As shown in Table 2, on average, monolingual participants scored slightly below age level on the English picture vocabulary test and at a 3-year-old level on the Spanish picture vocabulary test, indicating negligible productive vocabulary. No participant reported proficiency in another language.

Design. The experimental conditions formed a 2 (word frequency) × 10 (serial position) repeated-measures design. All words were presented in English. Each list had 10 items/serial positions, with half of the lists consisting of high-frequency and half consisting of low-frequency words. The dependent variables were recall rates, PFR, and forward and consecutive transition rates.

Apparatus. Experimental stimuli were presented on the monitor of an iMac computer. PsyScope X software was used to program the sequences of stimuli and their timing (Cohen, MacWhinney, Flatt, & Provost, 1993). A Sony digital recorder was used to record spoken recall for later transcription.

Materials. The experimental stimuli consisted of 160 high-frequency words and 160 low-frequency English words (see Appendix B in the online supplemental materials). High-frequency words had frequencies of at least 50 per million (Mdn = 160); low-frequency words had frequencies of no more than 15 per million (Mdn = 3.6; CELEX, Baayen, Piepenbrock, & Gulikers, 1995). The mean normative rated age of acquisition was 6.17 (SD = 2.02) for high-frequency and 6.99 (SD = 2.15) for low-frequency words (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). The median word length was six letters for high-frequency and seven letters for low-frequency words. Words at each frequency level were randomly assigned to 16 sets of 10 words each, to make 32 lists in all. The order of the words in each list was counterbalanced using a balanced Latin square, such that every word appeared in every serial position and every word followed every other word equally often across participants.

Table 2
Characteristics of Participants

Participant characteristic	Experiment 1	Experiment 2		Experiment 3	Experiment 4	
	English monolingual (N = 80)	English dominant (N = 40)	Spanish dominant (N = 40)	English monolingual (N = 80)	English dominant (N = 40)	Spanish dominant (N = 40)
Median age	21.0	22.0	22.0	19.0	20.0	20.5
Median age of acquisition—English	2.0	5.0	7.0	1.0	5.0	7.0
Median age of acquisition—Spanish	—	1.0	1.0	—	1.0	1.0
Age equivalency—English picture vocabulary[a]	18.8	16.5	11.0	15.6	15.1	10.2
Age equivalency—Spanish picture vocabulary[a]	3.3	11.1	14.5	3.4	10.0	13.6
Usage of English	89%	55%	38%	89%	62%	41%
Usage of Spanish	7%	29%	50%	7%	24%	42%
Usage of mixture	3%	16%	12%	4%	13%	17%
Percent reporting Latino ethnicity	71%	100%	100%	76%	100%	100%
Median participant education level	some college	some college	some college	some college	some college	some college
Median highest parent education level	graduated college	some college	graduated college	graduated college	graduated college	graduated college

[a] Based on picture vocabulary subtest of the WMLS-R (Woodcock et al., 2005).

Procedure. Participants were tested individually in 50-min sessions. After informed consent, the picture vocabulary subtest of the Woodcock-Muñoz Language Survey-Revised was administered in English and Spanish (Woodcock et al., 2005). While the experimenter scored this test, participants completed language background and demographics questionnaires.

In the computerized experiment, participants were asked to memorize and then free recall 32 lists of 10 words each. Words appeared on the computer monitor one at a time for 2 s each, with a blank screen for 1 s between words. At the end of each 10-word sequence, the word *Recall* appeared on the screen and participants recalled aloud as many words as they could remember. There were 16 study-recall cycles each for high- and low-frequency words. Cycles were blocked by frequency, changing after every eight cycles, and the order of frequency conditions was counterbalanced across participants. Participants were allowed to rest between lists whenever they felt the need to do so. All sessions were recorded and later transcribed.

Coding and scoring. In all experiments, recall sequences (and rehearsal sequences in Experiments 3 and 4) were transcribed into Excel spreadsheets. To reduce the likelihood of human errors or bias, coding and scoring of individual participant data were accomplished using customized templates with appropriate lookup and matching functions.

Results and Discussion

General approach to analysis. Binary dependent variables were analyzed using logistic mixed-effects regression, and quantitative dependent variables were analyzed using linear mixed effects regression to accommodate the random effects of participants and items simultaneously (Baayen, Davidson, & Bates, 2008). These analyses were implemented in R using the lme4 package. Binary predictors used effects coding, and continuous predictors were centered and standardized. In the logistic regression analyses, untransformed coefficients are reported, which refer to linear differences in log-odds across conditions.

All initial models included random intercepts for participants and items, random slopes across participants for all categorical within-subjects variables and their interactions, and random slopes across items for all categorical within-item variables and their interactions. Thus, maximal random effects structures were incorporated when possible (as recommended by Barr, Levy, Scheepers, & Tily, 2013), but they often had to be simplified to allow models to converge and avoid overparameterization. When models did not converge, random-effects structures were simplified by first removing random slopes for the highest-order interactions, then the next highest order, and so forth. When necessary for model convergence, removing parameters for correlations between random slopes and random intercepts does not increase Type I error rates (Barr et al., 2013). If a model had random slope terms with correlations $\geq .95$, the higher order term was removed from the model to avoid redundancy. (Fixed effects were never removed).

Recall rates and serial position curves. Monolingual serial position curves are illustrated in Figure 3. For the analysis of recall accuracy, each item was coded as correctly recalled (1) or not recalled (0). To examine the effects of serial position, we tested for linear and quadratic trends across positions rather than an unspecified main effect. The linear trend is sensitive to an overall increase or decrease in performance across positions, and the quadratic trend is sensitive to a U-shaped function typical of the serial position curve. While this approach to analysis is unconventional in the memory literature, it better captures the shape of the serial position curve than approaches that involve pooling sets of items from early, middle, and late positions. We therefore simplified the serial position variable by coding only two essential orthogonal components. One contained the linear trend coefficients (which served to center the 10 serial positions) and the other contained the quadratic trend coefficients. These components were not crossed in

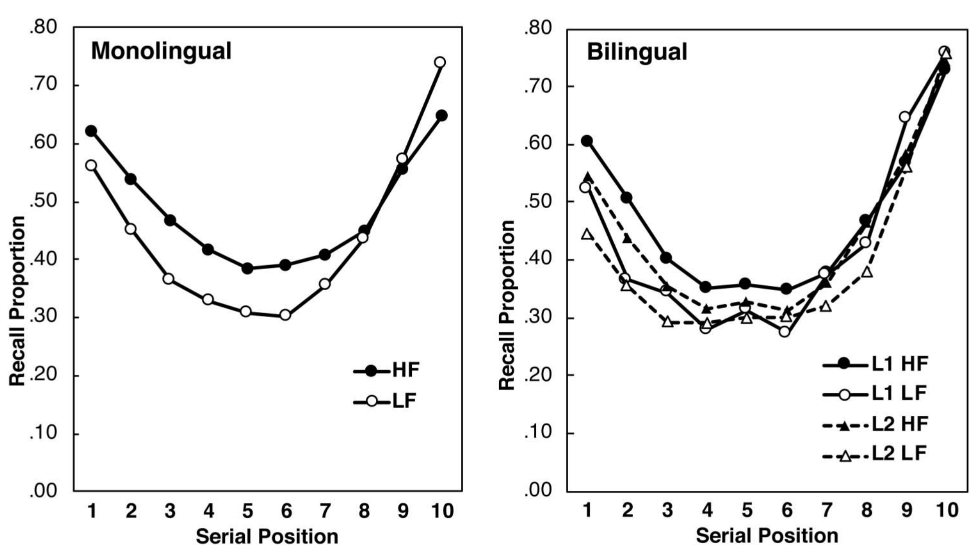

Figure 3. Serial position curves for recall accuracy in Experiments 1 and 2. Mean recall rates at each serial position in monolinguals and bilinguals. HF = high frequency; LF = low frequency; L1 = more proficient language; L2 = less proficient language.

the model, because they are orthogonal components of the same variable.

A logistic mixed-effects regression model was fit to the monolingual recall data, with word frequency and the linear and quadratic trends on serial position as within-subjects fixed effects and whether items were recalled or not as the dependent variable. The model included random intercepts for participants and items and random slopes for word frequency across participants. Monolingual serial position curves showed the typical U-shaped function, as indicated by a significant quadratic trend across serial positions ($b = .484$, $SE = .014$, $z = 35.486$, $p < .001$). There was also a linear trend across serial positions ($b = .124$, $SE = .013$, $z = 9.250$, $p < .001$), indicating that the recency effect was stronger than the primacy effect. Low-frequency words were less accurately recalled than high-frequency words ($b = -.214$, $SE = .061$, $z = -3.494$, $p < .001$). However, an interaction of word frequency and the linear trend on serial position ($b = .212$, $SE = .027$, $z = 7.930$, $p < .001$) indicated that the high-frequency advantage was observed for items from earlier but not later serial positions, consistent with previous research (e.g., Raymond, 1969; Sumby, 1963; Tan & Ward, 2000; Ward et al., 2003). Word frequency also interacted with the quadratic trend ($b = .164$, $SE = .027$, $z = 6.018$, $p < .001$), indicating that the curvature of the serial position function was greater for low-frequency word lists.

Position of first item recalled. The relative frequency distribution of the first item recalled across serial positions is shown in Figure 4. The dependent variable of PFR was analyzed as a quantitative variable in a linear mixed-effects model that included

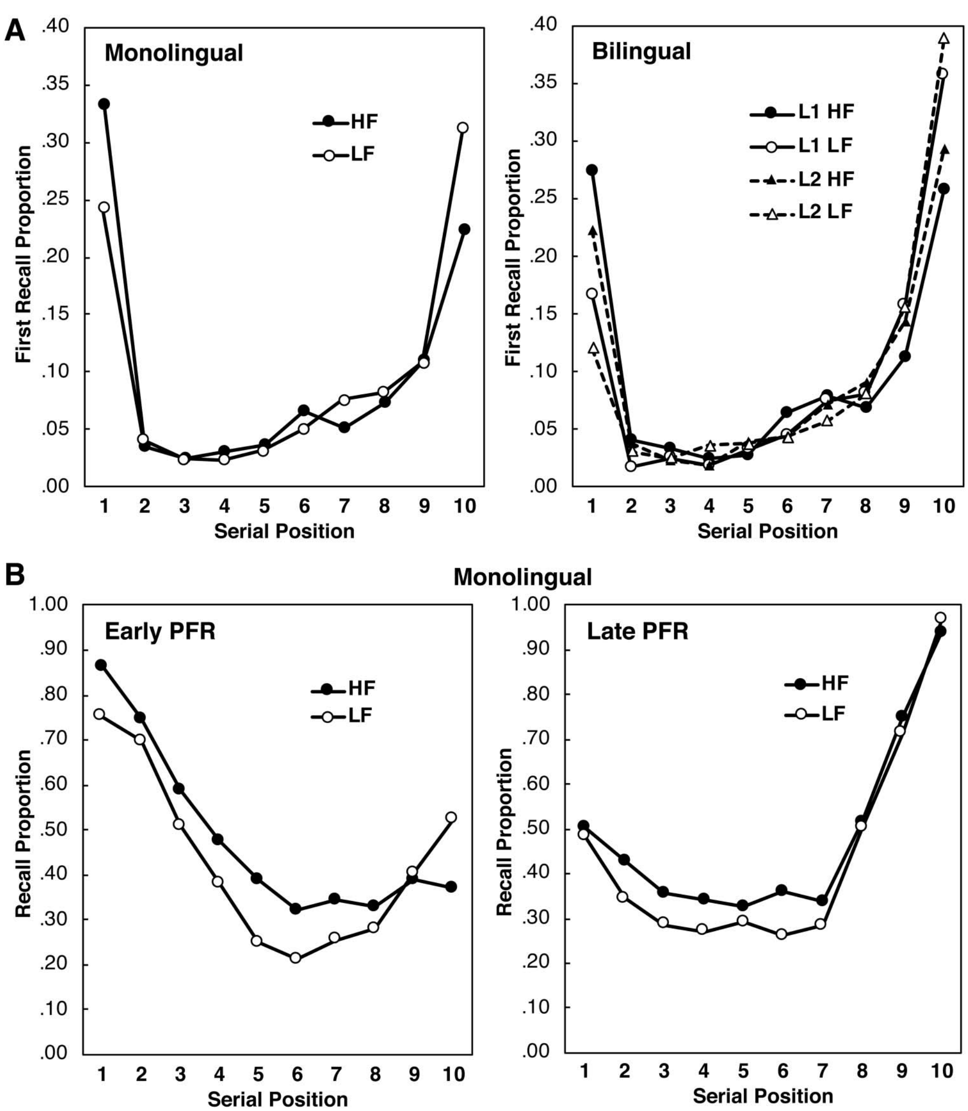

Figure 4. Serial position of first item recalled (PFR) and accuracy as a function of PFR in Experiments 1 and 2. (A) Relative frequency of first items recalled at each serial position as a function of word frequency and language proficiency. (B) Mean monolingual recall rates at each serial position as a function of word frequency for trials initated with items from early serial positions (1–3) or late serial positions (8–10). HF = high frequency; LF = low frequency; L1 = more proficient language; L2 = less proficient language.

word frequency as a fixed effect, random intercepts for participants and items, and random slopes for word frequency across participants. Items from the first and last serial positions were most frequently recalled first, consistent with previous research (e.g., Howard & Kahana, 1999; Kahana et al., 2002; Polyn et al., 2009). Recall of low-frequency lists began with items from later positions than for high-frequency lists ($b = .848$, $SE = .154$, $t = 5.488$, $p < .001$), as expected if items from early positions were not rehearsed as far into the study sequence.

Serial position effects as a function of PFR. The form of the serial position curve depends on whether the first item recalled comes from an early or late serial position (Ward, Tan, & Grenfell-Essam, 2010). Specifically, recall sequences that begin with items from early serial positions show stronger primacy and weaker recency effects, whereas those that begin with items from late serial positions show weaker primacy and stronger recency effects. Because a higher proportion of high-frequency than low-frequency recall sequences began with early list items, we considered the possibility that the greater primacy effect for high-frequency word lists could be an artifact of earlier PFR. To examine this possibility, PFR scores were categorized as early for the first three serial positions and late for the last three serial positions (other positions were not classified). The original serial position analysis was repeated with early versus late PFR included as a fixed categorical factor, and random slopes added for PFR category across participants and items. If PFR differences across frequency levels were the proximal cause of differences in serial position functions, conditionalizing recall accuracy on PFR category would eliminate the frequency effect. Serial position curves conditionalized on PFR are illustrated in Figure 4.

Overall accuracy did not differ for recall sequences that began with items from earlier and later serial positions ($z < 1$). However, serial position effects differed markedly for lists with early and late PFR, as indicated by significant interactions with linear ($b = 1.496$, $SE = .034$, $z = 43.377$, $p < .001$) and quadratic trends ($b = .257$, $SE = .035$, $z = 7.450$, $p < .001$). Separate analyses of sequences with early and late PFR showed that when recall began with items from early serial positions, the primacy effect was stronger than the recency effect ($b = -.774$, $SE = .025$, $z = 30.805$, $p < .001$), but when recall began with items from late serial positions, the recency effect was stronger than the primacy effect ($b = .708$, $SE = .023$, $z = 30.353$, $p < .001$). These effects are consistent with Figure 2n, Ward, Tan, and Grenfell-Essam (2010), and previous research showing that the tendency to recall the last item first is highly correlated with the recency effect (e.g., Deese & Kaufman, 1957; Kahana, 1996). The frequency effect persisted ($b = -.231$, $SE = .067$, $z = 3.437$, $p < .001$), and its size and interactions with serial position trends did not depend on PFR ($ps > .1$). Thus, differences in serial position effects based on PFR did not explain the differential effects of word frequency across serial positions.

Temporal association encoding. Analyses of the temporal relationships among items in recall output included data from the first four recall output positions only (as in Kahana et al., 2002). The reason not to include data from later output positions is that with recall accuracy averages from 40% to 50%, data beyond the fourth position were sparse. Lag-recency functions are shown in Figure 5. We analyzed transition direction and temporal proximity separately in logistic mixed-effects model with word frequency and linear and quadratic trends on serial position as within-subjects fixed effects, random intercepts for participants and items, and random slopes for word frequency across participants.

Overall, 63% of recall transitions were in the forward direction, meaning that the next word recalled was more likely to be from a later rather than earlier serial position (see Table 3), consistent with previous research (e.g., Howard & Kahana, 2002; Laming,

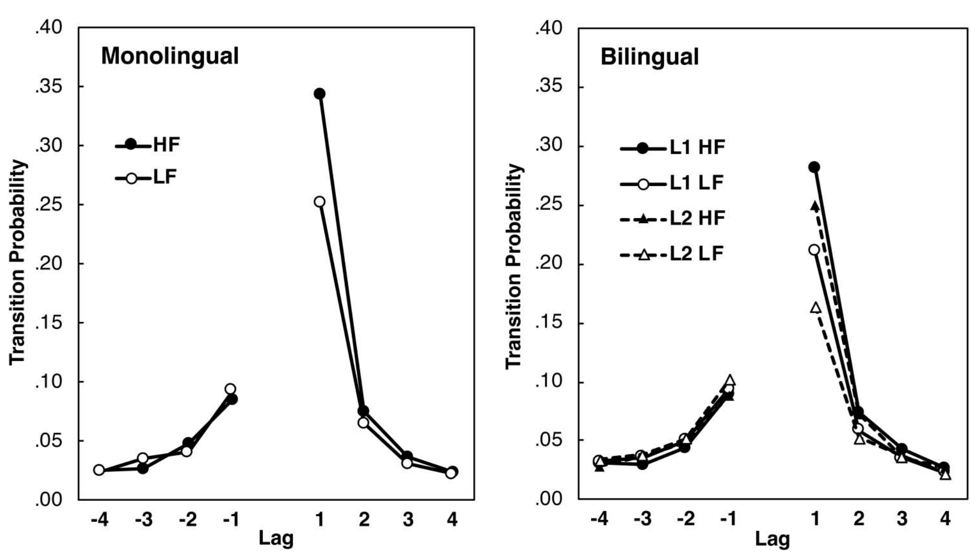

Figure 5. Lag recency functions (conditional response probability curves) in Experiments 1 and 2. Transition probabilities as a function of word frequency and language proficiency in monolinguals and bilinguals. Lag indicates the relative serial positions at study of items recalled consecutively at test. Probabilities are averaged across the first four output positions. HF = high frequency; LF = low frequency; L1 = more proficient language; L2 = less proficient language.

Table 3
Overall Recall Accuracy and Proportions of Forward and Consecutive Recall Transitions in Experiments 1 and 2

Group / Task language / Word frequency	Accuracy *M*	Accuracy *SE*	Forward[a] *M*	Forward[a] *SE*	Consecutive *M*	Consecutive *SE*
Monolingual English						
English						
HF	.489	.011	.724	.017	.429	.013
LF	.442	.012	.666	.019	.345	.013
English dominant						
English						
HF	.491	.017	.682	.031	.413	.017
LF	.446	.017	.641	.034	.338	.018
Spanish						
HF	.436	.016	.662	.032	.332	.017
LF	.398	.017	.562	.032	.258	.019
Spanish dominant						
English						
HF	.453	.013	.667	.032	.332	.015
LF	.404	.012	.597	.032	.272	.014
Spanish						
HF	.451	.013	.717	.027	.345	.015
LF	.415	.012	.579	.029	.274	.016

Note. HF = high frequency; LF = low frequency.
[a] Excluding transitions from Serial Positions 1 and 10.

2009; Ward et al., 2003). In fact, forward transitions were more frequent than backward transitions for every serial position except the final position, where a forward transition was not possible. The effect of word frequency on this pattern was examined in an analysis with direction (forward or backward) as a binary dependent variable. Items from the first and last serial positions were excluded, because they allow transitions in only one direction. The proportion of forward transitions was higher for items from earlier serial positions ($b = -.609$, $SE = .038$, $z = 16.123$, $p < .001$), and the curvature of this function was marginal ($b = .062$, $SE = .034$, $z = 1.846$, $p = .065$). The proportion of forward transitions was lower for low-frequency than for high-frequency word lists ($b = -.219$, $SE = .088$, $z = -2.480$, $p = .013$), replicating previous research (Ward et al., 2003). No other effects were significant ($ps > .20$).

When two items were recalled consecutively at test, their temporal proximity or lag at encoding could range from one position apart to nine positions apart. The full set of transition rates for all lags from one to nine are given in Table A.3 (in the online supplemental materials). However, we treated proximity in a binary manner, consecutive or nonconsecutive, for three reasons. First, about half of the transitions were to consecutive items (lags of one in either direction) and this was by far the most frequent lag. Second, lags greater than five were not possible for items with serial positions near the middle of the study sequence. Third, some explanations of the phenomenon claim that transitions to items two or more positions away are not based on a graded degree of temporal association as the distance increases (Laming, 2008, 2009; we return to this issue in the General Discussion). The proportion of consecutive transitions in recall output was higher for words from earlier relative to later serial positions ($b = .178$, $SE = .025$, $z = 7.079$, $p < .001$) and the function was curved ($b = .090$, $SE = .024$, $z = 3.803$, $p < .001$). The proportion of consecutive transitions was lower for low-frequency than for high-frequency lists ($b = .240$, $SE = .064$, $z = 3.772$, $p < .001$) as in Ward et al. (2003), with a stronger effects at earlier serial positions ($b = -.170$, $SE = .048$, $z = 3.532$, $p < .001$). No other effects were significant ($z < 1$). These patterns indicate that temporal associations formed at study are fewer or weaker for low-frequency relative to high-frequency word lists.

Experiment 2

The first goal of Experiment 2 was to replicate the effects of word frequency observed in Experiment 1 with a bilingual population to increase generalizability. The second goal was to examine the effects of language proficiency on the same four free recall measures and to compare and contrast proficiency and frequency effects. Participants attempted to memorize 16 sets of 10 English words and 16 sets of 10 Spanish words and free recall each list aloud. Half of the word sets in each language had only high-frequency and half had only low-frequency words.

Method

Participants. Participants were 80 Spanish-English bilingual students at the University of Texas at El Paso who were compensated with credit toward a course research requirement or a payment of $10. Half of the participants were English dominant and half were Spanish dominant, with dominance determined by their scores on the picture vocabulary subtest of the Woodcock-Muñoz Language Survey-Revised (Woodcock et al., 2005), a standardized objective language assessment that was developed, normed, and calibrated with English speakers and Spanish speakers in the Americas. The language with the higher score was considered the dominant language, and age-equivalency scores for the nondominant language had to be at least 8 years (or over 7 years with clear conversational proficiency). Characteristics of the research participants are summarized in Table 2.

Design. The experimental conditions formed a 2 (language) × (word frequency) × 10 (serial position) repeated-measures design. Half of the word lists were in English and half in Spanish. Within each language, half of the lists had high- and half had low-frequency words. Each list had 10 items/serial positions. Dependent variables were the same as in Experiment 1.

Materials. The experimental stimuli were 160 high-frequency words and 160 low-frequency words in English and their translation-equivalents in Spanish (listed in Appendix B, supplemental materials). The English words and lists were the same as in Experiment 1. Spanish high-frequency words had frequencies of at least 50 per million ($Mdn = 155$) and low-frequency words had frequencies of no more than 15 per million ($Mdn = 4.0$; Alameda & Cuetos, 1995). Mean normative rated age of acquisition in Spanish was 5.52 ($SD = 2.11$) for high-frequency and 5.95 ($SD = 1.87$) for low-frequency words (Alonso, Fernandez, & Díez, 2015). The median word length was six for high-frequency and seven for low-frequency words. The high- and low-frequency word sets were randomly assigned to appear in English or Spanish, with the assignment of sets to languages counterbalanced across participants. Thus, for each Language × Frequency combination, there were eight sets of 10 words each. No participant saw translation equivalents across lists.

Procedure. Participants were tested individually in 50-min sessions using the same assessments, questionnaires, and apparatus as in Experiment 1. The computerized experiment required participants to attempt to memorize and then free recall 32 lists of 10 words each, with the same instructions, trial structure, and timing as in Experiment 1. There were eight study-recall cycles for each Language × Frequency combination, blocked by language and frequency, with the order of conditions counterbalanced across participants. Participants were allowed to rest between lists whenever they felt the need to do so. All sessions were recorded and later transcribed.

Results and Discussion

Approach to analysis. The approach to analysis was the same as for Experiment 1 with the addition of language proficiency as a fixed factor. The inferential analyses reported in the following paragraphs treat language proficiency as a continuous variable, but the same patterns are found when proficiency is treated in a binary manner (L1 vs. L2). L1 and L2 means are used in illustrative graphs to better visualize the effects. Language proficiency scores were based on the objective WMLS-R assessment, using the W measure, which has better psychometric properties than age-equivalency scores. These scores were standardized, and z scores for each participant were entered according to the task language. Specifically, English WMLS-R scores were used for English recall and Spanish scores were used for Spanish recall trials.

Recall rates and serial position curves. Bilingual serial position curves are illustrated in Figure 3. Bilingual serial position curves for all list types showed the typical U-shaped function, as indicated by a significant quadratic trend across serial positions (b = .544, SE = .014, z = 39.170, p < .001). The linear trend across serial positions was also significant (b = .296, SE = .014, z = 21.844, p < .001), indicating that the recency effect was stronger than the primacy effect. As in Experiment 1 and previous monolingual research, high-frequency words were recalled more accurately than low-frequency words (b = −.193, SE = .053, z = 3.668, p < .001), and this effect was observed for earlier but not later list items (b = .160, SE = .027, z = 5.908, p < .001).

Higher proficiency in the task language was associated with more accurate recall (b = .082, SE = .017, z = 4.742, p < .001), and a marginal interaction with the linear trend (b = −.025, SE = .014, z = 1.856, p = .063) suggests that higher proficiency increased recall of items from earlier but not later serial positions (as in Yoo & Kaushanskaya, 2016). A three-way interaction of word frequency, language proficiency, and the linear trend on serial position (b = −.126, SE = .054, z = −2.318, p = .020), indicated that the greater primacy effect with higher proficiency was stronger for high-frequency than for low-frequency words and significant only for high-frequency words. No other effects were significant (ps > .2).

Position of first item recalled. The relative frequency distribution of PFR is shown in Figure 4. As in Experiment 1, low-frequency list recall had later PFR than high-frequency recall (b = 1.078, SE = .155, t = 6.950, p < .001). Similarly, lower proficiency was associated with later PFR (b = −.327, SE = .072, t = 4.560, p < .001). The effects of word frequency were stronger when proficiency was higher, and the effects of proficiency were stronger for high-frequency words (b = .299, SE = .123, t = 2.435, p = .015). Thus, items from early serial positions were less often recalled first when lists consisted of low-frequency words or when proficiency in the task language was lower.

Serial position effects as a function of PFR. We tested whether the patterns observed in Experiment 1 for recall sequences initiated with items from early and late serial positions would replicate in bilinguals. In contrast to Experiment 1, overall accuracy was higher for recall sequences that began with items from later serial positions (b = .120, SE = .046, z = 2.602, p = .009). However, the markedly different serial position effects for lists with early and late PFR replicated, as indicated by significant interactions with linear (b = 1.274, SE = .035, z = 35.979, p < .001) and quadratic trends (b = .257, SE = .036, z = 7.162, p < .001). Separate analyses showed that with early PFR, the primacy effect was stronger than the recency effect (b = −.566, SE = .028, z = 20.142, p < .001), but with late PFR, the recency effect was stronger than the primacy effect (b = .696, SE = .021, z = 33.267, p < .001). Consistent with Experiment 1, the sizes of the frequency effect, the language effect, their interaction, and their interactions with serial position contrasts did not depend on PFR (ps > .1). Thus, the primacy advantages for high-frequency words and higher language proficiency cannot be explained by differences in PFR.

Temporal association encoding. Proportions of forward transitions are given in Table 3. As in Experiment 1, the majority of transitions were in the forward direction, and averaging across conditions, the forward direction was more frequent at every serial position except the final position. Again, the proportion of forward transitions was stronger for items from earlier serial positions (b = −.634, SE = .039, z = 16.330, p < .001), and the serial position function was curved (b = .140, SE = .035, z = 4.059, p < .001). Again, low-frequency word lists elicited a lower proportion of forward transitions than high-frequency word lists (b = −.333, SE = .072, z = 4.614, p < .001). Similarly, lower language proficiency was associated with a lower proportion of forward transitions (b = .107, SE = .041, z = 2.634, p = .008). The effect of language proficiency changed across serial positions, as indicated by an interaction with the linear trend (b = .081, SE = .037, z = 2.158, p = .031). No other effects were significant (ps > .2). Thus, the rate of forward transitions was lower with low-frequency words and lower proficiency, indicating that weaker or less numerous directional associations are formed under these conditions.

Binary consecutive transition scores (see Table 3) were submitted to the same analysis as the forward transition scores. (Lag-recency functions for bilingual participants are shown in Figure 5, and the full set of transition probabilities in Table A.3 in the online supplemental materials.) Consistent with Experiment 1, a marginal linear trend suggested that the proportion of consecutive transitions in recall output was higher for words from earlier rather than later serial positions (b = .047, SE = .025, z = 1.888, p = .059), and the serial position function was curved (b = .066, SE = .024, z = 2.728, p = .006). As in Experiment 1, the rate of consecutive transitions was lower for low-frequency than for high-frequency lists (b = .209, SE = .055, z = 3.773, p < .001), and this effect was stronger at earlier serial positions (b = −.163, SE = .049, z = 3.340, p < .001). Language proficiency showed a similar pattern in which lower language proficiency was associated with a lower rate of consecutive transitions (b = −.059, SE = .028, z = 2.112,

p = .035), and a marginal interaction with the linear trend suggested that this effect was stronger at earlier serial positions (b = .045, SE = .024, z = 1.856, p = .064). No other effects were significant (ps > .20). Thus, the proportion of forward transitions was lower with low-frequency words lists and lower proficiency in the task language, indicating that weaker or less numerous interitem temporal associations were formed.

Analysis of Combined Data From Experiments 1 and 2

The combined data of Experiments 1 and 2 were analyzed to test hypotheses about differences between monolingual and bilingual recall dynamics. When comparing bilingual and monolingual samples, it is important to take steps to have some confidence that any effects of language group are due to differential language experience rather than other factors known to affect cognition, such as age, education, or socioeconomic status. As shown in Table 2, the monolingual and bilingual samples were equivalent on age, education, and highest parent education level (the median for the two bilingual groups combined was "graduated college"), which was used as a proxy for socioeconomic status (Galobardes, Shaw, Lawlor, Lynch, & Davey Smith, 2006). All bilingual participants and 71% of monolingual participants reported Latino/Hispanic ethnicity. To control for differences in any idiosyncratic properties of English and Spanish words, only English trials were included in the analyses. Thus, the participants were monolingual speakers of English, L1 speakers of English (English-dominant bilinguals), or L2 speakers of English (Spanish-dominant bilinguals). Both monolingual and bilingual protocols alternated word frequency levels after every eight study-test cycles to make them identical except for the language change after the sixteenth cycle in the bilingual protocol.

Results and Discussion

The same general approach was used as for the separate analyses of Experiment 1 and 2 data. For each analysis, there were two mixed-effects regression models, one to compare bilingual L1 and monolingual and one to compare bilingual L2 and monolingual performance in English. To avoid redundancy, we report only effects involving language group.

Recall rates and serial position curves. Monolingual and bilingual English recall accuracy (Table A.1 in the online supplemental materials) were compared in logistic mixed-effects regression models with group (monolingual or bilingual), word frequency, and linear and quadratic trends on serial position as fixed effects. The models included random intercepts for participants and items, and random slopes for word frequency across participants and for group across items.

Accuracy was higher for monolingual than for bilingual L2 (b = .170, SE = .084, z = 2.018, p = .044) but not L1 recall (z < 1). However, serial position slopes were more positive in bilingual L1 and L2 than in monolingual recall (L1: $b = -.214$, SE = .031, z = 6.945, p < .001; L2: $b = -.089$, SE = .030, z = 2.957, p = .003), indicating a monolingual advantage for earlier (primacy) list items but not later (recency) serial positions relative to both L1 and L2 bilingual recall. The serial position function showed less curvature for monolingual relative to bilingual L2 ($b = -.085$, SE = .031, z = 2.757, p = .006) but not L1 recall (p > .20).

The three-way interactions of language group, word frequency, and the quadratic trend on position were reliable (L1: b = .195, SE = .062, z = 3.131, p = .002; L2: b = .135, SE = .061, z = 2.197, p < .028), indicating that the effect of greater curvature for low-frequency words relative to high-frequency words was stronger in monolingual than in bilingual recall. No other effects involving group were reliable (p > .2).

Position of first item recalled. PFR patterns for monolingual and bilingual recall (see Table A.2 in the online supplemental materials) were compared. The numerical tendency for bilinguals to have later PFR than monolinguals was not reliable for L1 or L2 speakers (L1: $b = -.770$, SE = .446, t = 1.727, p = .087; L2: p > .20). These effects were not moderated by word frequency (ts < 1).

Temporal association encoding. Forward transition rates were compared across language groups (see Table 3). Marginal effects suggested that bilingual L2 speakers had fewer forward transitions than monolinguals (b = .269, SE = .141, z = 1.911, p = .056), particularly for items from early positions (b = .157, SE = .085, z = 1.844, p = .065), but L1 speakers did not (z < 1). No other effects involving language group were significant (ps > .1).

Bilingual L2 recall had fewer consecutive transitions than monolingual recall ($b = -.310$, SE = .088, z = 3.511, p < .001). Consecutive transitions for items from early but not late serial positions occurred less often in bilingual L1 than in monolingual recall (b = .114, SE = .057, z = 2.001, p = .045). These results indicate that temporal associations formed among items from early serial positions at study were weaker or less numerous for bilingual than for monolingual speakers. No other interactions involving group were significant (zs < 1).

Interim Discussion of Experiments 1 and 2

Experiments 1 and 2 supported the hypothesis that the primacy effect would be weaker with less efficient rehearsal; primacy effects were weaker under conditions expected to elicit less efficient rehearsal, specifically with low-frequency words, lower language proficiency, and bilingual status. The hypothesis that less efficient rehearsal would elicit recall sequences that began with items from later serial positions was supported by the novel finding that the PFR is later with lower word frequency and lower language proficiency. The hypothesis that less efficient rehearsal would reduce the strength or number of temporal associations formed at study was supported by the finding that forward and consecutive transition rates were lower for low-frequency words, lower language proficiency, and bilingual L2 status.

Although previous research shows that monolingual speakers indeed exhibit less efficient rehearsal of low-frequency words than high-frequency words, bilingual rehearsal patterns have not been reported in previous research. Explanations of the effects of language proficiency and bilingualism on various aspects of free recall performance are difficult to verify in the absence of rehearsal data, which is the impetus for Experiments 3 and 4.

Introduction to Experiments 3 and 4

Rehearsal dynamics at study appear to impact several aspects of free recall performance. However, to support the hypothesis that rehearsal efficiency mediates the effects of word frequency, language proficiency, and bilingualism on recall performance, it is important to more directly establish both the hypothesized rehearsal patterns and the hypothesized relationships between rehearsal and recall patterns.

The traditional nominal serial position does not predict recall accuracy as well as the *functional* serial position, which is a measure of recency of rehearsal (Tan & Ward, 2000; Ward et al., 2003). We hypothesized that the functional serial position would be a better predictor than nominal serial position for both recall accuracy (Tan & Ward, 2000; Ward et al., 2003) and the first item recalled. Thus, if items from early serial positions are rehearsed to the last rehearsal cycle, they would often be recalled first, allowing a single mechanism to explain primacy and recency effects in the PFR function. Because rehearsal and recall sequences exhibit similar patterns of transitions to items from later and adjacent serial positions, a causal link has been proposed (Laming, 2008, 2009). However, we found no investigations that directly mapped the transition patterns in rehearsal and recall sequences generated by the same participants for items in the same lists.

We hypothesized that less efficient rehearsal would change rehearsal patterns, which would in turn reduce the accuracy of later free recall and change the order of recall output. Relative to high-frequency word lists, low-frequency word lists do exhibit less efficient rehearsal with different output patterns (Tan & Ward, 2000; Ward et al., 2003) and less accurate recall with different output patterns (Experiments 1 and 2; Delosh & McDaniel, 1996; Raymond, 1969; Sumby, 1963; Tan & Ward, 2000; Ward et al., 2003). However, exploration of word-frequency effects on rehearsal patterns and the association between rehearsal and retrieval patterns has been limited, but recency of rehearsal was shown to contribute to the word-frequency effect on recall accuracy (Tan & Ward, 2000; Ward et al., 2003). As explained in the introduction, only two published studies investigated the effects of language proficiency or bilingualism on temporal patterns in free recall (Francis & Baca, 2014; Yoo & Kaushanskaya, 2016), and no previous study has reported bilingual rehearsal patterns or their contributions to recall performance.

To support the claim that rehearsal efficiency is the primary mechanism by which the effects of word frequency, language proficiency, and bilingualism impact recall performance, it is important to verify that (a) these factors impact rehearsal efficiency (Figure 2a–c) and rehearsal patterns; (b) rehearsal efficiency predicts rehearsal patterns (Figure 2d–h); and (c) these rehearsal patterns predict the accuracy and sequence of free recall (Figure 2i–n). Rehearsal at study is typically silent, and its impact is inferred based on recall patterns, such as the primacy effect. However, the link between rehearsal dynamics and recall output dynamics can be more directly examined using a rehearse-aloud protocol (e.g., Rundus & Atkinson, 1970; Tan & Ward, 2000). The number and sequence of rehearsals in a rehearse-aloud protocol may differ somewhat from those of spontaneous covert rehearsal, participants may not produce all rehearsals aloud, and participants might rehearse more under these instructions than without such instructions (Ward et al., 2003). However, presenting items to participants according to the rehearsal sequences generated by other participants yielded similar serial position functions to those of the participants who originally generated the sequences (Tan & Ward, 2000). The rehearse-aloud protocol offers the advantage of a direct link between rehearsal and retrieval dynamics. Experiments 3 and 4 use a rehearse-aloud protocol to investigate how word frequency, bilingual proficiency, and bilingual experience impact rehearsal patterns, the temporal dynamics of free recall, and rehearsal-contingent recall dynamics.

Experiment 3

Experiment 3 examined the effects of word frequency on the temporal dynamics of rehearsal and free recall. Rehearse-aloud experiments typically use longer word lists. Participants attempted to memorize eight sets of 20 high-frequency words and eight sets of 20 low-frequency words and free recalled each list aloud. The effects of word frequency were measured for several rehearsal and recall measures to test the hypothesized impacts of less efficient rehearsal. Associations between the rehearsal and retrieval patterns were also examined to investigate whether rehearsal patterns mediate the effects of word frequency on free recall performance.

Method

Participants. The participants were students at the University of Texas at El Paso who were compensated with credit toward a course research requirement or a payment of $20. According to the language proficiency assessment used in Experiment 1, participants had negligible productive vocabulary in Spanish, and none reported proficiency in another language. Other participant characteristics are given in Table 2.

Design. The experimental conditions formed a 2 (word frequency) $\times$ 20 (serial position) repeated-measures design. All words were presented in English in lists of either high- or low-frequency words. Each list had 20 items/serial positions. For rehearsal, the dependent variables measured for each word for each participant included the number of rehearsals, the last rehearsal set, maximum spacing of repeated rehearsals, and the numbers of forward and consecutive transitions. For recall, dependent variables included recall accuracy, PFR, and forward and consecutive transition rates in recall.

Materials. The set of 320 words used as stimuli was the same as in Experiment 1, but for greater comparability to previous rehearse-aloud studies (e.g., Ward et al., 2003), the list length was 20. Words were randomly assigned to eight high- and eight low-frequency sets. Item order within each list was counterbalanced using a balanced Latin square, such that every word appeared in every serial position and followed every other word in the set equally often across participants.

Procedure. Participants were tested individually in 2-hr sessions using the same assessments, questionnaires, and apparatus as in Experiment 1. The computerized experiment had 16 study-test cycles in which participants rehearsed and then recalled 20-item word lists. At study, participants were instructed to say each word aloud and also speak aloud any other list items that came to mind while committing the words to memory. Words appeared on the computer monitor one at a time for 1 s each, with 3 s between words in which the screen was blank. At the end of each 20-word sequence, the word *Recall* would appear on the screen and participants performed a free-recall task. There were eight rehearsal-recall cycles each for high- and low-frequency word lists. The cycles were blocked by frequency, changing after every four cycles, and condition order was counterbalanced across participants. Participants were allowed to rest between lists when they felt the need to do so. All sessions were recorded and later transcribed.

Results and Discussion

Approach to analysis. The approach to analysis was the same as in Experiment 1.

Rehearsal rates. The rehearsal rate (average rehearsals per cycle) was defined as the total number of rehearsals for a list divided by 20 cycles. As expected, rehearsal rates were lower for low-frequency (M = 3.15) than for high-frequency (M = 3.72) word lists (see Table 4, Figure 2a); this effect was significant in a linear mixed-effects model with frequency as fixed effect and random intercepts for participants and items (b = −.561, SE = .038, t = 14.63, p < .001).

Emergent rehearsal properties. The five emergent rehearsal measures were analyzed using linear mixed effects models with word frequency and linear and quadratic trends on serial position as fixed effects, random intercepts for participants and items, and random slopes for frequency across participants.

Rehearsals as a function of serial position. Rehearsals of items from each serial position are illustrated in Figure 6. For individual items, the number of rehearsals ranged from zero to 23, but data for numbers above 10 were relatively sparse. To keep extreme high values from unduly influencing analyses with the number of rehearsals as a dependent variable, all scores over 10 were coded as 11. Words from early serial positions were rehearsed more than items from later positions (b = −1.798, SE = .014, t = 128.349, p < .001), and the function was curved (b = 1.061, SE = .014, t = 75.740, p < .001). Low-frequency words were rehearsed less than high-frequency words (b = −.435, SE = .051, t = 8.847, p < .001), and this effect was stronger for items from earlier serial positions (b = .328, SE = .028, t = 11.704, p < .001). The curvature of the function was stronger for high-frequency words (b = −.573, SE = .028, t = 2.053, p = .04). This analysis confirms the prediction that items from early serial positions would be rehearsed fewer times in low- than in high-frequency lists (as in Tan & Ward, 2000; Ward et al., 2003).

Last rehearsal set. The final cycle in which each item was rehearsed is a measure of how recently an item was rehearsed and was analyzed as a function of serial position. The serial position function (see Figure 6) had an asymmetric U shape in which the scores systematically decreased across early serial positions (about the first five), before the upward linear increase from Positions 8 to 20. A linear trend indicated that items from later serial positions were rehearsed to a later point (b = 2.785, SE = .024, t = 114.461, p < .001), but this upward trajectory was qualified by a quadratic trend (b = 1.833, SE = .024, t = 75.361, p < .001). Low-frequency words were not rehearsed as far into the sequence as high-frequency words (b = −.485, SE = .085, t = 5.705, p < .001), with a stronger effect for items from early serial positions (b = −.396, SE = .049, t = 8.141, p < .001). A marginal interaction of frequency and the quadratic trend suggests greater curvature for low-frequency words (b = .094, SE = .048, t = 1.936, p = .053). This analysis confirms the prediction that in low-frequency word lists, items from early serial positions would not be rehearsed as far into the sequence (as in Ward et al., 2003).

Rehearsal spacing. A maximal spacing measure was derived for each item (within each participant as the maximum lag from one rehearsal to the next). The maximum spacing ranged from one to 80, but scores over 10 were sparse, and there were extreme scores, so this variable was also recoded such that scores over 10 were recoded as 11. Mean maximum spacing lags are illustrated in Figure 7 as a function of serial position. A linear trend on serial position showed that spacing was greater for items from early serial positions (b = −1.282, SE = .023, t = 55.289, p < .001), and this function was curved (b = .048, SE = .023, t = 2.102, p = .036). Maximal spacing was narrower for low-frequency than for high-frequency lists (b = −.642, SE = .093, t = 6.924, p < .001), but the frequency effect did not vary as a function of serial position (ts < 1). This analysis confirms the prediction that spacing between repeated rehearsals of words from early serial positions would be narrower for low-frequency than for high-frequency lists.

Rehearsal transitions. Two transition measures were computed. First, the forward transition measure was the number of times each item elicited a transition to an item from a later serial position (scores over 10 were recoded as 11), and mean counts are illustrated in Figure 7 as a function of serial position. This measure has only one possible value (zero) for the last serial position, so these items were excluded from all analyses involving forward transitions. As expected, items from earlier serial positions showed more forward transitions (b = −1.332, SE = .014, t = 97.574, p < .001), and this function was curved (b = 1.180, SE = .014, t = 87.071, p < .001). Low-frequency words exhibited fewer forward transitions (b = −.279, SE = .037, t = 7.601, p < .001), and this effect was stronger for items from early serial positions (b = .265, SE = .027, t = 9.760, p < .001) and decreased in a curved function (b = −.105, SE = .027, t = 3.885, p < .001).

Second, the consecutive transition measure was the number of times each item elicited a transition to an item from an adjacent

Table 4
Rehearsal Efficiency in Experiments 3 and 4

Group / Task language / Word frequency	Rehearsals/cycle[a] M	SE
Monolingual English		
English		
High	3.72	.08
Low	3.15	.06
English dominant		
English		
High	3.75	.15
Low	3.21	.13
Spanish		
High	3.15	.11
Low	2.63	.09
Spanish dominant		
English		
High	3.56	.10
Low	3.10	.09
Spanish		
High	3.56	.12
Low	3.19	.11

[a] Total number of rehearsals ÷ 20 cycles.

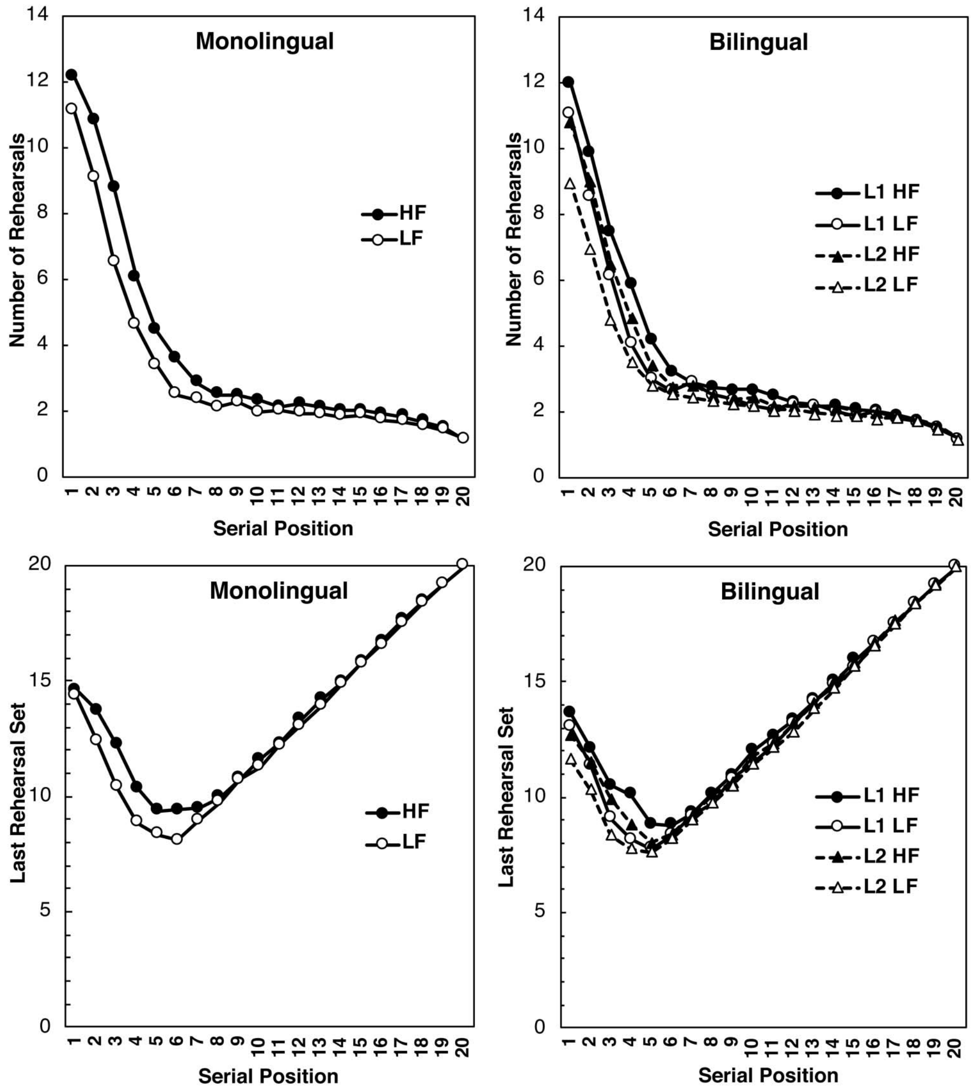

Figure 6. Number and recency of rehearsals as a function of serial position in Experiments 3 and 4. Mean number of rehearsals and last rehearsal set for items from each serial position. HF = high frequency; LF = low frequency; L1 = more proficient language; L2 = less proficient language.

serial position (the immediately previous or subsequent position), and means are illustrated in Figure 7. (Scores over 10 were recoded as 11.) Items from earlier serial positions showed more consecutive transitions ($b = -1.060$, $SE = .011$, $t = 100.338$, $p < .001$), and this function was curved ($b = .774$, $SE = .011$, $t = 73.237$, $p < .001$). Low-frequency words exhibited fewer consecutive transitions than high-frequency words ($b = -.271$, $SE = .028$, $t = 9.697$, $p < .001$), and this effect was stronger in early serial positions ($b = .283$, $SE = .021$, $t = 13.389$, $p < .001$), decreasing in a curved function ($b = -.141$, $SE = .021$, $t = 6.661$, $p < .001$). These analyses confirm the prediction that items from early serial positions would exhibit fewer forward and consecutive transitions in low-frequency lists than in high-frequency lists.

Emergent rehearsal properties as a function of rehearsal rate. Each of the preceding analyses was repeated with rehearsal rate added as a fixed factor. Note that rehearsal rate is orthogonal to serial position, because all 20 items in a trial have the same value. Consistent with Figure 2d–h, rehearsal rate was a significant predictor in all cases (ps $< .001$), and when rehearsal rate was controlled, the word frequency effect and its interaction with the linear trend on serial position were eliminated or reduced to less than half of their original sizes (see Table 5).

Recall accuracy. Recall accuracy was analyzed as a function of serial position and as a function of three rehearsal measures in logistic mixed-effects regression models with word frequency and serial position or the quantitative rehearsal measure as fixed effects. Rehearsal measures were centered. Models included random intercepts for participants and items and random slopes for frequency across participants. Whereas serial position was manipulated and can be causally associated with recall performance, the rehearsal characteristics were measured and their relationships to recall performance may not be causal.

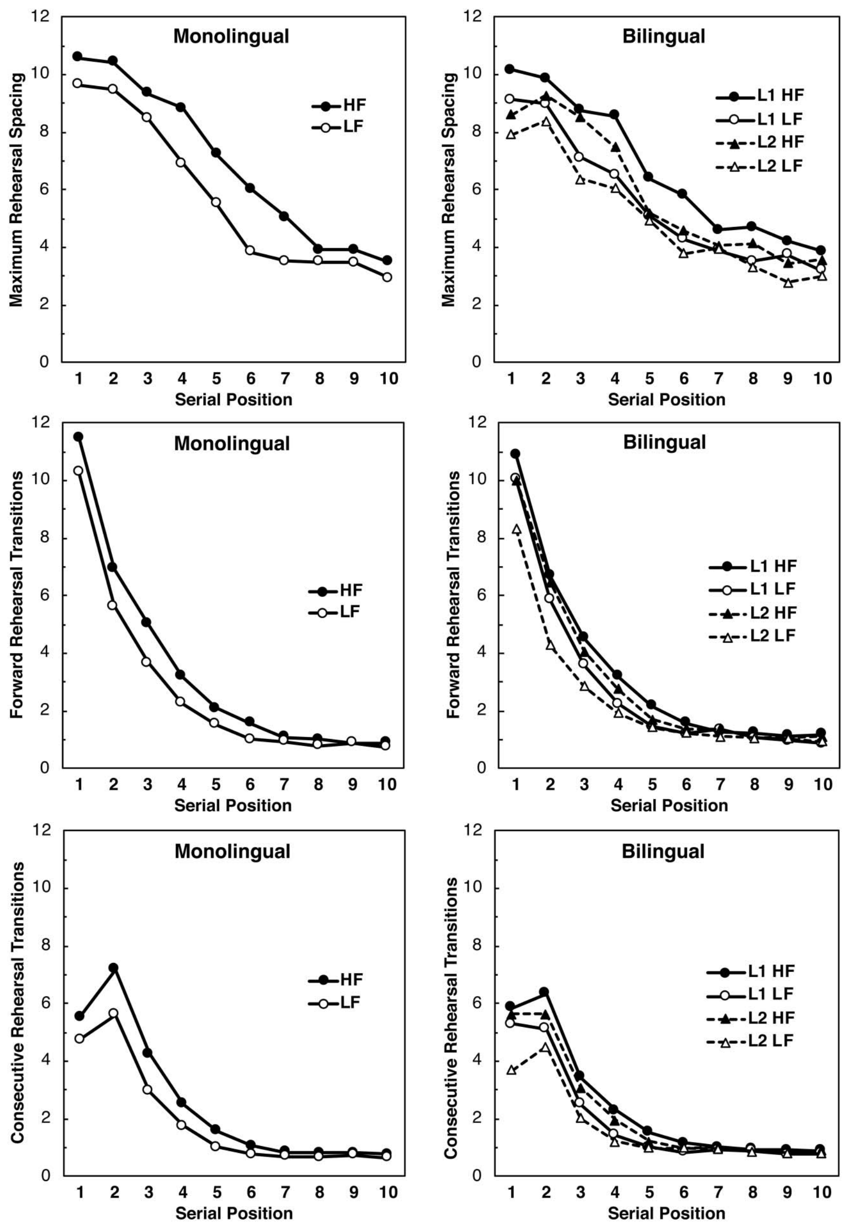

Figure 7. Rehearsal characteristics as a function of serial position in Experiments 3 and 4. Mean maximum rehearsal spacing, number of forward rehearsal transitions, and number of consecutive rehearsal transitions for items from Serial Positions 1–10. (Scores decreased or stayed the same across Serial Positions 11–20.) HF = high frequency; LF = low frequency; L1 = more proficient language; L2 = less proficient language.

Recall accuracy as a function of serial position. As illustrated in Figure 8, serial position effects on recall accuracy showed the typical U-shaped function, as indicated by a significant quadratic trend across serial positions (b = .776, SE = .015, z = 51.521, p < .001). In contrast to Experiment 1, a negative linear trend across serial positions indicated that the primacy effect was stronger than the recency effect (b = −.308, SE = .014, z = 22.207, p < .001), perhaps because of the rehearse aloud protocol. Overall, the high-frequency advantage was not reliable (b = −.092, SE = .067, z = 1.387, p = .165), and we suspect that the reason is that with twice as many items in each list, the items from early serial positions where the primacy effect is observed contribute a smaller proportion of the observations. Consistent with this interpretation, a significant interaction of word frequency and the linear trend on serial position (b = .113, SE = .028, z = 4.067, p < .001) indicates a high-frequency advantage for items from earlier but not later serial positions. Word frequency did not interact with the quadratic trend on serial position (p > .20). This analysis shows weaker primacy effects for low-frequency words relative to high-frequency words, consistent with Experiments 1 and 2 and previous research (e.g., Raymond, 1969; Tan & Ward, 2000).

Table 5
Mediation of Frequency and Proficiency Effects on Rehearsal Properties by Rehearsal Rate

Experiment (Predictor)	Main effect without rehearsal rate			Main effect with rehearsal rate			Main × Linear without rehearsal rate			Main × Linear with rehearsal rate		
Outcome	*b*	*t*	*p*	*b*	*t*	*p*	*b*	*t*	*p*	*b*	*t*	*p*
Experiment 3 (Frequency)												
Number of rehearsals	−.435	8.487	<.001	−.046	−1.005	.316	.328	11.704	<.001	−.022	.778	.436
Recency of rehearsal	−.485	5.705	<.001	.025	.306	.760	.396	8.141	<.001	−.148	2.987	.003
Maximum spacing	−.642	6.924	<.001	−.143	1.638	.103	.011	.243	.808	−.034	.710	.478
Forward transitions	−.279	7.601	<.001	−.063	1.785	.075	.265	9.760	<.001	.016	.570	.569
Consecutive transitions	−.271	9.697	<.001	−.094	3.272	.001	.283	13.389	<.001	.032	1.510	.131
Experiment 4 (Frequency)												
Number of rehearsals	−.381	9.748	<.001	.040	1.062	.289	−.319	11.702	<.001	.066	2.365	.018
Recency of rehearsal	−.419	6.156	<.001	−.005	.071	.944	.352	8.459	<.001	.067	1.347	.178
Maximum spacing	−.646	9.483	<.001	−.299	4.469	<.001	.115	2.607	<.001	.024	.513	.608
Forward transitions	−.288	8.958	<.001	−.102	3.166	.002	.267	9.923	<.001	.031	1.115	.265
Consecutive transitions	−.240	10.651	<.001	−.102	4.449	<.001	.277	14.125	<.001	.071	3.529	<.001
Experiment 4 (Proficiency)												
Number of rehearsals	.185	10.797	<.001	.027	1.690	.091	−.163	11.961	<.001	.069	4.811	<.001
Recency of rehearsal	.253	7.348	<.001	.056	1.780	.075	−.268	11.172	<.001	−.114	4.479	<.001
Maximum spacing	.309	11.286	<.001	.139	4.562	<.001	−.145	6.453	<.001	−.114	4.639	<.001
Forward transitions	.106	6.407	<.001	.014	.837	.402	−.114	8.446	<.001	−.024	1.690	.091
Consecutive transitions	.073	5.917	<.001	.003	.198	.843	−.102	10.366	<.001	−.018	1.804	.071
Experiments 3 & 4 (M vs. B1)												
Number of rehearsals	−.022	.227	.821	.005	.129	.897	−.112	3.586	<.001	−.122	3.897	<.001
Recency of rehearsal	.174	1.030	.305	.196	1.653	.101	−.283	5.203	<.001	−.282	5.160	<.001
Maximum spacing	.423	1.841	.068	.463	2.406	.018	−.014	.292	.771	−.015	.279	.780
Forward transitions	−.112	1.247	.215	−.095	1.218	.225	−.026	.849	.396	−.038	1.238	.216
Consecutive transitions	−.055	.788	.432	−.034	.499	.619	−.022	.925	.355	−.041	1.739	.082
Experiments 3 & 4 (M vs. B2)												
Number of rehearsals	.048	.514	.608	−.043	1.359	.174	−.267	8.558	<.001	−.207	6.526	<.001
Recency of rehearsal	.315	1.821	.071	.200	1.591	.114	−.460	8.454	<.001	−.334	6.013	<.001
Maximum spacing	.653	2.714	.008	.571	2.793	.006	−.185	3.837	<.001	−.141	2.845	.004
Forward transitions	−.001	.011	.991	−.035	.430	.668	−.134	4.411	<.001	−.087	2.813	.005
Consecutive transitions	−.003	.037	.970	−.028	.394	.694	−.129	5.517	<.001	−.093	3.894	<.001

Note. Main × Linear = interaction of specified effect with the linear trend on serial position. M = monolingual; B1 = bilingual dominant language; B2 = bilingual nondominant language.

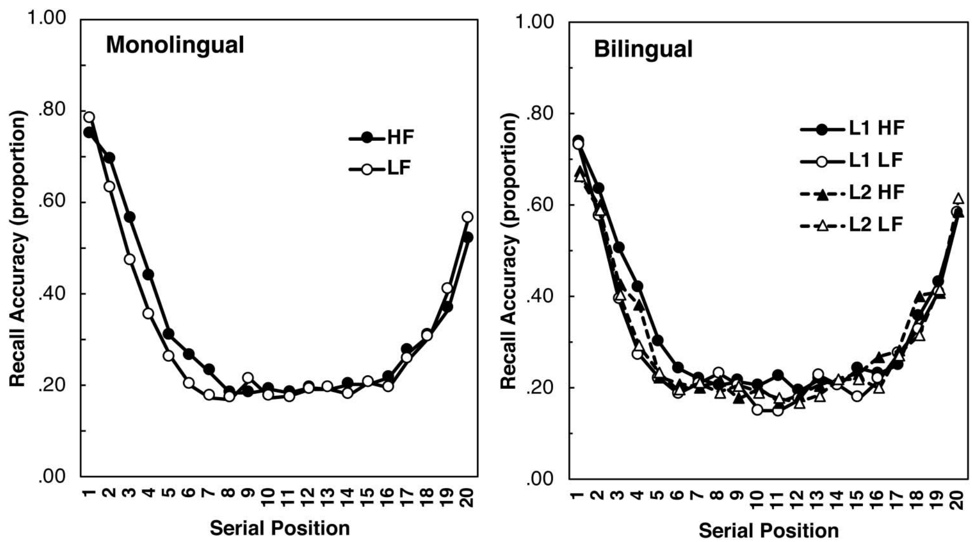

Figure 8. Serial position curves for recall accuracy in Experiments 3 and 4. Mean recall rates at each serial position. HF = high frequency; LF = low frequency; L1 = more proficient language; L2 = less proficient language.

Recall accuracy as a function of rehearsal properties. Three analyses were performed to find out how well rehearsal characteristics predicted recall accuracy (Figure 2i–k, 2n) and how they might mediate the effects of word frequency on recall accuracy. Because these measures were moderately correlated with each other, separate analyses were performed to examine the association of each measure with recall accuracy at the individual item level within participants. These measures were also moderately correlated with serial position, so serial position was not included in the models. Figure 9 shows recall accuracy as a function of the number of rehearsals, recency of rehearsal (functional serial position), and maximal rehearsal spacing.

Recall accuracy as a function of the number of rehearsals. Words that were rehearsed more times were more likely to be recalled (b = 1.460, SE = .021, z = 69.327, p < .001), and this effect was stronger for low-frequency than for high-frequency words (b = .140, SE = .041, z = 3.403, p < .001). Thus, the increase in recall accuracy with more rehearsals was greater for low-frequency words than for high-frequency words. In fact, putting the number of rehearsals into the model reversed the direction of the frequency effect (b = .163, SE = .066, z = 2.486, p = .013), with low-frequency words being recalled at a higher rate at higher numbers of rehearsals.

Recall accuracy as a function of recency of rehearsal. Recall was examined as a function of recency of rehearsal, as summarized by a functional serial position score in which higher scores indicate more recent rehearsal (Tan & Ward, 2000; Ward et al., 2003). Functional serial position scores were obtained by rank ordering the 20 words in each list according to their last rehearsal, assigning a 20 to the last word rehearsed, 19 to the second last word rehearsed, and so forth. Words that were more recently rehearsed (i.e., from later functional serial positions) were recalled at a higher rate than words that were less recently rehearsed (b = 1.210, SE = .019, z = 63.738, p < .001), and this function was curved with higher slope at later positions (b = .478, SE = .018, z = 26.218, p < .001). In this analysis the tendency for high-frequency words to be better recalled was not reliable (b = −.101, SE = .063, z = 1.602, p = .109), and did not interact with linear or quadratic serial position trends (ps > .2).

Recall accuracy as a function of rehearsal spacing. Words with greater maximal spacing between rehearsal repetitions were recalled at a higher rate (b = 1.151, SE = .024, z = 47.441, p < .001), but this effect did not interact with word frequency (z < 1). The word frequency effect became significant in the reverse direction, with low-frequency words exhibiting higher accuracy for equivalent degrees of spacing (b = .229, SE = .060, z = 3.800, p < .001).

Summary. The number, recency, and spacing of rehearsals were positively associated with recall accuracy (Figure 2i–k, 2n), and as explained previously, these three rehearsal measures had lower scores for low-frequency words. The high-frequency recall advantage (here nonsignificant) showed a significant reversal when number of rehearsals or maximum rehearsal spacing were controlled, which suggests that these rehearsal characteristics mediate the word-frequency effect on recall accuracy. Controlling for recency of rehearsal did not change the frequency effect. We will return to this issue in the General Discussion.

Position of the first item recalled. The serial position of the first item recalled was analyzed in the same manner as in Experiment 1. As shown in Figure 10, the position of the first item recalled tended to be later for low-frequency than for high-frequency lists (b = 1.040, SE = .521, t = 1.997, p < .049). We also analyzed the effect of word frequency on the *functional* serial position of the first item recalled, because the item presented first at study is often recalled first and is often rehearsed all the way to the last rehearsal set, and the first word recalled is usually one rehearsed in the last rehearsal set (93% of the time in Rundus, 1971; and 68% in Ward et al., 2003). Consistent with these findings, 66% of recall trials began with words that were rehearsed in the last rehearsal set. By this measure, the frequency effect was

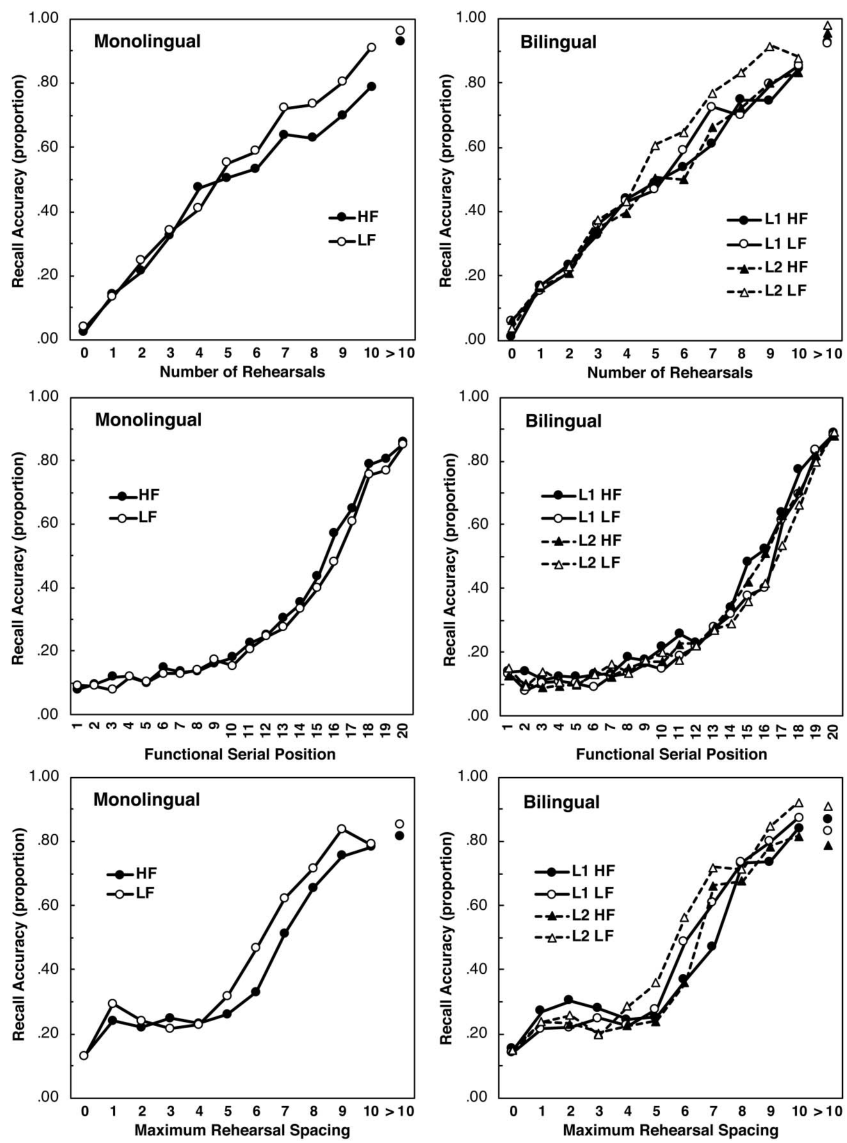

Figure 9. Recall accuracy as a function of number, recency, and spacing of rehearsals in Experiments 3 and 4. Mean recall rates as a function of number or rehearsals, recency of rehearsal (functional serial position), and maximal rehearsal spacing. A maximum spacing value of 0 indicates items rehearsed only once. HF = high frequency; LF = low frequency; L1 = more proficient language; L2 = less proficient language.

more robust ($b = .627$, $SE = .180$, $t = 3.492$, $p < .001$). This effect was evident primarily in a greater proportion of first recalls from the final two functional serial positions. The proportion of recall trials initiated with an item from one of the last five nominal serial positions was 45% and the proportion in the last five *functional* serial positions was 83%. Items from early serial positions were only recalled first if they continued to be rehearsed beyond the first five items (see Figure 10). Thus, the monotonic functional serial position or recency of rehearsal function was an excellent predictor of which item would be recalled first (Figure 2k), and it was more sensitive to the word-frequency effect than the nominal serial position function.

Temporal association encoding. Scores for three measures of temporal encoding (see Table 6) were analyzed in logistic mixed effects models with frequency and linear and quadratic trends on serial position as fixed effects, random intercepts for participants and items, and random slopes for frequency across participants.

Forward recall transitions. As in Experiments 1 and 2, items from the first and last serial positions were excluded. The proportion of forward transitions was higher for items from earlier serial positions ($b = -.771$, $SE = .037$, $z = 20.877$, $p < .001$). In contrast to Experiments 1 and 2, the frequency effect was not significant ($z < 1$), nor were any other effects ($ps > .20$).

Consecutive recall transitions. Overall, 30% of transitions during recall were to items in consecutive serial positions. The proportion of consecutive transitions followed a U-shaped pattern across serial positions ($b = -.367$, $SE = .032$, $z = 11.593$, $p <$

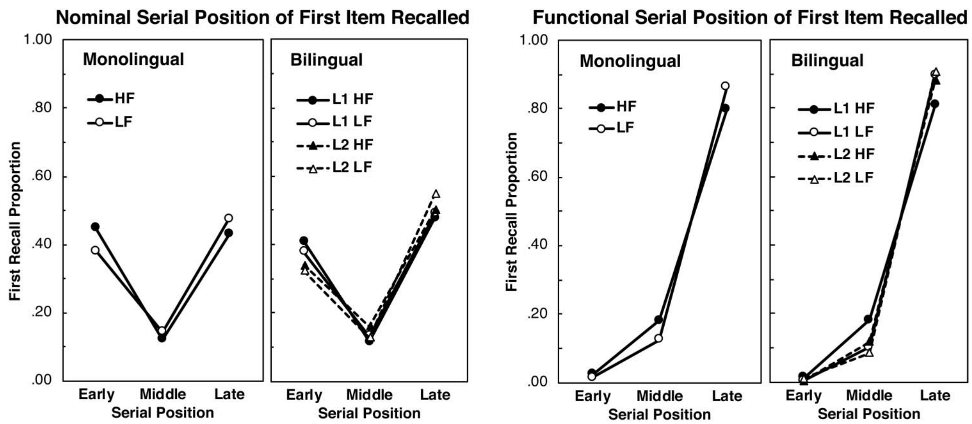

Figure 10. Serial position and functional serial position of the first item recalled in Experiments 3 and 4. Mean proportions of recall trials beginning with items from early (1–5), middle (6–15), and late (16–20) nominal serial positions (left) or functional serial positions (right). HF = high frequency; LF = low frequency; L1 = more proficient language; L2 = less proficient language.

.001), with a higher proportion for items from earlier serial positions than items from later serial positions (b = .268, SE = .031, z = 8.675, p < .001). Transitions among low-frequency words were less likely to be consecutive than transitions among high-frequency words (b = .229, SE = .071, z = 3.228, p = .001), indicating that weaker or less numerous temporal associations are formed among low-frequency than among high-frequency words. No other effects were significant (zs < 1).

Recall transitions as a function of rehearsal properties. Because consecutive rehearsals of two items has been claimed to be a precursor to recalling the same two items consecutively at test (Laming, 2008, 2009), we created a new functional proximity measure based on the minimum distance between rehearsals of the two items recalled consecutively. Overall, 67% of transitions during recall were functionally consecutive (rehearsed consecutively). Here, the proportion of consecutive transitions was higher for items from earlier serial positions (b = .655, SE = .032, z = 20.722, p < .001), and the serial position function was curved (b = −.279, SE = .030, z = 9.393, p < .001). Functionally consecutive transitions occurred at a lower rate for low-frequency words than for high-frequency words (b = .251, SE = .062, z = 4.025, p < .001) but did not interact with the serial position components (ps > .20). Thus, the impact of inefficient rehearsal was more evident when consecutive transitions were defined by

Table 6
Overall Recall Accuracy and Proportions of Forward and Consecutive Transitions in Experiments 3 and 4

Group / Task language / Word frequency	Overall accuracy M	Overall accuracy SE	Forward transitions[a] M	Forward transitions[a] SE	Consecutive transitions M	Consecutive transitions SE	Func consec transitions M	Func consec transitions SE
Monolingual English								
English								
High	.323	.009	.545	.020	.317	.013	.706	.014
Low	.305	.007	.520	.018	.270	.010	.646	.014
English dominant								
English								
High	.337	.015	.551	.030	.306	.022	.700	.019
Low	.310	.012	.522	.037	.292	.018	.662	.025
Spanish								
High	.304	.016	.534	.026	.264	.019	.631	.031
Low	.286	.013	.540	.029	.257	.024	.613	.026
Spanish dominant								
English								
High	.327	.016	.534	.029	.308	.020	.692	.021
Low	.285	.014	.560	.031	.307	.022	.664	.025
Spanish								
High	.326	.016	.515	.027	.271	.022	.665	.025
Low	.313	.021	.551	.028	.270	.022	.643	.030

[a] Excluding transitions from Serial Positions 1 and 20.

consecutive rehearsals generated by the participant rather than consecutive presentations from an external source.

To further document the correspondence between forward and consecutive transitions from rehearsal to recall, we plotted recall outcomes as a function of the number of forward or consecutive rehearsal transitions at study, using data pooled across Experiments 3 and 4 (see Figure 11). Two additional analyses were conducted on recall transitions with the corresponding rehearsal transition measures and word frequency as fixed effects. Consistent with Figure 2l, recalled items that had elicited more forward transitions in rehearsal were more likely to elicit forward transitions in recall ($b = 1.113$, $SE = .045$, $z = 24.576$, $p < .001$). Items that elicited forward transitions at recall had an average of 5.4 forward rehearsal transitions; items that elicited backward recall transitions had an average of 1.4 forward rehearsal transitions. Similarly, consistent with Figure 2m, recalled items that elicited more consecutive transitions in rehearsal had were more likely to elicit consecutive transitions in recall ($b = -1.040$, $SE = .035$, $z = 29.745$, $p < .001$). Items that elicited consecutive transitions at recall had an average of 4.6 consecutive rehearsal transitions; items that elicited nonconsecutive recall transitions had an average of 1.6 consecutive rehearsal transitions. Thus, as predicted by Laming (2008), specific transitions made in rehearsal were repeated in recall. No effects involving word frequency were reliable ($zs < 1$). Thus, the word frequency effect observed for consecutive recall transitions in Experiments 1, 2, and 3 was eliminated, which suggests that transition patterns in rehearsal mediate the effects of word frequency on transition patterns in recall.

Experiment 4

The results of Experiment 2 suggest that bilinguals have less efficient rehearsal in their less proficient language. Experiment 4 replicates tests of the language proficiency effects found in Experiment 2 and directly tests whether rehearsal patterns differ as a function of language proficiency using a rehearse-aloud protocol. Similarly, tests of the word frequency effects found in Experiment 3 are replicated. Finally, Experiment 4 replicates tests of the relationships between rehearsal and retrieval dynamics and the sensitivity of rehearsal-dependent recall measures to word frequency and further tests their sensitivity to language proficiency.

Method

Participants. Participants were 80 Spanish-English bilingual students at the University of Texas at El Paso who were compensated with course research credit or a payment of $20. Half of the participants were English dominant and half were Spanish dominant, with dominance determined as described in Experiment 2. Participant characteristics are summarized in Table 2.

Design. The experimental conditions formed a 2 (language) × 2 (word frequency) × 20 (serial position) repeated-measures design. Half of the lists were presented in English and half in Spanish. In each language, half had high-frequency and half had low-frequency words. Each list had 20 items/serial positions. Dependent variables were the same as in Experiment 3.

Materials. The experimental stimulus words were the same as in Experiment 2 and were randomly assigned to eight sets of 20 words each. Sets were randomly assigned to English and Spanish languages, with this assignment counterbalanced across participants. For each language and frequency combination, there were four sets of 20 words each, and the order of the words within each set was counterbalanced using a balanced Latin square, such that every word appeared in every serial position and every word followed every other word in the set equally often across participants. No participant saw translation equivalents across lists.

Procedure. Participants were tested individually in 2-hr sessions using the same assessments, questionnaires, and apparatus as in Experiment 3. The computerized experiment required participants to rehearse and then free recall 16 lists of 20 words each, with the same instructions, trial structure, and timing as in Experiment 3. There were four study-test cycles for each Language × Frequency combination, blocked by language and frequency with the order counterbalanced across participants. Participants were allowed to rest between lists when they felt the need to do so. All sessions were recorded and later transcribed.

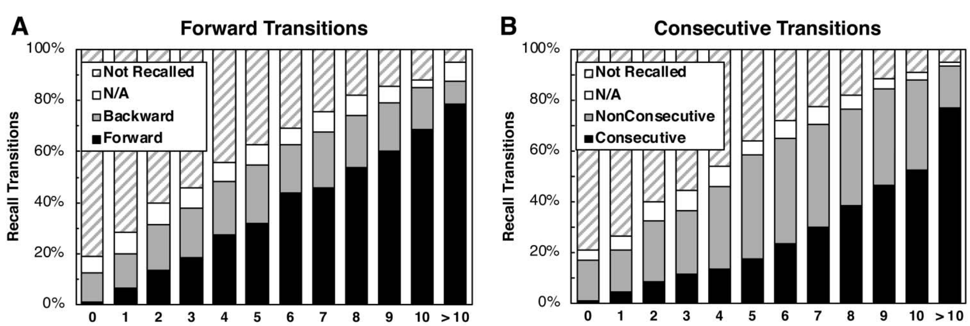

Figure 11. Transitions in recall as a function of forward and consecutive transitions in rehearsal. (A) Proportions of items with forward or backward transitions, recalled with no transition, or not recalled as a function of the number of forward transitions from that item in rehearsal at study. (B) Proportions of items with consecutive or non-consecutive transitions, recalled with no transition, or not recalled as a function of number of consecutive transitions from that item in rehearsal at study.

Results

Approach to analysis. The approach to analysis was the same as in Experiment 3. The language proficiency variable was treated in the same manner as for Experiment 2.

Rehearsal rate. As shown in Table 4, rehearsal rates were lower for low-frequency ($M = 3.03$) than for high-frequency lists ($M = 3.50$) and lower for L2 lists ($M = 3.11$) than for L1 lists ($M = 3.43$), consistent with Figure 2a–b. These effects were significant in a linear mixed-effects model with word frequency and proficiency as fixed effects, random intercepts for participants and items, and random effects for frequency across participants (frequency: $b = -.482$, $SE = .036$, $t = 13.578$, $p < .001$; proficiency: $b = .235$, $SE = .018$, $t = 13.115$, $p < .001$).

Emergent rehearsal properties. Rehearsal properties were analyzed using the truncated variables described for Experiment 3. Each rehearsal measure was analyzed in a linear mixed-effects model with word frequency, language proficiency, and linear and quadratic trends on serial position as fixed effects and random intercepts for participants and items.

Rehearsals as a function of serial position. The number of rehearsals for items at each serial position are illustrated in Figure 6. Words from earlier positions were rehearsed more ($b = -1.519$, $SE = .014$, $t = 111.626$, $p < .001$), and the serial position function was curved ($b = .861$, $SE = .014$, $t = 63.236$, $p < .001$). As in Experiment 3, there were fewer rehearsals in low-frequency than in high-frequency word lists ($b = -.381$, $SE = .039$, $t = 9.748$, $p < .001$), with a stronger effect for early serial positions ($b = .319$, $SE = .027$, $t = 11.702$, $p < .001$). Lower proficiency in the task language was also associated with fewer rehearsals ($b = .184$, $SE = .017$, $t = 10.797$, $p < .001$), with a stronger effect for earlier serial positions ($b = -.163$, $SE = .014$, $t = 11.961$, $p < .001$). Word frequency and language proficiency interacted with the quadratic trend across serial positions, with lower frequency and lower proficiency being associated with less curvature (frequency: $b = -.108$, $SE = .027$, $t = 3.996$, $p < .001$; proficiency: $b = .095$, $SE = .014$, $t = 6.960$, $p < .001$). No other effects were significant (zs < 1). This analysis confirms the prediction that items from early serial positions would be rehearsed fewer times with low word frequency (as in Experiment 3; Tan & Ward, 2000; Ward et al., 2003) or lower proficiency.

Last rehearsal set. The final cycle in which items were rehearsed is illustrated in Figure 6. There was a positive linear trend across serial positions ($b = 3.269$, $SE = .024$, $t = 136.508$, $p < .001$), and a quadratic trend due to early items being rehearsed further into the sequence than middle items ($b = 1.467$, $SE = .024$, $t = 61.290$, $p < .001$). As in Experiment 3, low-frequency words were not rehearsed as far as high-frequency words ($b = -.418$, $SE = .068$, $t = 6.156$, $p < .001$) with a greater effect for early serial positions ($b = .352$, $SE = .048$, $t = 7.348$, $p < .001$). Lower proficiency was also associated with not rehearsing as far ($b = .253$, $SE = .030$, $t = 8.459$, $p < .001$) with a greater effect for early serial positions ($b = -.268$, $SE = .024$, $t = 11.172$, $p < .001$). Both frequency and proficiency interacted with the quadratic trend, with less curvature for low frequency and lower proficiency (frequency: $b = -.104$, $SE = .048$, $t = 2.181$, $p = .029$; proficiency: $b = .160$, $SE = .024$, $t = 6.692$, $p < .001$). This analysis confirms the prediction that items from early serial positions would not be rehearsed as far with low word frequency (as in Experiment 3; Ward et al., 2003) or lower language proficiency.

Rehearsal spacing. Maximum rehearsal spacing lags (see Figure 7) were analyzed as a function of serial position. As expected, spacing between repeated rehearsals was greater for items from early serial positions ($b = -1.163$, $SE = .022$, $t = 52.410$, $p < .001$), and the function was curved ($b = .065$, $SE = .022$, $t = 2.947$, $p = .003$). Spacing among rehearsals was not as wide for low-frequency as for high-frequency words ($b = -.646$, $SE = .068$, $t = 9.483$, $p < .001$), with a stronger effect for early serial positions ($b = .115$, $SE = .044$, $t = 2.607$, $p = .009$), decreasing in a curved function ($b = .134$, $SE = .044$, $t = 3.037$, $p = .002$). Lower language proficiency was also associated with less spacing among repetitions ($b = .308$, $SE = .027$, $t = 11.286$, $p < .001$), with a stronger effect for early serial positions ($b = -.145$, $SE = .023$, $t = 6.453$, $p < .001$). No other effects were significant (ts < 1). This analysis confirms the prediction that the maximum spacing between repeated rehearsals of words from early serial positions would be narrower for low-frequency word lists and lists in a less proficient language.

Rehearsal transitions. Forward and consecutive transition rates are illustrated in Figure 7. Words from early serial positions elicited more forward rehearsal transitions ($b = -1.177$, $SE = .013$, $t = 87.568$, $p < .001$), and the function was curved ($b = 1.004$, $SE = .013$, $t = 74.652$, $p < .001$). There were fewer forward rehearsal transitions for low- than for high-frequency lists ($b = .288$, $SE = .032$, $t = 8.958$, $p < .001$), with a stronger effect for early serial positions ($b = .267$, $SE = .027$, $t = 9.923$, $p < .001$), decreasing in a curved function ($b = -.151$, $SE = .027$, $t = 5.631$, $p < .001$). Lower language proficiency was also associated with fewer forward transitions ($b = .106$, $SE = .016$, $t = 6.407$, $p < .001$), with a stronger effect for words from early serial positions ($b = -.114$, $SE = .013$, $t = 8.446$, $p < .001$), decreasing in a curved function ($b = .103$, $SE = .013$, $t = 7.655$, $p < .001$). The three-way interactions were not reliable ($p > .2$).

More consecutive rehearsal transitions were made for words from early serial positions ($b = -.910$, $SE = .010$, $t = 92.970$, $p < .001$), and the function was curved ($b = .641$, $SE = .010$, $t = 65.482$, $p < .001$). There were fewer consecutive transitions for low- than for high-frequency lists ($b = -.240$, $SE = .023$, $t = 10.651$, $p < .001$), with a stronger effect in early serial positions ($b = .276$, $SE = .020$, $t = 14.125$, $p < .001$), decreasing in a curved function ($b = -.180$, $SE = .020$, $t = 9.191$, $p < .001$). Lower proficiency was also associated with fewer consecutive transitions ($b = .073$, $SE = .012$, $t = 5.917$, $p < .001$), with a stronger effect in early serial positions ($b = -.102$, $SE = .010$, $t = 10.366$, $p < .001$), decreasing in a curved function ($b = .087$, $SE = .010$, $t = 8.868$, $p < .001$). The three-way interactions were not reliable ($p > .20$). These analyses confirm the prediction that items from early serial positions would exhibit fewer forward and consecutive rehearsal transitions in low-frequency lists and lists in a less proficient language.

Emergent rehearsal properties as a function of rehearsal rate. Each of the preceding analyses was repeated with rehearsal rate added to the model. Consistent with Figure 2d–h, rehearsal rate was a significant predictor in all cases (ps $< .001$), and when it was controlled, the effects of word frequency and language proficiency and their interactions with the linear trend on serial position were eliminated or reduced to less than half of their original size (see Table 5).

Recall accuracy. Recall was analyzed as a function of serial position and as a function of the rehearsal measures using the same models as in Experiment 3 with language proficiency added as a fixed effect. All rehearsal measures were centered.

Recall accuracy as a function of serial position. Serial position effects on recall accuracy are illustrated in Figure 8. Recall accuracy showed the typical U-shaped serial position function (b = .731, SE = .015, z = 49.348, $p < .001$), and the primacy effect was stronger than the recency effect (b = −.155, SE = .014, z = 11.292, $p < .001$). Recall was less accurate for low- than for high-frequency lists (b = −.132, SE = .054, z = 2.453, $p < .001$), and a marginal interaction suggests that the low-frequency disadvantage was stronger in early serial positions (b = .051, SE = .027, z = 1.862, p = .062). The overall effect of language proficiency was not reliable (b = .023, SE = .018, z = 1.268, p = .205), but lower proficiency was associated with less accurate recall in early but not late serial positions (b = −.079, SE = .014, z = 5.722, $p < .001$) as in Experiment 2. No other interactions reached significance (ps > .2). This analysis shows that primacy effects are weaker with low word frequency (as in Experiments 1, 2, and 3; Raymond, 1969; Tan & Ward, 2000) and lower language proficiency (as in Experiment 2; Yoo & Kaushanskaya, 2016).

Recall accuracy as a function of rehearsal properties. Three analyses were performed to test hypotheses about how rehearsal characteristics impact recall accuracy (Figure 2j–k, 2n) and mediate the effects of word frequency and language proficiency on recall accuracy. Figure 9 shows recall accuracy as a function of number, recency, and spacing of rehearsals.

Recall accuracy as a function of the number of rehearsals. Words rehearsed more times were more accurately recalled (b = 1.593, SE = .028, z = 57.848, $p < .001$). The reversal of the word frequency effect did not reach significance (b = .063, SE = .054, z = 1.177, p = .239), but the effect of word frequency on accuracy increased with more rehearsals, a greater increase for low- than for high-frequency words with additional rehearsals (b = .243, SE = .054, z = 4.508, $p < .001$). Controlling for the number of rehearsals significantly reversed the language proficiency effect, with lower proficiency associated with higher accuracy (b = −.081, SE = .020, z = 3.956, $p < .001$). This effect indicates that although lower proficiency is associated with fewer rehearsals, at a given number of rehearsals, lower proficiency is associated with higher recall. There was a marginal interaction of language proficiency with the number of rehearsals, such that the increase in accuracy with additional rehearsals was greater when proficiency was lower (b = −.049, SE = .026, t = 1.850, p = .064). No other effects were significant (zs < 1).

Recall accuracy as a function of recency of rehearsal. Recall was examined as a function of recency of rehearsal, as summarized by a functional serial position score in which higher scores indicate more recent rehearsal. Words that were more recently rehearsed (i.e., from later functional serial positions) were more accurately recalled (b = 1.146, SE = .019, z = 61.770, $p < .001$), and this function was curved with higher slope at later positions (b = .528, SE = .018, z = 29.343, $p < .001$). Low-frequency words were less accurately recalled than high-frequency words (b = −.155, SE = .056, z = 2.761, p = .006), and there was a marginal trend suggesting a larger effect for items with more recent rehearsals (b = −.063, SE = .037, z = 1.732, p = .083). There was a marginal effect suggesting that lower proficiency was associated with lower accuracy (b = .058, SE = .031, z = 1.834, p = .067). Lower proficiency was associated with greater curvature (b = −.047, SE = .018, z = 2.632, p = .008). There was also a marginal three-way interaction of frequency, proficiency, and the linear trend on serial position (b = .062, SE = .035, z = 1.744, p = .081). No other effects were significant (ps > .2).

Recall accuracy as a function of rehearsal spacing. Words with greater spacing between repeated rehearsals were more accurately recalled (b = 1.506, SE = .039, z = 38.659, $p < .001$). Having the rehearsal spacing variable in the model reversed the word frequency effect, such that with spacing controlled, recall accuracy was higher for low- than for high-frequency words (b = .133, SE = .053, z = 2.534, p = .011), and this effect was stronger with greater spacing (b = .342, SE = .073, z = 4.693, $p < .001$). The proficiency effect also reversed, such that with spacing controlled, lower proficiency was associated with more accurate recall (b = −.078, SE = .024, z = 3.226, p = .001), and this effect was stronger with greater spacing (b = −.096, SE = .037, z = 2.583, p = .010). No other effects were significant (z < 1).

Summary. As in Experiment 3, the number, recency, and spacing of rehearsals were positively associated with recall accuracy (Figure 2i–k and 2n), and as explained previously, all of the rehearsal measures had higher scores for high-frequency words and words in a more proficient language. Consistent with the results of Experiment 3, the effects of word frequency and language proficiency on recall accuracy were reversed when number or maximal spacing of rehearsals was controlled, which supports the contention that rehearsal patterns mediate the effects of word frequency and language proficiency on free recall accuracy. However, in the recency-of-rehearsal analysis, frequency and proficiency effects did not change.

Position of the first item recalled. Frequency distributions for the nominal and functional serial positions of the first items recalled are illustrated in Figure 10. The serial position of the first item recalled was analyzed in a linear mixed-effects regression model with word frequency and language proficiency as fixed effects and random intercepts for participants. The effects of frequency and language proficiency were in the direction of later PFR for low-frequency words and for lower language proficiency but were not statistically reliable (frequency: b = .584, SE = .373, t = 1.564, p = .118; proficiency: b = −.399, SE = .232, t = 1,715, p = .087). However, when the same analysis was conducted based on the *functional* serial position of the first item recalled, consistent with Figure 2k and the results of Experiment 3, low-frequency and lower proficiency were associated with significantly later PFR (frequency: b = .475, SE = .149, t = 3.192, p = .002; proficiency: b = −.182, SE = .080, t = 2.261, p = .024). The effects of frequency and proficiency did not interact (ps > .10). The proportion of recall trials initiated with an item from one of the last five nominal serial positions was 51% and the proportion from the last five *functional* serial positions was 87%. Thus, consistent with the results of Experiment 3, recency of rehearsal was an excellent predictor of the item to be recalled first, and it was more sensitive than nominal serial position to the effects of frequency and proficiency.

Temporal association encoding. Temporal measures of recall performance are given in Table 6 and analyzed using the same logistic mixed effects models as in Experiment 3 with language proficiency added as a fixed effect.

Forward recall transitions. The proportion of forward transitions was higher for items from earlier serial positions ($b = -.666$, $SE = .036$, $z = 18.361$, $p < .001$), and the serial position function was curved ($b = -.187$, $SE = .034$, $z = 5.514$, $p < .001$). High-frequency words were associated with greater curvature of the serial position function ($b = -.160$, $SE = .067$, $z = 2.378$, $p = .017$), and there was some suggestion that higher proficiency was associated with greater curvature ($b = .061$, $SE = .034$, $z = 1.780$, $p = .075$). No other effects were significant (zs < 1).

Consecutive recall transitions. The serial position function for the proportion of consecutive transitions was U-shaped ($b = -.312$, $SE = .032$, $z = 9.890$, $p < .001$), but higher for items from earlier serial positions than for items from later positions ($b = .184$, $SE = .031$, $z = 5.968$, $p < .001$). Lower proficiency was associated with greater curvature ($b = -.090$, $SE = .032$, $z = 2.856$, $p = .004$). No other effects were significant ($p > .10$).

Recall transitions as a function of rehearsal properties. The proportion of functionally consecutive transitions was higher for items from earlier serial positions ($b = .487$, $SE = .032$, $z = 15.435$, $p < .001$), and the serial position function was curved ($b = -.171$, $SE = .030$, $z = 5.717$, $p < .001$). The rate of functionally consecutive transitions was lower for low- than for high-frequency lists ($b = .147$, $SE = .064$, $z = 2.297$, $p = .022$) and lower with lower language proficiency ($b = -.107$, $SE = 035$, $z = 3.058$, $p = .002$). The curvature of the serial position function was greater for high-frequency words and for lower language proficiency (frequency: $b = -.151$, $SE = .060$, $z = 2.538$, $p = .011$; proficiency: $b = -.086$, $SE = .030$, $z = 2.872$, $p = .002$). No other effects were significant ($p > .20$). Consistent with the results of Experiment 3, these analyses show that the measure that considers consecutive rehearsal rather than consecutive serial position is more sensitive to word-frequency and language-proficiency effects.

To further establish the correspondence between forward and consecutive transitions from rehearsal to recall, in addition to the pooled plot in Figure 11, two recall transition analyses were conducted with the corresponding rehearsal transition measures, word frequency, and language proficiency as fixed effects. Consistent with Figure 2l, recalled items with more forward transitions in rehearsal were more likely to elicit forward transitions in recall ($b = 1.844$, $SE = .069$, $z = 26.893$, $p < .001$). Items with forward transitions at recall had an average of 5.2 forward rehearsal transitions; items with backward recall transitions had an average of 1.4 forward rehearsal transitions. In this analysis, a reversed frequency effect emerged, such that with forward rehearsal transitions controlled, low-frequency lists had higher forward transition rates in recall ($b = .372$, $SE = .089$, $z = 4.177$, $p < .001$). Also, the correspondence between forward transitions at rehearsal and recall were stronger for low-frequency words ($b = .304$, $SE = .134$, $z = 2.261$, $p = .024$). No effects involving language proficiency were significant (zs < 1).

Consistent with Figure 2m, recalled items that elicited more consecutive transitions in rehearsal were more likely to elicit consecutive transitions in recall ($b = -.983$, $SE = .040$, $z = 24.738$, $p < .001$). Items with consecutive transitions at recall had an average of 4.0 consecutive rehearsal transitions; items with nonconsecutive recall transitions had an average of 1.5 consecutive rehearsal transitions. There was a marginal effect suggesting that with the number of consecutive rehearsal transitions controlled, low-frequency lists had higher consecutive transition rates in recall ($b = -.118$, $SE = .069$, $z = 1.714$, $p = .087$). No other effects were significant (ps $> .20$). Thus, specific transitions in rehearsal were repeated in recall.

Analysis of Combined Data From Experiments 3 and 4

As with Experiments 1 and 2, the monolingual participants of Experiment 3 and the bilingual participants of Experiment 4 were matched on age, education, and parent education. The analyses reported below include only English language data. To avoid redundancy, only effects involving group are reported.

Results and Discussion

Approach to analysis. The approach to analysis was the same as in the combined analysis of Experiments 1 and 2.

Rehearsal rate. The rehearsal rate did not differ reliably across groups (ts < 1) and frequency effects on rehearsal rate did not differ across groups (ps $> .2$).

Emergent rehearsal properties. The rehearsal measures were analyzed in linear mixed-effects regression models with group, word frequency, and linear and quadratic trends on serial position as fixed effects, random intercepts for participants and items, and random slopes for word frequency across participants.

Rehearsals as a function of serial position. Bilingual L1 and L2 speakers rehearsed items from early but not later serial positions less than monolingual speakers (L1: $b = -.112$, $SE = .031$, $t = 3.586$, $p < .001$; L2: $b = -.267$, $SE = .031$, $t = 8.558$, $p < .001$; Table A.4 in the online supplemental materials). Curvature of the function was less for bilingual L1 and L2 than for monolingual rehearsal (L1: $b = .105$, $SE = .031$, $t = 3.356$, $p < .001$; L2: $b = .180$, $SE = .031$, $t = 5.756$, $p < .001$). No other effects involving group were significant ($p > .10$). This analysis confirms the prediction that bilinguals would rehearse items from early serial positions fewer times than monolinguals.

Last rehearsal set. Monolinguals rehearsed items from earlier but not later serial positions further into the study sequence than bilingual L1 and L2 speakers (L1: $b = -.283$, $SE = .054$, $t = 5.203$, $p < .001$; L2: $b = -.460$, $SE = .054$, $t = 8.454$, $p < .001$; see Table A.5 in the online supplemental materials), with a marginal main effect of group between monolingual and bilingual L2 speakers ($b = .315$, $SE = .173$, $t = 1.821$, $p = .071$). Curvature of the function was greater for monolingual than for bilingual L1 and L2 speakers (L1: $b = .257$, $SE = .054$, $t = 4.748$, $p < .001$; L2: $b = .310$, $SE = .054$, $t = 5.708$, $p < .001$). There were no other significant effects involving group (ps $> .1$). This analysis confirms the prediction that bilinguals would not rehearse items from early serial positions as far into the study sequence as monolinguals.

Rehearsal spacing. Maximum spacing among repeated rehearsals was not as great in bilingual L2 as in monolingual rehearsal ($b = .653$, $SE = .241$, $t = 2.714$, $p = .008$) particularly for items from early serial positions ($b = -.185$, $SE = .048$, $t = 3.837$, $p < .001$). Maximal rehearsal spacing was only marginally less for bilingual L1 relative to monolingual rehearsal ($b = .423$,

SE = .230, t = 1.841, p = .068). There was an unexpected marginal three-way interaction of the monolingual versus bilingual L2 comparison with word frequency and the quadratic trend ($b = -.174$, SE = .096, t = 1.820, p = .069.), but no other effects involving group were significant (ps $> .1$). For L2 (but not L1) rehearsal sequences, this analysis confirms the prediction that spacing between repeated rehearsals of items from early serial positions would be narrower for bilingual than for monolingual rehearsal.

Rehearsal transitions. Monolinguals exhibited more *forward transitions* in rehearsal of items from early but not later serial positions relative to bilingual L2 ($b = -.134$, SE = .030, t = 4.411, $p < .001$) but not relative to bilingual L1 speakers ($p > .2$). This decrease in the difference between monolingual and bilingual L2 forward transitions across serial positions followed a curved function, and there was a curvature difference between monolingual and L1 forward transitions (L1: b = .081, SE = .030, t = 2.669, p = .008; L2: b = .157, SE = .030, t = 5.182, $p < .001$). Monolinguals made more *consecutive transitions* in rehearsal of items from early but not later serial positions relative to bilingual L2 speakers ($b = -.129$, SE = .023, t = 5.517, $p < .001$), and the curvature of the serial position functions differed (b = .113, SE = .023, t = 4.831, $p < .001$). An unexpected marginal three-way interaction suggested that the curvature difference depended on word frequency. No other effects involving group on rehearsal transitions were significant (ps $> .1$). For L2 (but not L1) sequences, these analyses confirm the prediction that items from early serial positions would have fewer forward and consecutive transitions in bilingual than in monolingual rehearsal.

Emergent rehearsal properties as a function of rehearsal rate. The preceding analyses were repeated with rehearsal rate added as a fixed factor. When rehearsal rate was controlled, group differences and their interactions with the linear trend on serial position did not change substantially (see Table 5). These results may indicate that unlike frequency and proficiency, the effects of bilingual status on rehearsal patterns are not mediated by rehearsal efficiency. However, it should be noted that the original effects of bilingual status were smaller than the frequency and proficiency effects.

Recall accuracy. Recall accuracy was analyzed using the same logistic mixed-effects regression models as for the rehearsal measures. Monolingual and bilingual groups did not differ in overall accuracy (zs < 1; see Table 6). However, bilingual L1 and L2 recall of items from earlier but not later serial positions was less accurate than monolingual recall (L1: $b = -.102$, SE = .031, z = 3.314, $p < .001$; L2: $b = -.133$, SE = .031, z = 4.278, $p < .001$). The curvature of the serial position function was less for bilingual L2 than for monolingual recall (b = .145, SE = .062, z = 2.339, p = .019). No other effects involving group reached significance (ps $> .20$). As in the comparison of Experiment 1 and 2 recall, bilingual recall rates for items from early serial positions were lower than monolingual recall rates.

Three analyses were performed to test hypotheses about how rehearsal characteristics impact recall accuracy and mediate the effects of bilingual status on recall accuracy. Controlling for the number or recency of rehearsals did not change patterns of group effects, which indicates that the effects of bilingual status on recall accuracy are not mediated by the number or recency of rehearsals. However, controlling the maximum spacing increased the monolingual advantage in recall accuracy.

Position of the first item recalled. PFR was analyzed as in the Experiment 1 and 2 comparisons. No effects involving group were indicated (zs < 1). When this analysis was performed using functional serial positions, there was again little evidence for group differences, only marginal effects suggesting later functional PFR for bilingual L1 than for monolingual recall ($b = -.528$, SE = .313, t = 1.685, p = .094) and a larger frequency effect for monolinguals than for bilingual L2 recall (b = .665, SE = .393, t = 1.692, p = .092). This analysis provides only weak evidence suggesting that in L1 (but not L2), bilinguals initiate recall with items from later functional serial positions than monolinguals.

Temporal association encoding. Temporal order measures (see Table 6) were analyzed using logistic mixed-effects models with group, word frequency, and linear and quadratic trends on serial position as fixed effects and random intercepts for participants and items.

Forward recall transitions. The negative linear trend showing decreased rates of forward transitions across serial positions was not as steep for bilingual L1 and L2 recall as for monolingual recall (L1: $b = -.200$, SE = .080, z = 2.512, p = .012; L2: $b = -.199$, SE = .080, z = 2.487, p = .013). The curvature of the serial position function was greater for monolinguals than for bilingual L1 recall ($b = -.193$, SE = .075, z = 2.589, p = .010) but the comparison to bilingual L2 recall was only marginal ($b = -.132$, SE = .075, z = 1.758, p = .079). No other effects were significant (ps $> .1$).

Consecutive recall transitions. Serial position functions for the proportion of consecutive transitions had less curvature for bilingual L2 speakers than for monolinguals ($b = -.146$, SE = .069; z = 2.118, p = .034). There was a suggestion that the word frequency effect was weaker for bilingual L2 speakers than for monolinguals (b = .233, SE = .134; z = 1.743, p = .081). No other effects involving group were significant (ps $> .1$).

Recall transitions as a function of rehearsal properties. The proportion of functionally consecutive transitions exhibited a negative linear trend showing a decrease in functionally consecutive transitions across serial positions. This trend was less steep for bilingual L1 and L2 recall than for monolingual recall, with bilinguals showing a lower rate of functionally consecutive transitions following words from earlier serial positions (L1: b = .182, SE = .070, z = 2.623, p = .009; L2: b = .192, SE = .069, z = 2.771, p = .006). The serial position function had greater curvature for monolinguals than for bilingual L2 speakers ($b = -.190$, SE = .066; z = 2.771, p = .006), but the effect was only marginal relative to bilingual L1 speakers ($b = -.110$, SE = .066, z = 1.670, p = .095). No other effects involving group were significant (ps $> .1$). This analysis shows that bilinguals made fewer functionally consecutive transitions after recalling items from early serial positions and that the functional measure is more sensitive to the effects of bilingual status than the nominal measure.

To examine the correspondence between forward and consecutive transitions from rehearsal to recall (see Figure 11), two analyses of recall transitions were conducted with the corresponding rehearsal transition measures, word frequency, and language group as fixed effects. As in the original analyses, there were no effects of group on transition rates in recall (zs < 1). The association between rehearsal and recall transitions in the forward direction were marginally weaker in bilingual L2 than in monolingual recall ($b = .194$, $SE = .101$, $z = 1.926$, $p = .054$). In the bilingual L1 versus monolingual comparison, this pattern was stronger for high-frequency words, as indicated by a significant three-way interaction ($b = -.541$, $SE = .226$, $z = 42.398$, $p = .017$). The association between rehearsal and recall transitions to items from consecutive serial positions was weaker for bilinguals than for monolinguals (L1: $b = -.204$, $SE = .076$, $z = 2.667$, $p = .008$; L2: $b = -.231$, $SE = .077$, $z = 3.019$, $p = .003$). These analyses show a weaker correspondence between rehearsal and recall patterns in bilinguals.

Interim Discussion of Experiments 3 and 4

In Experiments 3 and 4, conditions with less efficient rehearsal showed rehearsal patterns consistent with the predictions of the model. Specifically, with lower word frequency and lower language proficiency, items from early serial positions were rehearsed fewer times, not as far into the rehearsal sequence, with less spacing between repetitions, and with fewer forward and consecutive transitions. Consistent with these patterns, bilinguals rehearsed items from early serial positions fewer times, not as far into the study sequence, and with less spacing between repetitions than monolinguals. The effects of word frequency and language proficiency (but not the effects of bilingual status) on these rehearsal patterns were mediated by rehearsal rate.

In recall accuracy, the effects of word frequency, language proficiency, and bilingualism replicated those of Experiments 1 and 2. Low frequency, lower proficiency, and bilingual status were associated with later functional PFR. The hypothesis that less efficient rehearsal would reduce the strength or number of temporal associations formed at study was supported by the lower rates of functionally consecutive recall transitions with low word frequency, lower language proficiency, and bilingual status.

The contention that rehearsal patterns mediate the effects of word frequency and language proficiency on recall accuracy is also supported by the associations of several rehearsal characteristics with characteristics of recall performance. Items that were rehearsed more times, rehearsed later into the sequence (i.e., more recently), or rehearsed at more widely spaced intervals were more accurately recalled. Furthermore, putting each of these rehearsal variables into the recall accuracy model reduced, eliminated, or reversed the effects of word frequency and language proficiency (but not bilingual status) relative to the serial position analyses. The nominal and functional PFR analyses showed that the first item recalled was more likely to be one of the last five items rehearsed than one of the last five items presented, and the functional approach was more sensitive to the effects of word frequency and language proficiency. A similar analysis of consecutive recall transitions showed that transitions between items with consecutive rehearsals were more sensitive to these effects than transitions between items from consecutive serial positions.

General Discussion

We examined the effects of word frequency, language proficiency, and bilingual experience on the temporal dynamics of rehearsal and retrieval in free recall in monolingual and bilingual samples matched on age, education, and socioeconomic status. The approach was to measure the effects of these factors on rehearsal patterns, recall measures, and the relationships between them. In the sections to follow, we discuss the theoretical implications of the results for a modified variable-context model, the levels of representation involved in temporal encoding, bilingual associative memory, and qualitative differences among different types of associations.

Summary and Interpretation of Results Within a Modified Variable-Context Model

As explained in the introduction, our proposed refinement to the variable context and retrieved context models that considers rehearsals as episodes equivalent to additional presentations (as in Laming, 2009; Tan & Ward, 2000) fully accommodates the phenomena explained by the original models as well as a broader range of rehearsal and recall phenomena observed in the present results and in the literature. This approach, illustrated in Figure 1, leads to specific hypotheses about the impact of rehearsal efficiency on rehearsal characteristics, which would in turn have an impact on recall characteristics, illustrated in Figure 2. We tested several hypotheses derived from this revised model by indirectly manipulating the efficiency of rehearsal through experimental manipulation of word frequency and variation of participants' bilingual language proficiency levels and bilingual or monolingual status. The experimental outcomes are summarized in Table 7 and interpreted in the following paragraphs.

Rehearsal patterns. We tested whether systematic variation of word frequency, language proficiency, and bilingual status was a valid method to manipulate rehearsal efficiency and then tested hypotheses about the impact of rehearsal efficiency on emergent rehearsal patterns using these manipulations. Indeed, with low frequency and lower proficiency, fewer rehearsals intervened between successive item presentations (Figure 2a–b). Low-frequency words from early serial positions were rehearsed fewer times (Figure 2d) and not as far into the study sequence (Figure 2f) as high-frequency words (as in Tan & Ward, 2000; Ward et al., 2003), and these patterns replicated with lower proficiency and bilingualism. We examined three additional hypothesized consequences of less efficient rehearsal on rehearsal patterns that might impact the accuracy or sequence of later recall: spaced repetitions, forward transitions to items from later serial positions, and transitions to items from consecutive serial positions. Particularly for items from early serial positions, we showed that repeated rehearsals were less widely spaced (Figure 2e) and rehearsal sequences had fewer forward and consecutive transitions (Figure 2g–h) with low frequency, lower proficiency, and bilingualism. Rehearsal rate was a reliable predictor of all five rehearsal patterns (Figure 2d–h), and it mediated the effects of word frequency and language proficiency (but not bilingual status) on these patterns. These consistent findings across five rehearsal measures support the assumptions about the impact of rehearsal efficiency that were used to make predictions about free recall performance.

Recall accuracy across serial positions. We tested the hypothesis that under conditions eliciting less efficient rehearsal, recall of items from early serial positions would be less accurate. Recall of items from early but not late serial positions was indeed less accurate for low-frequency than for high-frequency word lists, as in previous research (e.g., Raymond, 1969; Sumby, 1963; Tan & Ward, 2000; Ward et al., 2003). Similarly, recall of items from early but not late serial positions was less accurate with lower

Table 7
Summary of the Effects of Word Frequency, Language Proficiency, and Bilingualism on Rehearsal and Free Recall Measures

Measure	Word frequency	Language proficiency	Monolingual versus bilingual L2[a]	Monolingual versus bilingual L1[a]
Rehearsal rate	LF < HF	LP < HP	B2 = M	B1 = M
Rehearsal of items from early positions				
Number of rehearsals	LF < HF	LP < HP	B2 < M	B1 < M
Last rehearsal set	LF < HF	LP < HP	B2 < M	B1 < M
Spacing of repetition in rehearsal	LF < HF	LP < HP	B2 < M	B1 < M
Forward transitions in rehearsal	LF < HF	LP < HP	B2 < M	B1 = M
Consecutive transitions in rehearsal	LF < HF	LP < HP	B2 < M	B1 = M
Recall of items from early serial positions				
Accuracy	LF < HF	LP < HP	B2 < M	B1 < M
Serial position of first item recalled[b]	LF > HF[4]	LP > HP[4]	B2 = M	B1 = M
Functional serial position of first item recalled	LF > HF	LP > HP	B2 = M	B1 = M
Forward transition rate in recall (early)	LF < HF[4]	LP < HP[4]	B2 < M[1v2]	B1 < M[1v2]
Consecutive transition rate in recall (early)	LF < HF[4]	LP < HP[4]	B2 < M[3v4]	B1 = M
Functionally consecutive transition rate (early)	LF < HF	LP < HP	B2 < M	B1 < M

Note. Superscripts indicate experiments where indicated effect is not reliable; 3v4 indicates the comparison of Experiments 3 and 4. LF = low frequency; HF = high frequency; LP = low proficiency; HP = high proficiency; B1 = bilingual L1; B2 = bilingual L2; M = monolingual.
[a] English data only. [b] Here, a low number means initiating recall with an item from an earlier serial position, which is associated with more efficient rehearsal.

proficiency (consistent with Yoo & Kaushanskaya, 2016) and for bilinguals relative to monolinguals. As explained in the introduction, previous studies of these effects were underpowered and sometimes had language-related confounds, leading to inconsistent results. The present study unconfounded nominal language and proficiency, matched groups closely, and had greater power to detect differences; because there were several study-test cycles in each condition, the sample size was much larger, and serial position was included in the design.

Recall accuracy as a function of rehearsal patterns. Having established the impact of rehearsal efficiency on both rehearsal patterns and recall accuracy, we conducted additional analyses to link the rehearsal patterns to recall accuracy (Figure 2i–k, 2n), using corresponding scores at the item level for every participant. This approach is more sensitive than examining the correlation of overall rehearsal and overall accuracy across participants, because in addition to providing 320 times as many observations, the associations are much stronger at the item level ($r = .51$) than at the participant level ($r = .13$).

The results support the contention that rehearsals are encoding events encoded in a manner similar to presentations. Specifically, as in previous research, recall accuracy was a monotonic function of the number of rehearsals and recency of rehearsal, with more rehearsals and more recent rehearsal associated with higher accuracy (Tan & Ward, 2000; Ward et al., 2003). Also, greater spacing between rehearsals was associated with more accurate recall, supporting the idea that contextual features associated with both presentations and rehearsals are cued by reinstated contexts of recalled items. Further supporting the conclusion that rehearsal properties mediate the effects of word frequency and language proficiency on recall accuracy, the effects of word frequency and language proficiency on recall accuracy were substantially reduced, eliminated. or even reversed when the number of rehearsals or rehearsal spacing was controlled.

The argument that these associations of rehearsal and recall measures arise simply because the same items are more easily retrieved for both rehearsal and free recall is countered by strong evidence that the serial position is a more important determinant of these rehearsal characteristics than inherent ease of retrieval (see Brodie & Murdock, 1977; Murdock & Metcalfe, 1978; Ward et al., 2003). Also, when spacing of rehearsals is manipulated rather than measured, it contributes to recall accuracy above and beyond the number of rehearsals and the recency of the last rehearsal (Tan & Ward, 2000). Consistent with this contention, the present data set shows a strong impact of serial position on the number of rehearsals, with assignment of items to serial positions counterbalanced.

Serial position of the first item recalled. The rate at which particular items were recalled first was a monotonic function of recency of rehearsal, supporting the status of rehearsals as encoding events similar to presentations. More recall sequences began with recently rehearsed items than with items from recent (late) serial positions, supporting the hypothesis that recency of rehearsal drives the selection of the first item recalled. Based on the idea (verified in Experiments 3 and 4) that rehearsal would not continue as far into the study sequence, we hypothesized that inefficient rehearsal would cause recall to begin with items from later nominal and functional serial positions. Indeed, these tests showed for the first time that low word frequency and lower language proficiency (but not bilingualism) were associated with later nominal and functional PFR. The novel analysis based on the *functional* serial position of the first item recalled was more sensitive than the nominal PFR to these factors, further supporting the status of rehearsals as encoding events and the claim that recency of rehearsal mediates these effects. With low frequency or lower proficiency, items from early serial positions were less likely to be rehearsed to the last few rehearsal cycles and less likely to be recalled first.

Recall trials with early PFR had stronger primacy effects, and trials with late PFR had stronger recency effects, consistent with previous research (see Figure 4; Deese & Kaufman, 1957; Kahana, 1996). This finding supports the idea that the rate of late PFR is an "excellent" indicator of the recency effect (Figure 2n; Kahana et al., 2002). However, even with early PFR some items from late positions were recalled, and with late PFR some items from early positions were recalled. This result is not easily accommodated under the original variable context and retrieved context models. However, under the revised model, reinstating the contexts of items from late serial positions would cue the retrieval of other items *presented or rehearsed* late in the sequence, include those from early serial positions. Conversely, when an early item is rehearsed far into the sequence, recalling it can cue the retrieval of items from later positions. These results show that when an item from an early rather than a late serial position is recalled first, recall of other items from early positions is more likely. This finding, along with the findings that recency of rehearsal is associated with both recall accuracy and the probability that an item will be recalled first, suggests that the probability of an item being recalled first mediates the relationship between its recency of rehearsal and recall accuracy (Figure 2k and 2n).

Transitions among items in recall output. The majority of recall transitions were among items rehearsed consecutively at study, supporting the hypothesis that the reinstated presentation and rehearsal contexts of recalled items at test would cue the retrieval of items presented or rehearsed consecutively at study. As explained above, less efficient rehearsal was associated with fewer forward and consecutive transitions in rehearsal, which we hypothesized would decrease the formation of interitem associations. Consistent with this hypothesis, forward and consecutive transition rates were lower for low-frequency than for high-frequency words (as in Ward et al., 2003). Similarly, lower language proficiency was associated with lower forward transition rates (as with order scores in Francis & Baca, 2014), and lower consecutive transition rates in Experiment 2, but in Experiment 4, transition effects were reliable only when considering *functionally* consecutive recall transitions. While bilingual recall showed some evidence for lower forward and consecutive transition rates relative to monolingual recall, the group differences were consistently reliable only for functionally consecutive transitions.

Across all key comparisons, measures based on consecutive rehearsals were more sensitive to differences in rehearsal efficiency than the measures based on consecutive serial positions. To better establish the link between rehearsal and recall transitions (Figure 2l–m) noted by Laming (2008, 2009), we examined the recall transitions conditionalized on rehearsal transitions (see Figure 11). Items that elicited higher numbers of forward or consecutive transitions in a rehearsal sequence more often elicited forward or consecutive transitions in recall, supporting the contention that rehearsals are encoding events similar to presentations and impact the recall sequence. This finding, along with elimination or reversal of frequency effects when the number of rehearsal transitions was controlled, supports the conclusion that rehearsal transition patterns mediate the effects of word frequency on recall transitions.

Related approaches. Rehearsal patterns include many transitions among items from adjacent serial positions, particularly in the forward direction (Experiments 3 and 4; Laming, 2008). Laming (2008, 2009) claimed that this rehearsal characteristic underlies the tendency to recall items from neighboring serial positions (Figure 2m), which he concludes is more accurately described as recalling items from neighboring rehearsals. Indeed, forward and consecutive transitions in rehearsal were predictive of forward and consecutive transitions, respectively, for the same items in recall output (see Figure 11). Also, effects associated with rehearsal efficiency were stronger when recall transitions were considered in terms of consecutive rehearsals rather than consecutive presentations (see Table 6). Laming (2008, 2009) further suggested that the graded and bidirectional lag-recency function is an artifact of the local reordering of items in rehearsal sequences and that only forward rehearsal transitions between adjacent items are diagnostic in predicting patterns of forward and consecutive transitions in the recall sequence. The present results are not inconsistent with this contention.

An alternative explanation for forward and consecutive transitions in recall output is that items presented in a study sequence are grouped into smaller clusters or chunks. In one version, items within clusters are strictly serially ordered, each cluster is bound to a temporal context shared by all component items, and retrieval occurs at the cluster level (Farrell, 2012). In another version, associations form among items within each cluster, with the first item being more strongly associated to the temporal context; with an all-or-nothing cluster retrieval process, retrieval of the first item cues recall of subsequent items in the same cluster (Lehman & Malmberg, 2013). If these approaches are applied to both presentations and rehearsals, then clusters would consist of items that are *presented or rehearsed* consecutively. This idea is supported by the finding that transitions in rehearsal were repeated in free recall output (Experiments 3 and 4; Laming, 2008, 2009). Forward, consecutive, and functionally consecutive transition rates in recall output were lower with low frequency and lower proficiency. These lower rates are consistent with the contention that when less efficient rehearsal leads to lower rates of forward and consecutive transitions in rehearsal (Experiments 3 and 4), the clusters formed are either less numerous or have weaker encoding.

Later instantiations of the retrieved context framework, including the *context maintenance and retrieval model* (CRM2, Lohnas et al., 2015; CRM, Polyn et al., 2009), accommodate other sorts of interitem associations, such as semantic relatedness and presentation source, and the occurrence of interlist intrusions to provide a more comprehensive account of the mechanisms underlying recall output sequences. Although semantic relationships and source information clearly influence the sequence of recall (e.g., Golomb et al., 2008; Lohnas et al., 2015; Polyn et al., 2009), further investigation of the role of semantic and source associations was beyond the scope of the present study.

An Associative Strength Theory of Bilingual Proficiency and Memory

In the present study, word frequency and language proficiency had parallel effects on rehearsal and recall patterns (as shown in Table 7). The nearly perfect alignment of word-frequency and language proficiency effects suggests that the bases of these two effects are one and the same, with the language manipulation perhaps being a somewhat weaker manipulation of experience than the word frequency manipulation. Because Experiments 3 and 4 show that both word frequency and language proficiency impact recall dynamics through rehearsal efficiency and rehearsal patterns, this conclusion is justified. Nevertheless, conceptualizations of language proficiency effects by analogy to word frequency effects are inadequate to explain a growing body of bilingual episodic memory research. Although language proficiency and word frequency effects exhibit striking parallels in recognition memory (Francis & Gutiérrez, 2012; Francis & Strobach, 2013) and repetition priming (Francis, Augustini, & Sáenz, 2003; Francis et al., 2008), they do not always follow the same patterns (e.g., Francis & Baca, 2014; Francis et al., 2019).

We propose an approach to bilingual memory that is similar to the *associative deficit hypothesis* of memory and aging (Naveh-Benjamin, 2000), in which older adults have difficulty learning new semantic, source-context, and temporal-context associations. In our *associative strength theory* (Francis et al., 2019), bilingualism and bilingual proficiency impact encoding or retrieval of associations, but the impact varies across different types of associations.

Learning of new semantic associations would not be expected to vary as a function of bilingual status or language proficiency, because semantic/conceptual representations are shared by translation equivalents in different languages (for reviews, see Francis, 1999, 2005), and long-standing semantic associations appear to be language general (for a review, see Francis, 2018). Indeed, after learning unrelated paired associates, cued recall accuracy did not differ across monolingual and bilingual samples (Papagno & Vallar, 1995), and associative recognition accuracy did not differ across language proficiency levels or across monolingual and bilingual samples (Francis, Lara, & Strobach, 2018).

The source contexts in which words are studied are conjectured to be associated at the conceptual level rather than word form level (Buchler, Faunce, Light, Gottfredson, & Reder, 2011; Diana & Reder, 2006). Indeed, the strengths of new source context associations do not appear to vary with language proficiency. Although bilingual source discrimination was more accurate for low-frequency than high-frequency words, language proficiency had no effect (Francis et al., 2019). The three experiments showed consistent results across perceptual (spatial location and color; auditory/visual presentation modality) and temporal (first or second list membership) source context manipulations. In the absence of proficiency effects, the additional finding that source memory was consistently more accurate for bilinguals than for monolinguals was counterintuitive. After ruling out proficiency and strategic factors, we concluded that bilinguals more efficiently link meaningless information like source context to word concepts.

In the present study, measures meant to capture temporal memory showed consistent effects. Lower proficiency was associated with lower rates of forward transitions (consistent with lower relative order scores in Francis & Baca, 2014) and consecutive transitions in recall output. Similarly, bilinguals generally (albeit not with consistent statistical significance) had lower rates of forward and consecutive transitions in recall output than monolinguals. We conclude that these effects arise because, like lower word frequency, lower language proficiency is associated with slower retrieval of word forms for production, which leads to less efficient rehearsal and its associated consequences for rehearsal and recall patterns.

Conclusions

Similar effects were observed for word frequency, language proficiency, and bilingualism across several measures of rehearsal and recall dynamics. With lower word frequency, lower language proficiency, and bilingual status, rehearsal exhibited patterns less conducive to accurate recall. Also, recall of words from early serial positions was less accurate, with recall sequences that indicate less effective encoding of temporal associations. Along with analyses that used rehearsal-based measures to link rehearsal and recall patterns for individual items, these results support four main conclusions. First, manipulations that decrease rehearsal efficiency impact both rehearsal and recall patterns. Second, rehearsal efficiency differences and their impact on rehearsal patterns mediate the effects of word frequency and language proficiency on recall performance. Third, the findings support modifications of the variable context model and retrieved context framework that treat presentations and rehearsals as equivalent encoding events. These modifications extend the explanatory power of the model mechanisms to explain primacy effects and several characteristics of recall output sequences. Finally, complex patterns of bilingual performance across different verbal associative memory tasks can be explained by a theory based on how well bilinguals learn different types of associations through their two languages.

Context Paragraph

This research project began as part of the first author's recent NIH-funded program of research on bilingual associative memory, which addresses how language proficiency in bilinguals impacts the formation of new semantic, contextual, and temporal associations and the strength of long-standing episodic and semantic associations. In reviewing the literature on associative memory, she became fascinated with theories behind the temporal dynamics of memory and possible links between rehearsal patterns at study and the retrieval patterns in free recall. By measuring both rehearsal and recall sequences, each participant's rehearsal history for each item could be linked to its later accuracy and output position in recall. This link made it possible to test whether word frequency and language proficiency effects on free recall are mediated by rehearsal properties. This project became more about those issues than the original focus on how and why language proficiency impacts memory. It occurred to the authors that in addition to allowing simultaneous generalization across participants and items, mixed-effects regression analyses with crossed random factors would make it possible to analyze these fine-grained links between rehearsal and recall patterns and incorporate continuous measures of participant language proficiency separately for each language based on the language of each trial.

References

Alameda, J. R., & Cuetos, F. (1995). *Diccionario de frecuencias de las unidades linguisticas del castellano* [Dictionary of frequencies of Spanish linguistic units]. Oviedo, Spain: Servicio de Publicaciones de la Universidad de Oviedo.

Alonso, M. A., Fernandez, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 Spanish words. *Behavior Research Methods, 47,* 268–274. http://dx.doi.org/10.3758/s13428-014-0454-2

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (CDROM).* Philadelphia, PA: Linguistic Data Consortium.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal* *of Memory and Language, 68,* 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6,* 173–189. http://dx.doi.org/10.1016/0010-0285(74)90009-7

Brodie, D. A., & Murdock, B. B., Jr. (1977). Effects of presentation time on nominal and functional serial position curves in free recall. *Journal of Verbal Learning and Verbal Behavior, 16,* 185–200. http://dx.doi.org/10.1016/S0022-5371(77)80046-7

Buchler, N. G., Faunce, P., Light, L. L., Gottfredson, N., & Reder, L. M. (2011). Effects of repetition on associative recognition in young and older adults: Item and associative strengthening. *Psychology and Aging, 26,* 111–126. http://dx.doi.org/10.1037/a0020816

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, & Computers, 25,* 257–271. http://dx.doi.org/10.3758/BF03204507

Deese, J. (1960). Frequency of usage and number of words in free recall: The role of association. *Psychological Reports, 7,* 337–344. http://dx.doi.org/10.2466/pr0.1960.7.2.337

Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology, 54,* 180–187. http://dx.doi.org/10.1037/h0040536

DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1136–1146. http://dx.doi.org/10.1037/0278-7393.22.5.1136

Diana, R. A., & Reder, L. M. (2006). The low-frequency encoding disadvantage: Word frequency affects processing demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 805–815. http://dx.doi.org/10.1037/0278-7393.32.4.805

Ebbinghaus, H. (1885). Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. Leipzig, Germany: Duncker & Humboldt. Reprinted as Ebbinghaus, H. E. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger, Trans.). New York, NY: Dover.

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review, 119,* 223–271. http://dx.doi.org/10.1037/a0027371

Fernandes, M. A., Craik, F., Bialystok, E., & Kreuger, S. (2007). Effects of bilingualism, aging, and semantic relatedness on memory under divided attention. *Canadian Journal of Experimental Psychology, 61,* 128–141. http://dx.doi.org/10.1037/cjep2007014

Francis, W. S. (1999). Cognitive integration of language and memory in bilinguals: Semantic representation. *Psychological Bulletin, 125,* 193–222. http://dx.doi.org/10.1037/0033-2909.125.2.193

Francis, W. S. (2005). Bilingual semantic and conceptual representation. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 251–267). New York, NY: Oxford University Press.

Francis, W. S. (2018). Shared core meanings and shared associations in bilingual semantic memory: Evidence from research on implicit memory. *The International Journal of Bilingualism*. Advance online publication. http://dx.doi.org/10.1177/1367006918814375

Francis, W. S., Augustini, B. K., & Sáenz, S. P. (2003). Repetition priming in picture naming and translation depends on shared processes and their difficulty: Evidence from Spanish-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1283–1297. http://dx.doi.org/10.1037/0278-7393.29.6.1283

Francis, W. S., & Baca, Y. (2014). Effects of language dominance on item and order memory in free recall, serial recall and order reconstruction. *Memory, 22,* 1060–1069. http://dx.doi.org/10.1080/09658211.2013.866253

Francis, W. S., Corral, N. I., Jones, M. L., & Sáenz, S. P. (2008). Decomposition of repetition priming components in picture naming. *Journal of Experimental Psychology: General, 137,* 566–590. http://dx.doi.org/10.1037/0096-3445.137.3.566

Francis, W. S., & Gutiérrez, M. (2012). Bilingual recognition memory: Stronger performance but weaker levels-of-processing effects in the less fluent language. *Memory & Cognition, 40,* 496–503. http://dx.doi.org/10.3758/s13421-011-0163-3

Francis, W. S., Lara, B., & Strobach, E. N. (2018). *Effects of bilingual proficiency on the learning of new semantic associations: New evidence for global matching in associative recognition*. Manuscript submitted for publication.

Francis, W. S., & Strobach, E. N. (2013). The bilingual L2 advantage in recognition memory. *Psychonomic Bulletin & Review, 20,* 1296–1303. http://dx.doi.org/10.3758/s13423-013-0427-y

Francis, W. S., Strobach, E. N., Penalver, R. M., Martínez, M., Gurrola, B. V., & Soltero, A. (2019). Word-context associations in episodic memory are learned at the conceptual level: Word frequency, bilingual proficiency, and bilingual status effects on source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45,* 1852–1871. http://dx.doi.org/10.1037/xlm0000678

Francis, W. S., Taylor, R. S., Gutiérrez, M., Liaño, M. K., Manzanera, D. G., & Penalver, R. M. (2018). The effects of language proficiency on item recall and semantic clustering in free recall output: Evidence for shared semantic associations across languages. *Memory, 26,* 1364–1378. http://dx.doi.org/10.1080/09658211.2018.1476551

Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., & Davey Smith, G. (2006). Indicators of socioeconomic position (Part 1). *Journal of Epidemiology and Community Health, 60,* 7–12. http://dx.doi.org/10.1136/jech.2004.023531

Glenberg, A. M. (1984). A retrieval account of the long-term modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 16–31. http://dx.doi.org/10.1037/0278-7393.10.1.16

Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 3–15. http://dx.doi.org/10.1037/0278-7393.12.1.3

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always a means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58,* 787–814. http://dx.doi.org/10.1016/j.jml.2007.07.001

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition, 33,* 1220–1234. http://dx.doi.org/10.3758/BF03193224

Golomb, J. D., Peelle, J. E., Addis, K. M., Kahana, M. J., & Wingfield, A. (2008). Effects of adult aging on utilization of temporal and semantic associations during free and serial recall. *Memory & Cognition, 36,* 947–956. http://dx.doi.org/10.3758/MC.36.5.947

Gregg, V., Montgomery, D., & Castano, D. (1980). Recall of common and uncommon words from single and mixed lists. *Journal of Verbal Learning and Verbal Behavior, 19,* 138–140. http://dx.doi.org/10.1016/S0022-5371(80)90202-9

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language, 38,* 313–338. http://dx.doi.org/10.1006/jmla.1997.2547

Haritos, C. (2002). A developmental examination of memory strategies in bilingual six, eight, and ten year olds. *International Journal of Bilingual Education and Bilingualism, 5,* 197–220. http://dx.doi.org/10.1080/13670050208667756

Harris, J. G., Cullum, C. M., & Puente, A. E. (1995). Effects of bilingualism on verbal learning and memory in Hispanic adults. *Journal of the International Neuropsychological Society, 1,* 10–16. http://dx.doi.org/10.1017/S1355617700000059

Hogan, R. M. (1975). Interitem encoding and directed search in free recall. *Memory & Cognition, 3,* 197–209. http://dx.doi.org/10.3758/BF03212898

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 923–941. http://dx.doi.org/10.1037/0278-7393.25.4.923

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299. http://dx.doi.org/10.1006/jmps.2001.1388

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica, 127,* 277–288. http://dx.doi.org/10.1016/j.actpsy.2007.06.003

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition, 24,* 103–109. http://dx.doi.org/10.3758/BF03197276

Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 530–540. http://dx.doi.org/10.1037/0278-7393.28.3.530

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44,* 978–990. http://dx.doi.org/10.3758/s13428-012-0210-4

Laming, D. (2008). An improved algorithm for predicting free recalls. *Cognitive Psychology, 57,* 179–219. http://dx.doi.org/10.1016/j.cogpsych.2008.01.001

Laming, D. (2009). Failure to recall. *Psychological Review, 116,* 157–186. http://dx.doi.org/10.1037/a0014150

Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review, 120,* 155–189. http://dx.doi.org/10.1037/a0030851

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual variability in free recall. *Journal of Memory and Language, 64,* 249–255. http://dx.doi.org/10.1016/j.jml.2010.11.003

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review, 122,* 337–363. http://dx.doi.org/10.1037/a0039036

Mensink, G.-J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review, 95,* 434–455. http://dx.doi.org/10.1037/0033-295X.95.4.434

Murdock, B. B., Jr., & Metcalfe, J. (1978). Controlled rehearsal in single trial free recall. *Journal of Verbal Learning and Verbal Behavior, 17,* 309–324. http://dx.doi.org/10.1016/S0022-5371(78)90201-3

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1170–1187. http://dx.doi.org/10.1037/0278-7393.26.5.1170

Nott, C. R., & Lambert, W. E. (1968). Free recall of bilinguals. *Journal of Verbal Learning and Verbal Behavior, 7,* 1065–1071. http://dx.doi.org/10.1016/S0022-5371(68)80069-6

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology, 48,* 98–107. http://dx.doi.org/10.1080/14640749508401378

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116,* 129–156. http://dx.doi.org/10.1037/a0014420

Postman, L. (1970). Effects of word frequency on acquisition and retention under conditions of free-recall learning. *The Quarterly Journal of Experimental Psychology, 22,* 185–195. http://dx.doi.org/10.1080/00335557043000113

Potter, M. C., So, K. F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23,* 23–38. http://dx.doi.org/10.1016/S0022-5371(84)90489-4

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134. http://dx.doi.org/10.1037/0033-295X.88.2.93

Raymond, B. (1969). Short-term and long-term storage in free recall. *Journal of Verbal Learning and Verbal Behavior, 8,* 567–574. http://dx.doi.org/10.1016/S0022-5371(69)80106-4

Richardson, J. T. E., & Baddeley, A. D. (1975). The effect of articulatory suppression in free recall. *Journal of Verbal Learning and Verbal Behavior, 14,* 623–629. http://dx.doi.org/10.1016/S0022-5371(75)80049-1

Ross, B., & Landauer, T. (1978). Memory for at least one of two items: Test and failure of several theories of spacing effects. *Journal of Verbal Learning and Verbal Behavior, 17,* 669–680. http://dx.doi.org/10.1016/S0022-5371(78)90403-6

Rundus, D. (1971). An analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89,* 63–77. http://dx.doi.org/10.1037/h0031185

Rundus, D., & Atkinson, R. C. (1970). Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior, 9,* 99–105. http://dx.doi.org/10.1016/S0022-5371(70)80015-9

Sholl, A., Sankaranarayanan, A., & Kroll, J. F. (1995). Transfer between picture naming and translation: A test of asymmetries in bilingual memory. *Psychological Science, 6,* 45–49. http://dx.doi.org/10.1111/j.1467-9280.1995.tb00303.x

Sumby, W. (1963). Word frequency and serial position effects. *Journal of Verbal Learning and Verbal Behavior, 1,* 443–450. http://dx.doi.org/10.1016/S0022-5371(63)80030-4

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1589–1625. http://dx.doi.org/10.1037/0278-7393.26.6.1589

Ward, G., Tan, L., & Grenfell-Essam, R. (2010). Examining the relationship between free recall and immediate serial recall: The effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1207–1241. http://dx.doi.org/10.1037/a0020122

Ward, G., Woodward, G., Stevens, A., & Stinson, C. (2003). Using overt rehearsals to explain word frequency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 186–210. http://dx.doi.org/10.1037/0278-7393.29.2.186

Wheeldon, L. R., & Monsell, S. (1992). The locus of repetition priming of spoken word production. *Quarterly Journal of Experimental Psychology, 44,* 723–761. http://dx.doi.org/10.1080/14640749208401307

Woodcock, R. W., Muñoz-Sandoval, A. F., Ruef, M. L., & Alvarado, C. G. (2005). *Woodcock-Muñoz Language Survey-Revised*. Rolling Meadows, IL: Riverside Publishing.

Yoo, J., & Kaushanskaya, M. (2016). Serial-position effects on a free-recall task in bilinguals. *Memory, 24,* 409–422. http://dx.doi.org/10.1080/09658211.2015.1013557

Received January 11, 2018
Revision received October 13, 2019
Accepted November 21, 2019 ■