Applied Psycholinguistics **25** (2004), 293–321
Printed in the United States of America
DOI: 10.1017.S0142716404001146

Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition

GIOVANNA SPECIALE, NICK C. ELLIS, and TRACEY BYWATER
University of Wales, Bangor

ADDRESS FOR CORRESPONDENCE
Nick C. Ellis, School of Psychology, University of Wales, Bangor, Gwynedd LL57 2AS, United Kingdom. E-mail: n.ellis@bangor.ac.uk

ABSTRACT
Two studies examined individual cognitive differences affecting the acquisition of second language word forms. Experiment 1 measured 40 undergraduates' ability to learn phonological sequences, their phonological short-term store capacity as indexed by ability to repeat nonwordlike nonwords, and their learning of novel foreign language vocabulary (German) in an experimental task. Phonological sequence learning predicted receptive vocabulary learning. Phonological sequence learning and phonological store capacity made independent additive contributions to productive vocabulary learning. Experiment 2 determined the interactions of phonological sequence learning ability, phonological store capacity, and second language acquisition during a longitudinal field study of 44 novice undergraduate learners of Spanish during a 10-week course. Students' initial skill in phonological sequence learning predicted their final levels of Spanish receptive language and their eventual ability to repeat Spanish-wordlike nonwords. The results suggest that phonological store capacity and phonological sequence learning ability are initially separable constraints on second language vocabulary acquisition and that sequence learning ability underpins the acquisition of long-term phonological knowledge. Subsequent apprehension and consolidation of a novel word form is a product not only of phonological short-term store capacity but also of this long-term knowledge of the phonological regularities of language.

Some students are better than others at learning second language (L2) vocabulary. The reviews of Nation (2001), Schmitt and McCarthy (1997), and Laufer and Hulstijn (2001) illustrate that a considerable amount of recent research activity has been directed toward understanding individual differences in this ability. Yet the majority of this research has concentrated on the learning of L2 word meaning. It is clear that learning the forms of words, their input and output representations, involves processes that are quite separable from those involved in learning word meanings (Ellis, 1994a). This paper therefore focuses on individual differences in the processes involved in the acquisition of L2 word forms and the ability to recognize and produce novel phonological strings. Two strands of work within cognitive psychology served as our starting points. The first was the model of working memory developed by Baddeley and an associate (Baddeley & Hitch, 1974; Gathercole & Baddeley, 1993). The second was recent research on implicit induction of phonological sequences (Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Newport, Aslin, & Tunick, 1997).

The *working memory model* (Baddeley & Hitch, 1974) posits a phonological short-term memory system that is responsible for the temporary storage and manipulation of speech. It is composed of two subsystems: the phonological store, which holds phonological input for approximately 2 s, and the phonological loop, which can refresh the contents of the store by a process of subvocal articulation. A formidable body of evidence suggests that individual differences in the capacity of phonological short-term memory constrains L2 vocabulary development. Nonword repetition ability has been shown to predict ability to learn L2 (English) vocabulary in Finnish children (Service, 1992) and in Hong Kong seventh graders (Cheung, 1996). There is a relationship between nonword repetition and native language receptive vocabulary knowledge even in very young children (Gathercole & Adams, 1993, 1994). The ability to repeat nonwords is related specifically to the learning of novel phonological forms; it is not related to the learning of novel word meanings (Gathercole & Baddeley, 1990). Preventing use of the phonological loop by interference from a secondary task of articulatory suppression dimininshes the ability to learn novel foreign language words (Ellis & Sinclair, 1996; Papagno, Valentine, & Baddeley, 1991), and natural talent for learning foreign language is often accompanied by an enhanced phonological short-term memory capacity (Papagno & Vallar, 1995). In sum, as reviewed by Baddeley, Gathercole, and Papagno (1998) and Ellis (1996), a variety of evidence suggests that phonological short-term memory plays an important role in the acquisition of L2 word forms.

The original working memory model argued that the greater the quantity of information that could be held in the store and refreshed via the loop, the more complete and precise would be the short-term and long-term representations of a novel phonological form. Thus, it assumed that a novice learner comes to the task of L2 acquisition with a particular phonological short-term memory capacity, and it is this that limits the rate of L2 acquisition.

However, a further constraint on L2 vocabulary acquisition may lie in the ability to learn the phonological regularities of a language per se. Various sources suggest that phonological sequence learning, and the long-term knowledge of phonological regularities of a language that results, contribute to at least three facets of lexical development: the segmentation of speech into discrete word units, identification of the lexical units of language, and the development of automaticity in their processing.

Ability to segment a continuous stream of speech into discrete word units is an important precursor to vocabulary acquisition. There are no silences between words in continuous speech that serve to separate lexical tokens as do spaces on the printed page. Instead, speech must be segmented, by infants and adults alike, on the basis of knowledge of the phonological regularities of a language (Brent, 1999; Elman, 1990; Jusczyk, 1997; Saffran, et al., 1996; Saffran, Newport, & Aslin, 1996; Saffran et al., 1997). Infants manage this learning on the basis of statistical analysis of phonotactic sequences to determine "troughs" in the pattern of transitional probabilities between adjacent syllables (Gómez & Gerken, 2000). Such findings put this phenomenon firmly within the broad experimental area of sequence learning and theories of implicit learning (Ellis, 1994b; Reber, 1993).

Adults are sensitive to the sequential dependencies of their language in a similar way. Phonotactic competence emerges from using language, from the primary linguistic data of the lexical patterns that a speaker knows. Frisch, Large, Zawaydeh, and Pisoni (2001) asked native speakers to judge, using a 7-point rating scale, nonword stimuli for whether they were more or less like English words. The nonwords were created with relatively high or low probability legal phonotactic patterns. These expected probabilities were strongly related to the mean wordlikeness judgements for the nonword stimuli ($r = .87$). Frisch et al. (2001) argue that any new nonword is compared to the exemplars that are in the speaker's mental dictionary: the closer it matches their characteristics, the more wordlike it is judged to be.

The development of automaticity of word recognition and production is also an important facet of L2 lexical development because this determines the extent to which attentional resources can be freed to semantic and syntactic processing. Ellis (1996) argues that the development of automaticity of auditory word recognition is the result of learning sequential regularities in a language: speech sounds that frequently co-occur are further chunked into ever larger linguistic units in a hierarchical fashion so that the more frequently a sequence of phonemes/syllables is heard, the faster it is processed.

If the learning of phonological regularities does indeed affect lexical development in these ways, it follows that individual differences in the ability to learn phonological sequences might play a part in determining their aptitude in vocabulary acquisition. However plausible this potential relationship, there is to our knowledge no direct empirical evidence to support it, and the evaluation of this proposal thus constitutes the major goal of this paper.

This goal, however, is complicated by the fact that in the course of development these two potential limits on vocabulary acquisition (phonological short term memory and knowledge of phonological regularities) lose their independence and enter into interaction. The evidence for this is that knowledge of phonological regularities influences retention of material in short-term memory. Three types of effect are relevant. The first concerns the "wordlikeness" effect whereby phonotactically regular nonwords are easier to repeat than are nonwords or foreign language words that do not resemble native language words (Ellis & Beaton, 1993b; Gathercole, 1995a, 1995b; Gathercole, Willis, Emslie, & Baddeley, 1991; Hulme, Maughan, & Brown, 1991; Service, 1992; Service & Craik, 1993; Treiman & Danis, 1988). The second concerns lexicality effects: short-term memory span for words is better than for nonwords and it has been shown that this effect is related to the phonological properties of words rather than their semantic properties (Hulme et al., 1991; Hulme, Roodenrys, Brown, & Mercer, 1995). The third concerns differential short-term memorability within the population of words: Words of similar rhyme predominate in the constituency of phonological neighborhoods in English monosyllables (De Cara & Goswami, 2002), and children remember words from dense phonological neighborhoods better than those from sparse neighborhoods in short-term memory tasks (Thomson, Goswami, & Hazan, 2003). These three phenomena suggest that short-term memory performance is enhanced by wordlikeness and that this effect is dependent on long-term phonological knowledge which is derived from prior lexical experience (Bailey & Hahn, 2001; Metsala & Walley, 1998).

The original working memory model could not explain these effects: it simply argued that short-term retention is a function of short-term store capacity and made no reference to learners' long-term knowledge of language. Hence, Gathercole further developed the model, proposing two possible explanations for the wordlikeness effect. First, facilitation for wordlike nonwords is due to postlexical redintegrative processes whereby decayed partial representations of items in phonological short-term memory are filled in by analogy to similar existing long-term lexical or sublexical representations (Gathercole, 1995a; Gathercole & Adams, 1994; Gathercole et al., 1991). Second, the wordlikeness effect is due to prelexical perceptual processes with the perceptual system being tuned to more easily process nonwords that more closely resemble native language words (Gathercole, 1995a; Gathercole & Martin, 1996). Both of these forms of facilitation critically depend on knowledge of which phonemes or syllables commonly co-occur in a language, and it follows that knowledge of the regularities of a language may be important in moderating the limits of short-term store capacity in vocabulary learning.

So what is it that limits learners' vocabulary development? Is it the capacity of their phonological store; or is it the ability to learn the phonological sequences of language, both phoneme sequence information specific to particular words and the summed experience of phonological sequences, that collates into general knowledge of the phonotactic regularities of their languages? Alternately, are these two capacities inextricably entwined through experience?

The following two experiments were designed to throw light on these options. We measured two sources of individual difference, the ability to learn phonological regularities and, separately, the capacity of the phonological short-term store. We also measured L2 vocabulary learning. We then determined the extent to which these two sources of individual difference are independently related to the ability to learn L2 vocabulary, as well as the extent to which they are interrelated.

The capacity of the store and the ability to learn phonological regularities are generalized determinants of language learning ability; they should therefore contribute to the ability to learn both L2 and native language. In order to assess this subsidiary hypothesis, a test measuring the breadth of native language vocabulary was also included in Experiment 1.

The first experiment used two rather different vocabulary tasks: one emphasizing L2 vocabulary *learning* by having participants learn German foreign language vocabulary from computerized presentation, the other tapping the breadth of native language lexical *knowledge* by means of a lexical decision task involving very low frequency English words. The second experiment sought greater ecological validity by using a longitudinal field experiment in which we assess learners' acquisition of Spanish vocabulary during a 10-week novice-level course. In both of these experiments, a measure known as the Phonological Sequencing Index (PSI) was used to determine individual differences in the ability to learn phonological sequences. This task used a continuous recognition paradigm and measured the extent to which participants were able to distinguish between sequences of phonemes/syllables that occur only once and sequences that recur: how many exposures were necessary before an individual participant was sensitive to recurrence? Both of these experiments also involved nonword repetition tasks as an index of phonological short-term memory capacity. Experiments 1 and 2 both used nonwords guaranteed to be low in wordlikeness by using a speech synthesis to produce a relatively "pure" measure of the capacity of the store, unsullied by long-term phonological knowledge (Gathercole, 1995a). Experiment 2 added a measure of the ability to repeat nonwords that resemble Spanish words, a measure that taps both the capacity of the short-term store and long-term knowledge of the regularities of Spanish.

EXPERIMENT 1

Experiment 1 separately measured (a) participants' nonword repetition ability using speech-synthesized materials that are low in wordlikeness in order to provide a relatively pure measure of phonological store capacity and (b) their phonological sequence learning ability. Nonwords were constructed by randomly combining consonant–vowel (CV) syllables to form nonwords of various lengths and then producing these by means of computer speech synthesis. This was done in order to ensure that nonwords were both pronounceable and relatively nonwordlike. This experiment determined the degree to which these skills are related, and it then assessed their relative contributions in predicting L2 German vocabulary learning and native English vocabulary knowledge.

Method

Participants. Participants were 38 undergraduate students of psychology at the University of Wales, Bangor. They chose to partake in the experiment to fulfill course requirements. Their ages were between 18 and 45. None of the participants had studied German or professed to have any knowledge of German.

Design. A within subjects design was used. The two predictor variables were nonword repetition ability and phonological sequence learning. The three dependent variables were rate of learning German vocabulary for productive and receptive testing and accuracy on an English lexical decision task involving very low frequency words. A balanced Latin square design was used to determine the order of presentation of these tasks.

Apparatus. All testing was done individually using an Apple Macintosh computer with a 15-in. monitor programmed in HyperCard. The Apple Macintosh speech synthesis facility (Speech Text) was used in some of the tasks. The voice selected to speak the stimuli was "Junior," a synthesized voice intended to resemble the voice of a young American male. The speech rate for all experiments was 80 syllables/min. All auditory stimuli were presented using the external speakers on the computer. A Sony Dictaphone with an Altai microphone was used to record all verbal responses. Participants were tested individually and the experiment took approximately 1 hr.

Stimuli and procedure.

NONWORD REPETITION. The stimuli were the 32 nonwords listed in Appendix A, which were spoken using computer synthesized speech. Nonwords were composed of sequences of CV syllables constructed by randomly combining one of 12 consonants (g v ɾ f s p d m n k b l) with one of four vowels (ɔ o i ʊ). They varied from one to eight syllables in length, with four examples at each syllable length. These stimuli were spoken without prosody: there was no lexical stress or pitch change. Phonemes were not lengthened beyond the default length used by the speech synthesis program. Stimuli were presented at a rate of 80 syllables/min. A pretest established that these nonwords did not resemble English words.

Participants were instructed to listen to each nonword and, once it had finished to repeat it as accurately as possible into the microphone. Stimuli were presented in an ascending order of syllabic length. A dictaphone was used to record all responses.

PHONOLOGICAL SEQUENCING LEARNING. The stimuli (listed in Appendix B) were nonwords (12 targets and 96 foils) varying in length from two to four syllables. Target items were presented eight times during the task and foils only once. Stimuli were constructed using a HyperCard script in which consonants and vowels were randomly combined to produce CV syllables that were themselves randomly combined to produce nonwords. All stimuli were spoken using the Macintosh speech synthesis facility.

On each trial a nonword was presented. The first 24 trials comprised 12 target items and 12 foil items; these were presented in a fixed quasirandom order. The following 168 trials were presented in random order: on half of these trials, one of the 12 target items was presented and on the other half, a foil item was presented. Each target item was thus presented on eight trials, while foil items were presented just once. Participants indicated whether they had heard each stimulus previously by a mouse click to one of two buttons on the screen labeled "heard before" and "new sound."

Participants who rapidly acquired the phonological sequences of the target items were therefore able to discriminate between old and new items earlier on in the learning sequence.

SECOND LANGUAGE VOCABULARY ACQUISITION: THE GERMAN LEARNING TASK. The stimuli were the 24 German words and their English translation equivalents listed in Appendix C. These were divided into two sets of 12 word pairs. In each set there were six verbs (in the infinitive form) and six nouns. This task used digitized recordings of the German words spoken by a female native German speaker. Recordings were made using MacRecorder sampling in 8 bits at a rate of 22 kHz. These stimuli and procedures were used in a similar fashion to that of Ellis and Beaton (1993a).

The task consisted of three phases: the learning phase, the receptive test, and the productive test. On each trial of the learning phase, an English test word was visually presented in the center of the screen and at the same time a digitized recording of the paired German translation equivalent was spoken twice. When all of the 12 English words and their German translation equivalents had been presented in this way in random order, the procedure was repeated a second time. Thus participants heard all target German words four times. Participants were instructed that they should try to learn these words and their meanings.

For the receptive task, digitized recordings of the German test words were presented individually in random order. Following the presentation of each German word, participants were instructed to produce the English translation equivalent. Responses were typed into a dialogue box in the center of the screen. If participants were unable to answer within 30 s, they were verbally instructed to move on to the next item.

For the productive task, written English translation equivalents of German test words were presented individually, in random order, in the center of the screen. Following the presentation of each English word, participants were prompted to verbally produce the paired German translation equivalent; they were given as much time as needed to do this. All responses were recorded on audiotape.

NATIVE VOCABULARY KNOWLEDGE: THE ENGLISH LEXICAL DECISION TASK. Test items were the 80 low frequency English words and 80 nonwords listed in Appendix D. Low frequency words had a Thorndike–Lorge frequency value of less than 1 per million (Thorndike & Lorge, 1944).

Stimuli were presented on the Macintosh monitor individually in random order in a dialogue box at the top of the screen. Participants indicated whether they believed each item to be a word or a nonword by clicking on an appropriately labeled button on the screen.

Scoring.

NONWORD REPETITION TASK. Two measures of nonword repetition accuracy were computed from the audiorecorded data. The first of these was the number of nonwords that were correctly repeated. A repetition attempt was scored as correct only if all syllables were correctly repeated in the correct serial order. Because there were 32 nonwords, scores on this measure varied from 1 to 32. This is the scoring method favored by Gathercole and colleagues (e.g., Gathercole & Adams, 1993, 1994; Gathercole, Hitch, Service, & Martin, 1997; Gathercole, Willis, Emslie, & Baddeley, 1992). This measure will be known as "Nonword Repetition (words)."

The second of these measures of nonword repetition accuracy was the number of syllables correctly repeated. Service (1992) used this kind of measure. A syllable was scored as correct only if both the onset and rime portion of the syllable was correctly repeated in the correct serial position within the nonword. In order to assess serial position in terms of both initial position and final position, the nonword was divided into two halves. For the first half of the stimulus (initial syllables), serial position was adjudged in terms of left to right serial position. For the syllables in the second half of the stimulus (final syllables), serial position was adjudged in terms of right to left serial position. This system was used in order to give credit to those participants who correctly repeated final syllables in syllable final position but added syllables in the middle of the nonword. If serial position was merely adjudged from right to left, no such credit could be given. This measure will be known as "Nonword Repetition (syllables)."

Occasionally, there was a disparity between the consonant the speech synthesizer was programmed to produce and that commonly perceived by participants. For example, at the end of vowels, particularly those preceded by velar plosives (k and g), the speech synthesizer produced sounds that more resembled alveolar plosives. Participants were not penalized for incorrect repetition of perceptually ambiguous phonemes.

A random sample of 10% of the audiorecordings was assessed and scored by a second rater. The interrater reliability, according to a measure of Cronbach's alpha, was .99. A regression analysis was also conducted to measure interrater reliability. This indicated that 97% of the variance in the scores from the first rater can be accounted for by variance in the scores from the second rater ($r = .98$, $r^2 = .97$, $p < .01$).

PHONOLOGICAL LEARNING: THE PSI. A formula derived from signal detection theory was used to measure the degree to which participants were able to identify target words; this measure was designed to assess the contribution of guessing and subtract this from the final score (Anderson & Freebody, 1983; Meara & Buxton, 1987):

$$P(k) = \frac{P(\text{h}) - P(\text{fa})}{1 - P(\text{fa})}.$$

A hit is a correct "heard before" response. Here, $P(\text{h})$ represents the probability of making a "hit," and it is calculated by measuring the proportion of target item trials on which the participant made the correct "heard before" response. A false alarm is made on trials in which the "heard before" response is not correct, and $P(\text{fa})$ represents the probability of making a false alarm. It is calculated by measuring the proportion of foil trials in which the participant made an incorrect "heard before" response. Here, $P(\text{fa})$ reflects the contribution of guessing: the more false alarms the participant makes, the more likely it is that guessing will have played a role in correct identifications. Both $P(\text{fa})$ and $P(\text{h})$

Table 1. *Descriptive statistics for all tasks in Experiment 1*

Variable	*M*	*SD*
PSI *P*(*k*)	.66	.17
Nonword (words)/32	10.71	3.30
Nonword (syllables)/144	64.32	15.68
German receptive task/24	13.05	3.81
German productive task/96	43.92	17.63
English lexical decision *P*(*k*)	.47	.18

Note: PSI, Phonological Sequencing Index; $n = 38$.

were calculated on the basis of 168 trials. Thus, $P(k)$ measures the accuracy minus guessing.

THE GERMAN LEARNING TASK. The German learning task had two outcome measures that separately indexed receptive and productive learning. An item was scored as receptively correct if the correct English translation was produced; phonologically feasible misspellings of the correct translation equivalent were also considered correct but synonyms of the correct translation equivalent were not. Scores were summed to a total score.

Scoring the German productive task was a more complex matter because we wished to give credit for partially correct responses. A score of 4 was given to a totally correct response; with points being lost for each incorrect element within the response such that a score of zero represented a totally incorrect response. Two different types of error were itemized: omitting a phoneme that was present in the target German word and including a phoneme that was not present in the target. Correct responses were typically far from perfect productions of the German word; however, this license to participants was necessary because not all phonemes present in German are represented within the English phonological system and these were ab initio learners. Because this scoring system involved an element of subjectivity, a random sample of 10% of the audiorecordings was assessed and scored by a second rater. The interrater reliability, according to a measure of Cronbach's alpha, was .99.

THE ENGLISH LEXICAL DECISION TASK. The ability of participants to accurately make a lexical decision was measured using the $P(k)$ measure described above. A hit response was a correct "word" response and a false alarm was an incorrect "word" response.

Results

Table 1 shows performance levels for each task. Individual differences in PSI $P(k)$ were not significantly related to those in nonword repetition (words, $r = .13$; syllables, $r = .21$). German receptive acquisition correlated with German productive acquisition ($r = .86$, $p < .01$).

Table 2. *Pearson and partial correlations relating PSI performance, Nonword Repetition, and measures of lexical competence*

	German Receptive	German Productive	English Receptive
Pearson's correlation coefficients			
PSI $P(k)$	.45**	.46**	.04
Partial correlation coefficients controlling for Nonword Repetition			
PSI $P(k)$	.36*	.39*	.00
Pearson's correlation coefficients			
Nonword Repetition (words)	.32	.36*	.15
Nonword Repetition (syllables)	.29	.45**	.38*
Partial correlation coefficients controlling for PSI $P(k)$			
Nonword Repetition (words)	.29	.33*	.14
Nonword Repetition (syllables)	.22	.40**	.38*

Note: PSI, Phonological Sequencing Index; $df = 35$.
*$p < .05$. **$p < .01$.

Pearson correlations were performed to investigate the associations between these predictors and the lexical outcome variables. The coefficients are shown in Table 2. There were significant positive correlations between PSI $P(k)$ scores and both German receptive acquisition ($r = .45$, $p < .01$) and German productive acquisition ($r = .46$, $p < .01$). Partial correlations were also performed controlling for Nonword Repetition. The relationships between PSI $P(k)$ and performance in the German productive task and the German receptive task remained significant when performance on the nonword repetition measures were partialed out. There was no significant correlation between performance on the PSI and performance on the English lexical decision task.

Pearson's correlations revealed that Nonword Repetition (syllables) was significantly related to performance in the German productive task ($r = .45$, $p < .01$) but not to performance in the German receptive task ($r = .29$, *ns*). The difference between these r values was significant ($t = 2.042$, $p < .05$). Partial correlations were conducted controlling for performance on the PSI. Under these conditions the relationship between nonword repetition and the German productive task remained significant. Nonword Repetition (syllables) was significantly related to English lexical decision ($r = .38$, $p = .05$) and this relationship remained significant when PSI $P(k)$ was partialed out.

A series of stepwise multiple linear regression analyses were conducted in which PSI $P(k)$ and Nonword Repetition were entered into the model in different orders to ascertain whether they explain unique variance in each dependent variable (German receptive and productive vocabulary learning and English vocabulary knowledge). The reader should bear in mind when considering these results that, due to low degrees of freedom ($df = 38$), these analyses lack statistical power and thus some effects may be missed as a result.

In the first of these sets of multiple regression analyses, German receptive learning served as the dependent variable. The PSI $P(k)$ was entered in the first step and Nonword Repetition (syllables) was entered in the second step. PSI $P(k)$ accounted for a 19.8% of the variance in German receptive learning, $R^2 = .198$, $F(1, 36) = 8.92$, $p = .005$, whereas Nonword Repetition accounted for an additional 4% of the variance (R^2 change *ns*). When the order of entry was reversed such that Nonword Repetition was entered on the first step and PSI $P(k)$ was entered on the second step, Nonword Repetition accounted for 8.3% of the variance, $R^2 = .083$, $F(1, 36) = 3.25$, $p = .08$; and PSI $P(k)$ accounted for an additional 15.5% of the variance (R^2 change, $p = .011$). Regardless of the order of entry, the relationship between PSI $P(k)$ and German receptive learning remains significant. Nonword Repetition is not significantly independently related to German receptive learning, although the correlation approaches significance when Nonword Repetition is entered on the first step.

In the second set of analyses, German productive learning served as the dependent variable. Entered on the first step, PSI $P(k)$ accounted for 20.7% of the variance in German productive learning, $R^2 = .207$, $F(1, 36) = 9.4$, $p = .004$, whereas Nonword Repetition accounted for an additional 13.1% of the variance (R^2 change, $p = .001$). When the order of entry was reversed, with Nonword Repetition being entered on the first step, Nonword Repetition accounted for 20.1% of the variance, $R^2 = .201$, $F(1, 36) = 9.06$, $p = .005$ and PSI $P(k)$ accounted for an additional 13.7% of the variance (R^2 change, $p = .011$). Both PSI $P(k)$ and Nonword Repetition significantly predicted German productive learning, independently, regardless of their order of entry into the model.

Discussion

The lack of significant correlation between nonword repetition and PSI indicates that we achieved a fairly pure measure of phonological store capacity that is unrelated to phonological sequence learning ability. This is a consequence of using nonword repetition items that are extremely low in wordlikeness, whose repetition could thus not be supported by contributions from long-term knowledge of the linguistic regularities captured in lexis and phonotactics (Bailey & Hahn, 2001).

Scores on the PSI were significantly correlated with performance on both the German receptive and German productive tasks, and these relationships were independent of performance on the nonword repetition task. Thus, an identifiable part of the acquisition of both receptive and productive L2 vocabulary form does seem to depend on those cognitive factors measured by the PSI task. This task was designed as a measure of the specific ability to learn phonological sequences. Nevertheless, it is the only "learning" task in Experiment 1, and caution thus dictates that further research is necessary to assess the degree to which PSI does, indeed, specifically measure phonological abilities rather than a generalized "learning factor."

Performance on the Nonword Repetition task was significantly related to vocabulary acquisition in the German productive task but not in the receptive task. This corroborates the findings of Edwards and Lahey (1998) that nonword repetition is more closely associated with measures of productive rather than receptive competence in SLI children, and of Gathercole et al. (1997) that digit span correlates more highly with productive vocabulary. The existence of a relationship between nonword repetition and German productive learning independent of phonological sequencing skill demonstrates that the ability to accurately articulate a novel word is an additional, separable component of L2 productive vocabulary acquisition. This suggests that one component in the correlation between phonological store capacity and vocabulary acquisition that has been reported in the literature comes from the role of the phonological store in generating articulatory sequences. Ellis and Beaton (1993b) also illustrate this separability of factors involved in receptive and productive vocabulary learning, with pronunciability of the foreign word (in terms of native norms) affecting productive but not receptive vocabulary learning.

Individual differences in PSI were not, however, related to native language vocabulary knowledge as indexed by the English Lexical Decision performance. Why might the ability to learn the phonological regularities of a language constrain the learning of novel foreign language vocabulary but not knowledge of low frequency native language vocabulary in adult speakers? Three potential explanations come to mind. First, it may be that print exposure is the primary predictor of knowledge of low frequency native language words (see Stanovich & Cunningham, 1992). Words as rare and register specific as those illustrated in Appendix D are unlikely to have been experienced by all participants, and those students who have encountered them are likely to have done so through print rather than conversation. Second, the modality mismatch between the presentation of stimuli in the PSI and lexical decision tasks attenuate any relationship between them, particularly because these low frequency words are typically irregular in the print–sound correspondences. Third, it is possible that phonological short-term memory plays a determining role in acquiring new phonological structures only in the early stages of acquiring novel L2 vocabulary. Several studies suggest that in learners with considerable familiarity with a language, new vocabulary acquisition is mediated largely by existing lexical knowledge representations (existing vocabulary, cognates, and semantics), and that it is more in the earlier stages of learning words in a new language where phonological short-term memory plays a determining role in acquiring new phonological structures (Baddeley, Papagno, & Vallar, 1988; Gathercole et al., 1992; Masoura & Gathercole, 1999).

In sum, Experiment 1 demonstrated that phonological sequence learning and phonological store capacity as measured using repetition of nonwordlike nonwords are separable cognitive components. Phonological sequence learning is a significant predictor of receptive vocabulary learning. Moreover, phonological sequence learning and phonological store capacity make independent additive contributions to productive vocabulary learning.

EXPERIMENT 2

The findings of Experiment 1 warrant triangulation, if not replication, and it would be appropriate to do this in a situation that has higher ecological validity.

Experiment 2 therefore investigated a parallel set of hypotheses to those tested in Experiment 1, but this time within the context of longitudinal study of classroom Spanish foreign language vocabulary acquisition. Participants were all novice-level students of Spanish taking part in a 10-week Spanish course. As in Experiment 1, individual differences in the ability to learn phonological regularities were measured using the PSI, and the capacity of the store was measured using the nonwordlike Nonword Repetition task. These abilities were assessed at the beginning of the students' course (T1), and we could therefore determine how much individual differences in these abilities which our novices brought to the task of language learning affected their rate of Spanish acquisition over the subsequent months of study and their final performance in Spanish at the end of the course (T2).

In addition, however, we included at T2 a phonological short-term store measure that potentially tapped long-term knowledge of Spanish phonological sequences. This test of Spanish nonword repetition was included to provide a combined measure of capacity of the store and the extent to which phonological regularities of Spanish have been learned. We believed nonword repetition for Spanish wordlike material would additionally tap knowledge of the phonological regularities of Spanish because (a) native speakers can better repeat nonwords that are high rather than low in wordlikeness in their language (Gathercole, 1995b; Gathercole et al., 1991; Gathercole & Martin, 1996) and (b) bilingual children's short-term memory performance in each language mirrors their overall linguistic competence, with greater vocabulary knowledge being associated with higher levels of recall of both words and nonwords in that language (Masoura & Gathercole, 1999; Thorn & Gathercole, 1999).

Therefore, we expected that this test's combined sensitivity to individual short-term capacity and long-term knowledge should result in correlating more strongly with Spanish language proficiency than does ability on nonwordlike nonword repetition. Further, we expected that the difference between repetition of nonwords high in Spanish wordlikeness and nonwords low in wordlikeness should be greater for those of superior Spanish vocabulary knowledge than for those of lesser Spanish vocabulary knowledge.

Method

Participants. Participants were 44 first year undergraduate students. All were enrolled in the Spanish for nonspecialists course (novice level) at the University of Wales, Bangor. This course is designed for students whose main subject of study is not Spanish but who wish to study Spanish as part of the first year of a modular degree. This was a 10-week course which offered two 1-hr classes per week; in addition, students were also expected to attend a 0.5 hr conversation class with a native Spanish speaker and to spend 1 hr per week in the language laboratory. Students on the course were split into four classes and were taught by two different teachers. One teacher taught 38 of the participants and six were taught by another. The course syllabus and text were identical for all participants.

Participants were ages 18–50. Most spoke English as a first language, but this was not true of all: six spoke Welsh as a first language, but were nonetheless proficient in English, one spoke Greek, and two spoke German as their first language. Participants were recruited following a short talk by the experimenter at the end of one of their first Spanish lessons. They were paid £6 for their participation in this experiment.

Design. A within subjects design was used. There were three predictor variables: nonword repetition, phonological sequence learning ability, and initial levels of Spanish receptive vocabulary (close to floor for these ab initio learners). The dependent variables were student performance in their end of course Spanish exam and their Spanish nonword repetition performance at T2. This was a longitudinal study and participants were tested on two separate occasions: T1 was at the beginning of their Spanish course (2–3 weeks after the start of the course) and T2 was 7 weeks later (2–3 weeks before the end of the course). Participants sat the end of course Spanish exam 5 weeks after formal tuition had ended (this included a 3-week vacation period and was 15 weeks after T1 testing).

Stimuli and procedure. The nonword repetition task and the phonological sequence learning task were administered as described above in Experiment 1 at both T1 and T2 testing sessions. The Spanish nonword repetition task was administered at T2 and scored as outlined in the procedures of Experiment 1.

THE SPANISH RECEPTIVE VOCABULARY TASK. This task used a lexical decision paradigm to assess vocabulary knowledge; participants were asked to distinguish between real Spanish words and nonword foils. One hundred twenty target Spanish words were selected from the 2000 most frequent items in written Spanish according to the Rodrigues Bou frequency list. This list does not assign a frequency value to each word; instead, items are listed in order of frequency. The 120 items consisted of every 16th item from this list. The Spanish nonword foils were selected from those used in the LLEX 2.1 Spanish vocabulary test (Meara & Buxton, 1987). Target items and foils were presented on the Macintosh monitor individually in random order. Participants indicated whether they believed each item to be a word or a nonword by clicking on an appropriately labeled button on the screen.

THE SPANISH NONWORD REPETITION TASK. Spanish nonwords were created by changing letters of Spanish words randomly. These candidate nonword stimuli were then rated for wordlikeness by a group of eight native Spanish speakers. Those stimuli that were rated as being highest in Spanish wordlikeness were then selected for use as stimuli. These 32 nonwords varying in length from two to eight syllables are shown in Appendix E. An audiorecording of these stimuli was made using a Superscope cassette recorder and an Altai microphone. An adult male whose first language was Spanish spoke the stimuli slowly and clearly with an approximate 3-s interval between each item.

Participants were verbally instructed that they would hear a taperecording of some nonsense words that sound like Spanish words and that their task was to repeat each word as accurately as possible. Nonwords were presented individually, pausing the tape in between each nonword to allow the participant unlimited time to attempt to repeat the item. Target items were presented in ascending order of length.

EXAM PERFORMANCE. Twenty-five percent of the students' final marks for the Spanish course were dependent on performance in this written exam. The exam paper was composed of seven separate written questions, each contributing a separate percentage toward the total exam mark.

Question 1 was a video comprehension task in which students were shown a video of six short interviews with native Spanish speakers and were asked to note down information imparted in each interview.

Question 2 was a listening comprehension task in which students were presented with two short audiotaped conversations between two native Spanish speakers who described where they lived. Students were required to answer four questions in English about the content of these dialogues.

Question 3 was a reading comprehension task. Students read a short passage (approximately 250 words in length) and were required to answer four questions concerning the content of the passage.

Question 4 was a test of both productive and receptive vocabulary. Students were presented with a small town map; their task was to write two interrogative sentences concerning directions to locations on the map and to write one sentence in Spanish describing directions from one point on the map to another. Further, students were presented with a three sentence passage describing another location on the map and were required to identify this point.

Question 5 was a test of written production. Students were required to write a 50-word passage in Spanish describing the amenities in the town portrayed in the map used in Question 4.

Question 6 tested receptive vocabulary knowledge. Students were given a list of 28 high frequency words and asked to arrange words from this list into 10 pairs of "opposites" (words that can be reasonably contrasted).

Question 7 was a test of reading comprehension and written production. Students were presented with eight statements written in Spanish and were required to write an interrogative sentence in Spanish that could reasonably act as a precursor to each of these statements.

Students were allowed 1 hr to complete these exam questions that were later marked by the course teachers. Responses to Questions 1, 2, 6, and 7 each accounted for 12% of the overall exam grade; Questions 3, 4, and 5 accounted for 20% each. Marks awarded for each question on the exam paper were summed to provide an overall measure of exam performance, *Exam total*. In order to provide an overall measure of receptive competence, *Exam mean receptive*, the marks for Questions 1, 2, 3, and 6 were averaged.

Reliability. A random sample of 10% of the responses from both the nonword repetition task and the Spanish nonword repetition task were assessed and scored by a second rater. Interrater reliability was assessed in terms of scores on the nonword repetition (syllables) measure. The interrater reliability of the nonword

Table 3. *Descriptive statistics for all tasks in Experiment 2*

Variable	*M*	Max	*SD*	*n*
Spanish receptive vocabulary task (T1)				
T1 *P*(*k*)	.2	1	.14	42
PSI				
T1 PSI *P*(*k*) overall	.70	1	.16	42
T2 PSI *P*(*k*) overall	.80	1	.17	42
Nonword Repetition				
T1				
Words	10.26	32	3.43	35
Syllables	63.66	144	19.15	35
T2				
Words	11.93	32	3.60	42
Syllables	72.36	144	18.29	42
Spanish Nonword Repetition (T2)				
Words	11.98	32	5.70	41
Syllables	112.17	162	20.74	41
Exam performance (T2)				
Exam total	16.97	20	4.84	38
Exam mean receptive	9.39	14	2.35	38
Q1. Exam video comprehension	9.45	12	2.58	38
Q2. Exam listening comprehension	4.26	12	1.83	37
Q3. Exam reading	13.29	20	5.29	38
Q4. Exam productive and receptive	13.61	20	5.16	38
Q5. Exam writing	7.79	12	3.09	38
Q6. Exam receptive vocabulary	10.18	12	2.26	38
Q7. Exam receptive and productive	7.05	12	3.13	38

Note: Max, maximum score possible; PSI, Phonological Sequencing Index; Q1–Q7, Questions 1–7.

repetition task, using Cronbach's alpha, at T1 was .97 and at T2 was .99. The interrater reliability of the Spanish nonword repetition task using Cronbach's alpha was .99.

Results

There were a number of missing data points: two participants did not attend the T2 testing session, seven sets of T1 nonword repetition data were lost as a result of theft from the experimenter's car, and six of the students did not take the end of course exam.

Table 3 displays the mean scores and standard deviations of all the variables. As expected for beginning learners, T1 performance in the Spanish receptive vocabulary task was close to floor levels. Performance levels in the PSI task were reasonably stable across T1 and T2, as was performance on the nonword repetition task. The reliability of these measures over time demonstrated reasonable stability;

the test–retest reliability of PSI was .58 ($p < .01$), and the test–retest reliability of the nonword repetition measure was .73 ($p < .01$) when scored by syllables and .44 ($p < .01$) when scored by words. The pattern of intercorrelations across various submeasures of exam performance was not uniform, suggesting that different exam components were tapping a range of different abilities. Although intercorrelations between the various submeasures of exam performance were generally high (mean $r = .43$, $SD = .14$), two submeasures (Questions 3 and 7) did not correlate uniformly with all other measures of exam performance the r values for Question 3 ranging from .16 to .59 and those of Question 7 ranging from .16 to .47.

The relationship between PSI, Nonword Repetition, and Spanish receptive vocabulary at T1. As in our previous experiment, PSI was independent of the ability to repeat nonwordlike nonwords: the correlation between T1 PSI $P(k)$ and T1 nonword repetition (syllables) was $r = .14$ (*ns*). The beginners learners' knowledge of Spanish receptive vocabulary, being close to zero, was also independent of both T1 PSI $P(k)$ ($r = .09$, *ns*) and T1 nonword repetition (syllables; $r = .15$, *ns*). Thus, these three abilities are essentially independent at the beginning of the course.

T1 PSI, Nonword Repetition, and Spanish receptive vocabulary as predictors of T2 Spanish Nonword Repetition. Each of these T1 measures significantly predicted Spanish Nonword Repetition ability at the end of the course, the respective correlations being T1 PSI $P(k)$ at $r = .33$, $p < .05$, T1 Nonword Repetition (syllables) at $r = .38$, $p < .05$, T1 Spanish receptive vocabulary at $r = .34$, $p < .05$. The fact that these three abilities were independent at T1 suggests that they are making independent contributions to Spanish Nonword Repetition ability.

Stepwise forced-entry multiple regressions with Spanish Nonword Repetition ability as the dependent variable were used to determine this. When T1 Nonword Repetition (syllables) and T1 PSI $P(k)$ were entered at Stage 1 they jointly explained 22% of the variance ($R^2 = .215$, $p = .01$) and entering T1 Spanish receptive vocabulary at Stage 2 accounted for an additional 8.8% of the variance (R^2 change, $p = .04$). When T1 Nonword Repetition (syllables) and T1 Spanish receptive vocabulary were entered at Stage 1 they jointly explained 24% of the variance ($R^2 = .244$, $p = .006$), and entering T1 PSI $P(k)$ at Stage 2 accounted for an additional 5.9% of the variance (R^2 change, $p = .09$). When T1 PSI $P(k)$ and T1 Spanish receptive vocabulary were entered at Stage 1 they jointly explained 20% of the variance ($R^2 = .196$, $p = .018$), and entering T1 Nonword Repetition (syllables) at Stage 2 accounted for an additional 10.6% of the variance (R^2 change, $p = .025$).

The relationship between PSI and measures of L2 lexical competence. Pearson correlations were used to investigate the relationship between PSI and measures of L2 vocabulary competence. The T1 PSI $P(k)$ correlated significantly with two measures of exam performance: Question 1 (video comprehension, $r = .36$, $p < .05$) and Question 6 (receptive vocabulary, $r = .35$, $p < .05$). The correlations

Table 4. *Pearson correlations relating Nonword Repetition and performance in exam tasks*

	Total	Mean Recp.	Q1	Q2	Q3	Q4	Q5	Q6	Q7
T1 Nonword Repetition									
Words	−.02	−.06	−.07	−.13	.12	.11	.02	.06	−.25
Syllables	.18	.18	.17	.06	.41*	.20	.13	.14	−.14
T2 Nonword Repetition									
Words	.07	−.01	−.02	−.02	.16	−.00	.05	.04	−.35*
Syllables	.14	.09	.01	.07	.35*	.04	.12	.15	−.16
Spanish Nonword Repetition									
Words	.52*	.5**	.36*	.41*	48**	43**	42**	.35*	.11
Syllables	.61**	.61**	.46**	.49**	.55**	.57**	.47**	.42**	.19

Note: Total, exam total; mean recp., exam mean receptive; Q1, Question 1 (video comprehension); Q2, Question 2 (reading comprehension); Q3, Question 3 (listening comprehension); Q4, Question 4 (productive and receptive); Q5, Question 5 (writing); Q6, Question 6 (receptive vocabulary); Q7, Question 7 (receptive and productive).
$^{*}p < .05$. $^{**}p < .01$.

between T1 PSI $P(k)$ and both the mean of exam receptive tasks ($r = .30$, $p = .06$) and Question 4 (productive and receptive task, $r = .28$, $p = .09$) approached significance. The correlations between T2 PSI $P(k)$ and Question 2 (listening comprehension, $r = .27$, $p = .10$); Question 6 (receptive vocabulary, $r = .28$, $p = .09$), and exam mean receptive ($r = .27$, $p = .09$) all approach significance. Thus the language outcome measures to which the PSI is most closely related are those tasks that tap auditory word recognition skills, but not those that principally tap visual word recognition skills such as Question 3 (reading comprehension) and Question 7 (receptive and productive).

The relationship between Nonword Repetition and measures of lexical competence. Table 4 shows the coefficients relating Nonword Repetition and exam performance. Measures of the ability to repeat speech-synthesized nonwords were not significantly related to overall exam performance but were related to performance in Question 3 (reading comprehension) at both T1 and T2.

In contrast, measures of Spanish Nonword Repetition were significantly associated with virtually all measures of exam performance, with a correlation between Spanish Nonword Repetition (syllables) and overall exam performance of $r = .61$, $p < .01$.

Multiple regression analyses. A series of stepwise forced-entry multiple regression analyses was conducted in order to assess how much of the variance in exam performance could be accounted for by PSI and nonword repetition performance. Here, as in Experiment 1, the power of these analyses is low, given that $df = 42$. Under such conditions, large effects and low error is required in order to achieve significant results. Notwithstanding the lack of power, the pattern of results is nonetheless informative.

In the first set of analyses exam mean receptive performance acted as the outcome measure. When entered in the first step, T1 PSI $P(k)$ accounted for 8.6% of the variance in mean exam performance ($R^2 = .086$, $p = .07$); entering T1 PSTM in the second step explained a negligible 0.4% of variance (R^2 change, *ns*). When T1 PSTM was entered first it explained just 1.7% of the variance ($R^2 = .017$, *ns*), and entering T1 PSI $P(k)$ in the second step explained an additional 7.2% of the variance (R^2 change, $p = .10$). Neither variable is significantly related to mean receptive exam performance but, given the low power, it is noteworthy that PSI approaches significance.

In order to assess whether the relationship between PSI and exam mean receptive performance was due to a confound with initial knowledge of Spanish, T1 Spanish receptive vocabulary knowledge was entered as the first step in a second set of analyses, where it accounted for 6.7% of the variance ($R^2 = .067$, $p = .11$). When PSI $P(k)$ was entered at a second step it accounted for an additional 6.9% of the variance (R^2 change, $p = .10$). In contrast, when PSTM was added at a second step it added a negligible 3% of extra variance (R^2 change, *ns*).

Wordlikeness and Spanish competence as determinants of Nonword Repetition. Is it Spanish wordlike Nonword Repetition ability that discriminates successful Spanish learners, rather than Nonword Repetition ability per se? In order to make the performance on these tasks directly comparable, Nonword Repetition (syllables) and Spanish Nonword Repetition (syllables) scores were expressed as a proportion of their possible maxima. Participants were then classified as good or poor Spanish learners on the basis of whether their exam mark fell above or below the mean exam mark.

A two-way mixed samples analysis of variance (ANOVA) was conducted. There were two independent variables: wordlikeness (low/high “Spanish” wordlike; within subjects) and exam grade (low/high grade; between subjects). The dependent variable was Nonword Repetition performance. This analysis revealed a significant main effect of wordlikeness, $F(1, 36) = 64.06$, $p < .001$, and a significant interaction between wordlikeness and exam grades, $F(1, 36) = 5.05$, $p = .03$. Figure 1 shows that Spanish nonwords were repeated better than nonwordlike nonwords and that this was particularly the case for the better learners of Spanish.

Discussion

These results confirm that the conclusions of Experiment 1 apply in a more naturalistic context and, further, they extend our understanding of these relationships. As in Experiment 1, Nonword Repetition and PSI were not significantly associated. Here again, by using a Nonword Repetition task, the content of which was neither wordlike in Spanish nor in English, we produced a measure of phonological storage capacity that is unrelated to phonological sequence learning ability.

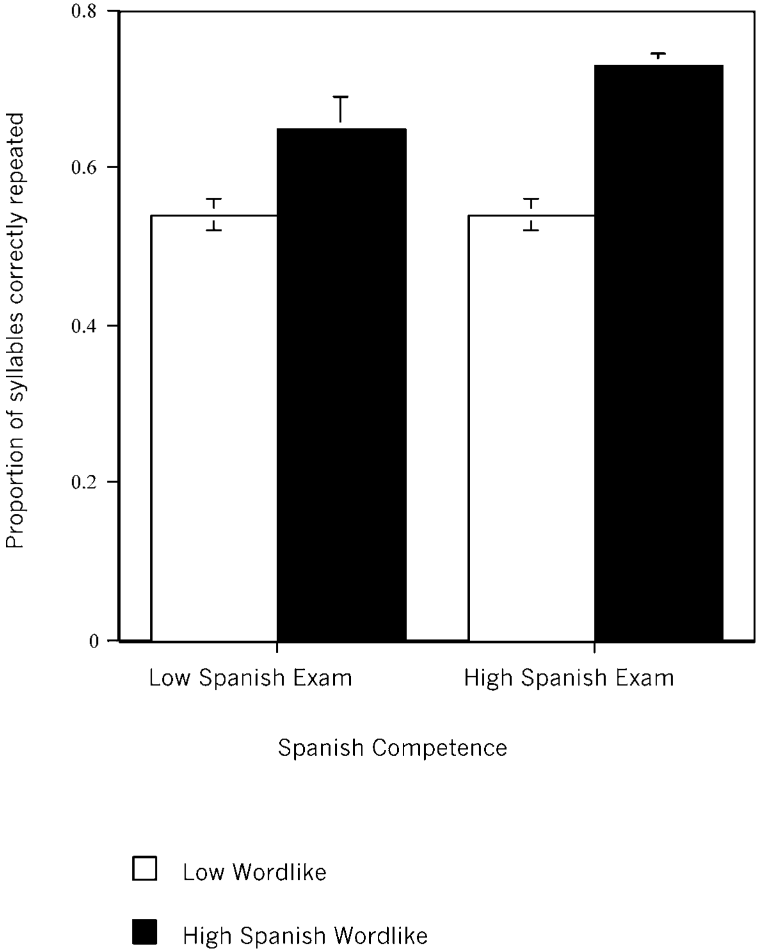

Figure 1. The effect of Spanish wordlikeness and student Spanish competence on Nonword Repetition ability ($n = 36$). Error bars represent standard errors.

Although PSI was not strongly related to the overall measure of exam performance, it did significantly correlate with a number of submeasures of exam performance and the common theme that unites these is that each required receptive competence. Here then, as in Experiment 1 (Table 2), PSI was independently correlated with receptive learning. These findings suggest that performance in the PSI measures a discrete source individual difference that is dissociable from short-term store capacity and that is a significant determinant of L2 vocabulary development. However, the interpretation of this result comes with a number of provisos. First, although PSI was designed to measure individual differences in the ability to learn phonological regularities, further research is needed in order segment performance on the PSI task into the portion attributable to a generalized learning factor, that attributable to an ability to learn sequences, and, further, that specific to the learning of phonological sequences. Second, the relationship between PSI and mean receptive exam performance merely approaches significance, but does not achieve it. Replication is required in the context that permits sufficient statistical power for necessary analyses. Third, unlike Experiment 1, learning is not measured directly in our second experiment, in that there is no single pre-/posttask to permit direct comparison of vocabulary knowledge at the beginning and the end of the course.

These findings also confirm previous reports that when the material to be remembered in phonological short-term memory tasks is language-like, then phonological short-term memory is not a language-independent system; rather, it functions in a highly language-specific way by capitalizing on relevant long-term phonological knowledge (Ellis, 1996; Masoura & Gathercole, 1999; Thorn & Gathercole, 1999; Vitevitch, Luce, Charles–Luce, & Kemmerer, 1997). Spanish Nonword Repetition was designed not only to tap the efficiency of short-term storage mechanisms but also to measure the extent to which participants had learned the regularities of Spanish over the 10-week course. It would, of course, have been better to also measure Spanish PSTM at T1 as well as at T2 so that knowledge of the regularities of Spanish gained prior to the course could then have been partialed out. Nonetheless, Spanish PSTM was significantly related to performance on both the T1 PSI task and the T1 PSTM task, with T1 Nonword Repetition and T1 Spanish receptive vocabulary making significant, and T1 PSI marginally significant independent contributions to T2 Spanish Nonword Repetition. Further performance on the Spanish Nonword Repetition task was significantly related to virtually all measures of Spanish ability, as assessed by exam performance at end of course. These patterns of results suggest that Spanish Nonword Repetition is a task that calls both upon the capacity of the phonological short-term store and upon knowledge of the regularities of Spanish.

There is no reason to expect this relationship between wordlike Nonword Repetition performance and knowledge of language to be all or none, indeed, gradations of association are more likely. Gathercole, Frankish, Pickering, and Peaker (1999) argued that if vocabulary competence is not only a function of the capacity of the short-term phonological store but also a function of the ability to exploit the regularities of the language, it follows that individuals of high lexical competence should be better able to exploit the regularities in wordlike items than individuals of lesser lexical competence. Gathercole et al. (1999) compared 16 children of high and low vocabulary knowledge on repetition for nonwords high, low, and very low in wordlikeness. Numerical trends in the data suggested that the difference between repetition of nonwords low in wordlikeness and repetition of nonwords high in wordlikeness was greater for those who were good vocabulary learners than for those who are poor vocabulary learners, although the statistical interaction proved insignificant on ANOVA. It seems possible that this null result may have been due to lack of statistical power. (Note that the Table 2 data for Gathercole et al., 1999, were later corrected.) Our data in Figure 1 indicate that by the end of the course, all our learners were better able to repeat nonwords high in Spanish wordlikeness than nonwords low in wordlikeness. Moreover, the advantage of Spanish nonwords over nonwordlike nonwords was more pronounced for those students whose grades were above average on the Spanish exam. These findings also concur with those of Thorn and Gathercole (1999) and Masoura and Gathercole (1999), who demonstrated that bilinguals' short-term memory performance mirrors their overall linguistic competence such that greater vocabulary knowledge is associated with higher levels of recall of both words and nonwords in that language. Taken together these results indicate that short-term memory for linguistic material is a product not only of phonological short-term store capacity but also of the efficiency with which an individual is able to call upon knowledge of the regularities of the language. Participants of greater lexical competence are better able to exploit these regularities to support short-term retention of novel words than are participants of lesser lexical competence.

Our account of the processes that underpin the overall pattern of results of our two experiments is as follows. At the very beginnings of learning a language, phonological store capacity and sequence learning ability are more readily separable. However, as exposure to the language increases, so does the degree to which a learner begins to recognize repeated phonological sequences and to abstract their regularities determines the extent of the long-term knowledge base. The greater is the learning rate, the greater the resultant receptive vocabulary, and thus the greater the possibility for this long-term memory contribution to short-term repetition of wordlike materials. By this account individual differences in the efficiency of both phonological sequence learning and phonological short-term storage mechanisms independently contribute to receptive vocabulary knowledge. Receptive vocabulary knowledge in its turn then contributes to the subsequent efficiency of both phonological storage of language-like material and further subsequent learning of similar phonological sequences. The process is one of sequence learning and chunking (Ellis, 1996, 2002) all the way up, repeated cycles of differentiation and integration (Studdert–Kennedy, 1991).

In conclusion, these results confirm the association between phonological short-term memory ability and L2 vocabulary acquisition. They illustrate that the capacity of the short-term phonological store places constraints on lexical acquisition. However, more importantly, they indicate that the combined effect of the capacity of the store and the ability to learn phonological regularities is more closely related to both productive and receptive L2 lexical competence than the capacity of the store alone. In analyzing the factors that contribute to phonological short-term memory for native-language words and nonwords either high or low in phonotactic frequency, Gathercole et al. (1999) estimated that approximately 50% of phonological short-term memory performance in their experiments was supported by the phonological store, 40% by contributions from long-term lexical knowledge, and 10% by phonotactic knowledge. Our experiments have shown that long-term knowledge of phonological sequences has a similarly large role in supporting short-term memory for novel L2 words and for the long-term consolidation of L2 vocabulary. Further research is clearly needed to understand the detailed processes that underlie this bootstrapping into language.

APPENDIX A

Nonword Repetition stimuli

No. Syllables				
1	sɔ	boɷ	bɔ	gi
2	nɔgi	lubu	pɔkɔ	boɷnu
3	mɔtudoɷ	vukudu	lɔkinu	moɷfubu
4	nugubɔvɔ	fɔsulipu	foɷnimɔki	kudoɷnoɷfoɷ
5	guvutɔfusu	pitidɔminɔ	siputɔnɔvu	poɷdoɷgɔnɔvoɷ
6	tɔgumusɔbuku	toɷdusoɷpɔmiki	boɷdufutivoɷmu	kɔfoɷdoɷmoɷbisi
7	voɷgupɔsoɷludubɔ	dukoɷsiboɷvoɷtulɔ	tipɔdoɷnɔboɷlisoɷ	kilupoɷnufusoɷvi
8	luvuboɷsumudufinoɷ	doɷsubivɔmoɷtɔluki	fisoɷgoɷpoɷboɷtimiloɷ	poɷgoɷdɔnitumusubi

APPENDIX B

Stimuli used in the phonological sequence learning task

Two Syllable	Three Syllable	Four Syllable
	Target Items	
θaɪbɛ	tæɹɔðaɪ	teɪθɛjϖθu
mɪɹaϖ	tʃivæbɪ	ɪəkɪmɪdϖ
ɹɔdʒɪ	hɪθaɪðɪϖ	tʃɔgɑɹɑtæ
tϖpɪ	θɛboϖpɛ	maɪpɔwɛhæ
	Foil Items	
ŋɔθu	koϖdʒeɪvɑ	dʌvaϖdʒɑgw
ɱoϖtʃu	kaɪbϖgeɪ	tɪdʒɔðɔvaϖ
wɪnɑ	tʃɔdʒɛdʒɑ	dϖθəsædʒɪ
dʒaɪvϖ	dʒɑʃɔɪfɛ	θɪfitʃɛti
ðiŋϖ	fægɔpɔ	sɔdʒeɪwəsɑ
θɑtʃeɪ	zeɪhɛɹaϖ	gϖnɑsaɪjϖ
dʒətæ	waɪpɔtdϖ	θʌkaϖʃəmɪ
dʒʌgɔ	bɪlitaϖ	poϖgɛdaϖbw
pϖdʒɑ	ziðoϖkɑ	səbigaɪlɑ
lakeɪ	dʒɑϖhɛkϖ	hʌmoϖŋɔli
dʒɔɪjæ	diwəgaɪ	dʒϖseɪɹɔɪneɪ
dʒɑɪlϖ	lɪpɔɪnaɪ	nɑkɑkaϖʃʌ
viʃɑ	jikɔɪwɪ	vigeɪkoϖŋɑ
dʒϖʃæ	bɑnϖdʒʌ	wɪvɛvϖzɔ
dʒeɪgʌ	seɪwɪɹɪ	lɪðɪtʃʌjaɪw
tʃϖtoϖ	ðʌfɪŋɛ	aϖgɔɪðϖjɛ
kaɪhɑ	jɪtʌgʌ	jϖzisɔgɑ
lɪʃæ	bɛθɪfʌ	wɑkoϖtaϖ
ʃaϖðɔ	tʃɑfɪgɔɪ	hϖdʒϖðɔɪgɑ
ʃəmɔɪ	tɔsɔŋɑ	zoϖdʒϖʃʌtɑ
dʒoϖtæ	miθaϖθɔ	mæjənɔɪbi
paɪzʌ	tætaϖkoϖ	lϖzoϖdʒidʒɪ
ɹoϖɹɔɪ	kaɪvədʒaɪ	dʒɪtʌmaɪɹu
dʒəwɪ	timɛdɑ	wəɹɔðϖfɛ
ɹ ɑwɑ	dʒϖtɔɪdʒʌ	zɔɪdʒɪzɑsɛ
ŋilaɪ	kæɹænɔ	seɪtʌfɪðɑ
dʒɑbɔɪ	jətutɛ	ŋaϖdʒaɪpɑʃi
dʒʌŋϖ	dʒɪdʒuɹæ	ŋiθɪlutæ
ɱoϖpϖ	ŋɛveɪzɔɪ	mənɛlɔɪhaɪ
wuɹoϖ	θɑhɛvi	jɔɪteɪpɪmϖ
θoϖzɔ	sæŋɔɪθu	jɪzϖkʌŋɔɪ

APPENDIX C

Words used in the German learning task and their English translation equivalents

Block 1	Block 2
Ecke–corner	Streichen–to paint
Haben–to have	Graben–to dig
Brauchen–to need	Klippe–cliff
Nehmen–to take	Schere–scissors
Dohle–jackdaw	Rasen–lawn
Hose–trousers	Teller–plate
Sperre–barrier	Rufen–to call
Fliegen–to fly	Zahlen–to pay
Stellen–to put	Mieten–to rent
Kaufen–to buy	Fahne–flag
Leiter–ladder	Stossen–to push
Friseur–hairdresser	Kuche–kitchen

APPENDIX D

The low frequency English words and nonwords used in the English vocabulary test of Experiment 1

Words: periwig, purview, libretti, nystagmus, kosher, derma, chlorosis, transporter, japonica, troth, raffish, gnostic, quotidian, palmate, nascent, sarcophagus, merganser, neoplasm, anachronic, abdicant, kudos, kemp, ogive, trauma, apotheosis, fluoresce, maelstrom, ochrous, protuberance, tyrannic, quiesce, datum, igniform, borzoi, hassock, uterine, traduce, iambic, sorcery, rood, beneficency, otiose, hustings, cerebration, oxymoron, nary, abhorrence, trapezoid, mordant, quisling, edelweiss, grenadine, vaporous, regurgitate, juxtapose, machete, oppugn, jocose, inched, galumph, bergamot, glandulous, urbane, finitude, revivalist, deist, recidivist, marrowbone, bicipital, lothario, rabbinic, anabolism, spheroid, regatta, fakir, beatitude, ratteen, fecund, furrier, lachrymal.

Nonwords: dunphy, artigan, sacrumate, vennard, ashill, pardoe, mynott, prelatoriat, moft, brind, youde, instere, lang, rudall, apsitis, jemmett, stemp, grandon, martlew, obsolation, ackrill, brimble, hamp, copner, cliss, prowt, concannon, gazard, pocock, deliction, keable, croath, murtagh, spraker, amphlett, aimler, whitrow, allard, ashment, carow, berrow, joyle, bendle, ancrum, verdon, acklon, aistrope, condick, nudd, alden, crayonal, picardine, bance, craddock, opinarchy, lanworn, allam, lamble, voule, sandry, wherp, disportal, condron, hebulate, bechelet, ainge, beament, boobier, bamber, auner, pinkard, putbrace, spedding, cordle, skine, fancett, eley, dowling, druce, beap.

APPENDIX E

Nonwords used in the Spanish Nonword Repetition task

No. Syllables	Spanish Nonword	Spanish Nonword in IPA	Spanish Wordlikeness Rating/5
2	termar	tɛrmar	3.6
2	baldor	bældʊr	3.2
2	ardal	ardæl	4
2	orda	orðə	4.4
3	taurete	tæwrɛtɛ	3.8
3	patanco	pætanco	3.8
3	mermallar	mermayiær	4.1
3	laderal	lædɛræl	3.9
4	movilido	mɔbIlido	3.1
4	sucursina	sʊkɔrsinə	3.1
4	antiembre	æntiɛmbrɛ	2.6
4	desboroto	ɾɛsbɔtato	3.5
5	cobrosamente	abrosæmɛntɛ	4.4
5	estancioso	ɛstænðioso	4.9
5	acrecentera	ækreðentɛrə	4
5	rabiosera	ræbiɔsɛrə	3.6
5	emitancia	ɛmItæðiə	2.6
5	oprimalmente	ɔprImælmɛnte	4
6	santificorado	sæntIkɔrædo	2.9
6	embulicioso	ɛmʊlIðIoso	3.5
6	eliminicio	ɛlImInIðIo	3.5
6	conespidiente	kɔnespIdienɾe	3.2
6	proseguienda	prɔsɛrIɛndə	3.2
6	abastologia	æbæsɾɔlɔχIə	3.5
7	decacuad-refecto	dɛkækwædrefekto	3.2 + 3.4
7	trasora-naderio	ɾræsɔrænædɛrio	2.4 + 3
7	curtillo-barajento	kɔrtiyobæræχɛnto	3.8 + 2.4
7	autido-<u>desampato</u>	awɾidodɛsapæto	2.9 + ?
8	crosar-partiferencia	krɔsarpartifɛrɛnðia	3.4 + 3.1
8	tironano-<u>civinista</u>	tirɔnænɔsivinIstə	3.4 + ?
8	ampato-<u>debicario</u>	æmpætodɛbiikarIyo	2.9 + ?
8	mangual-<u>solteramente</u>	mængwælsɔltæramɛntɛ	2.9 + ?

Note: Underlined portions of nonwords were not included in the pretest. IPA, International Phonetic Alphabet.

REFERENCES

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. *Advances in Reading Language Research, 2*, 231–256.

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

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

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

Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language, 27*, 586–595.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language, 44*, 568–591.

Brent, M. R. (1999). Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences, 3*, 294–301.

Cheung, H. (1996). Non-word span as a unique predictor of second-language vocabulary learning. *Developmental Psychology, 32*, 867–873.

De Cara, B., & Goswami, U. (2002). Statistical analysis of similarity relations among spoken words: Evidence for the special status of rimes in English. *Behavioral Research Methods and Instrumentation, 34*, 416–423.

Edwards, J., & Lahey, M. (1998). Non-word Repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics, 19*, 279–309.

Ellis, N. C. (1994a). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 211–282). London: Academic Press.

Ellis, N. C. (Ed.). (1994b). *Implicit and explicit learning of languages*. London: Academic Press.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking and points of order. *Studies in Second Language Acquisition, 18*, 91–126.

Ellis, N. C. (2002). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition (Target article). *Studies in Second Language Acquisition, 24*, 143–188.

Ellis, N. C., & Beaton, A. (1993a). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology. 46A*, 533–558.

Ellis, N. C., & Beaton, A. (1993b). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning, 43*, 559–617.

Ellis, N. C., & Sinclair, S. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *Quarterly Journal of Experimental Psychology, 49A*, 234–250.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Frisch, S. F., Large, N. R., Zawaydeh, B., & Pisoni, D. B. (2001). Emergent phonotactic generalizations in English and Arabic. In J. Bybee & P. Hopper (Eds.), *Proceedings of the symposium on frequency effects and emergent grammar* (pp. 159–180). Amsterdam: Benjamins.

Gathercole, S. E. (1995a). Is non-word repetition a test of phonological memory or long-term knowledge? It all depends on the non-words. *Memory and Cognition, 23*, 83–94.

Gathercole, S. E. (1995b). Non-word repetition: More than just a phonological output task. *Cognitive Neuropsychology, 12*, 857–861.

Gathercole, S. E., & Adams, A. (1993). Phonological working memory in very young children. *Developmental Psychology, 29*, 770–778.

Gathercole, S. E., & Adams, A. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language, 33*, 672–688.

Gathercole, S. E. & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*, 439–454.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language.* Hove, UK: Erlbaum.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 84–95. [table 2 correction, *25*, 562].

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33*, 966–979.

Gathercole, S. E., & Martin, A. J. (1996). Interactive processes in phonological memory. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 73–100). Hove, UK: Psychology Press.

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1991). The influence of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics, 12*, 349–367.

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 38*, 887–898.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Science, 4*, 178–186.

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

Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology, 86*, 527–536.

Jusczyk , P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*, 1–26.

Masoura, E. V., & Gathercole, S. E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology, 34*, 383–388.

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4*, 142–154.

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Mahwah, NJ: Erlbaum.

Nation, P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Erlbaum.

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

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology, 48*, 98–107.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8 month old infants. *Science, 274*, 1926.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

Saffran, J. R., Newport, E. L., Aslin, R. N., & Tunick, R. A. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*, 101–105.

Schmitt, N., & McCarthy, M. (Eds.), (1997). *Vocabulary: Description, acquisition, and pedagogy*. Cambridge: Cambridge University Press.

Service, E. (1992). Phonology, working memory and foreign-language learning. *The Quarterly Journal of Experimental Psychology, 45*, 21–50.

Service, E., & Craik, F. I. M. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language, 32*, 608–623.

Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition, 20*, 51–68.

Studdert–Kennedy, M. (1991). Language development from an evolutionary perspective. In N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch, & M. Studdert–Kennedy (Eds.), *Biological and behavioral determinants of language development*. Hillsdale, NJ: Erlbaum.

Thomson, J., Goswami, U., & Hazan, V. (2003). *The effect of neighbourhood density on phonological awareness in dyslexic and normally-developing readers*. Poster session presented at the First Workshop of the Centre for Human Communication, University College London.

Thorn, A. S. C., & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 52A*, 303–324.

Thorndike, E. L., & Lorge, I. (1944). *The teachers word book of 30,000 words*. New York: Columbia University Teachers College.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 145–152.

Vitevitch, M. S., Luce, P. A., Charles–Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken words. *Language and Speech, 40*, 47–62.