

Application of the Testing and Spacing Effects to Name Learning

SHANA K. CARPENTER* and EDWARD L. DeLOSH

Colorado State University, USA

SUMMARY

Four experiments investigated the effects of testing and spacing on the learning of face-name stimulus-response pairs. Experiments 1a and 1b compared the recall of names following intervening tests versus additional study opportunities and found that testing produced better retention of names. Experiments 2 and 3 explored the effects of repeated tests versus study for massed, uniform, or expanded spacing intervals. Tested names were better retained than studied names, spaced names were better retained than massed names, and memory was best for items tested at spaced intervals. Contrary to past findings, expanded schedules did not yield better memory than uniform schedules in either experiment. Theoretical implications for the testing and spacing effects are discussed, along with effective name-learning techniques based on these principles. Copyright © 2005 John Wiley & Sons, Ltd.

Two reliable memory phenomena that are applicable to everyday learning situations are the *testing effect* and the *spacing effect*. The testing effect refers to the advantage in retention for material that is tested over material that is presented for additional study (Allen, Mahler, & Estes, 1969; Kuo & Hirshman, 1996). For example, Carrier and Pashler (1992) investigated the learning of paired-associate target items as a function of whether the target was retrieved on an intervening cued-recall test or was re-presented for additional study. On a final cued-recall test over all items, retention was enhanced for targets that were retrieved previously relative to those that were given a second study opportunity. Other studies have also observed the beneficial effects of testing on memory using materials such as pictures (Wheeler & Roediger, 1992), general knowledge facts (McDaniel & Fisher, 1991), and prose passages (Glover, 1989; Nungester & Duchastel, 1982).

The spacing effect refers to the retention advantage for information that is repeated in a distributed fashion relative to information that is repeated in a massed fashion (Hintzman, 1974; Melton, 1970). For example, Dempster (1987) repeated paired-associate English vocabulary words and definitions three times, with each repetition occurring either in immediate succession (massed) or with other items in-between (spaced). A significantly greater number of vocabulary words were retained for the spaced schedule than the massed schedule. Bahrick and Phelps (1987) found that Spanish-English vocabulary

*Correspondence to: Dr S. K. Carpenter, Department of Psychology, Colorado State University, Fort Collins, CO 80523, USA. E-mail: shcarp@lamar.colostate.edu

words were better retained after 8 years when studied and re-learned at 30-day intervals, as compared to 1-day or 0-day intervals. The spacing effect has also been demonstrated using non-verbal materials such as pictures (Hintzman & Rogers, 1973).

The effects of testing and spacing can also be combined to enhance memory beyond the individual effects of either factor alone. For example, in studies investigating the effects of multiple test and study trials, test trials produced better memory than study trials, and spaced trials produced better memory than massed trials, such that the greatest overall retention occurred for tested items presented at spaced intervals (Cull, 2000; Izawa, 1992; Landauer & Eldridge, 1967). Thus, not only can testing and spacing improve memory independently (see Dempster, 1996; Schmidt & Bjork, 1992), their effectiveness can also be enhanced by combining factors in the form of repeated testing at spaced intervals.

One everyday task in which the testing and spacing effects may be useful is in the learning of peoples' names. Name-learning situations often include multiple opportunities for test or study, such that in the former case one attempts to retrieve a person's name, whereas in the latter case one asks the person to repeat their name. Multiple encounters with a person's name can also occur in massed or spaced fashion. Many everyday name-learning situations require a person's name to be retrieved in response to their face, so empirical research aimed at approximating realistic conditions should include the use of face-name stimulus-response pairs.

There is, however, relatively little research on testing and spacing effects using face-name stimulus-response pairs. Such research is, however, of theoretical importance because explanations of testing and spacing are largely based on verbal learning studies, and there is good reason to question whether findings from verbal learning studies generalize to face-name learning. The primary goal of the current study is to determine whether or not testing and spacing effects do generalize to learning situations in which a person's face serves as a cue and their name as the target, and to consider the general implications of our findings for theories of testing and spacing effects.

FACE-NAME LEARNING

In studies using unfamiliar face-name stimulus-response pairs, it is well-documented that names are more difficult to recall than other types of biographical information about a person (Cohen & Burke, 1993; Cohen & Faulkner, 1986; Stanhope & Cohen, 1993). The advantage in recall for a person's occupation relative to a person's name is obtained even when the same word (e.g. *baker*) is encoded as an occupation for one group and as a surname for a second group (McWeeny, Young, Hay, & Ellis, 1987). One popular explanation for the deficit in recall of names relative to other personal identity information is that names are less meaningful than other words used in everyday communication, and are therefore less likely to be represented in semantic memory (Cohen, 1990; Cohen & Burke, 1993). Cohen (1990) measured the rate of recall for both a name and personal possession associated with a face, as a function of whether the possession was a meaningful item (e.g. *dog*) or a meaningless non-word item (e.g. *wesp*). The rate of recall for names was similar to that of meaningless possessions and inferior to that of meaningful possessions.

Manipulations designed to increase the meaningfulness of the association between a face and name reduce the usual deficit in name recall. Cohen (1990), for example, presented photographs of people paired with names and occupations, and varied the names

and occupations so that they represented meaningful words or meaningless non-words (e.g. *baker* vs. *hapton* for names, and *pilot* vs. *wesp* for occupations). The disadvantage for name recall was reduced when the meaningfulness of the name was highlighted by presenting the meaningful name in the context of a meaningless occupation. Name-learning mnemonics that make use of semantic-based imagery also improve the recall of names from faces. One example is the mnemonic suggested by McCarty (1980), in which names are encoded by imagining the interaction between a person's face and a concrete transformation of their name (e.g. imagining a *bride* and an *ant* interacting with the face of Mr. *Bryant*). Learners usually do not employ such strategies on their own, however, and typically find it more difficult to associate a face to a name than to associate a face to meaningful information such as an occupation or possession. Thus, when learners are left to their own devices, they usually do not establish meaning-based associations between unfamiliar faces and names (Cohen, 1990; WcWeeny et al., 1987).

In short, unfamiliar face-name pairs represent associations between items that are pictorial and difficult to verbalize (a face) and items that are verbal yet relatively meaningless (a name). As such, it is not clear whether the testing and spacing effects observed with meaningful verbal material generalize to unfamiliar face-name pairs. This empirical question has potential implications for theories of testing and spacing effects, particularly with regard to the role of semantic processing in these effects. As an illustration, consider one well-known explanation of the spacing effect.

In his two-process theory, Greene (1989, 1990) proposed that for cued-memory tests, participants intentionally allocate rehearsal to items based on their perception of how easily the items will be remembered in the future. Because massed items seem more familiar than spaced items, they are rehearsed less, resulting in deficient processing of massed items relative to spaced items. Consistent with this view, Greene obtained a spacing advantage on cued-memory tests for intentional but not incidental learning conditions. In a follow-up study, Challis (1993) also observed spacing effects for cued-memory tests given under intentional learning conditions. However, Challis also found a spacing advantage on cued-memory tests under incidental conditions when the encoding task directed attention to semantic features of items (i.e. a pleasantness or imagery rating task), but not when the encoding task directed attention to non-semantic features of items (i.e. a letter identification task). Challis' modified two-process theory therefore specifies that the spacing advantage on cued-memory tests results from deficient *semantic* processing for massed items relative to spaced items, rather than deficient processing per se. Based on this view, one would not expect a spacing effect with materials that are unlikely to engender semantic processing. Thus, if a robust spacing effect is observed with unfamiliar face-name stimulus-response pairs, this would suggest that deficient semantic processing for massed items does not fully account for the spacing advantage on cued-memory tests.

PAST RESEARCH ON FACE-NAME LEARNING

Despite the empirical and theoretical issues outlined above, very few studies on the testing effect have used face-name stimuli, and to our knowledge, no studies on the spacing effect have used face-name stimuli. The few testing-effect studies that do exist were not constrained to face-name stimulus-response pairs or did not clearly differentiate test trials from study trials (see following page). Thus, it is not clear from these studies whether

testing reliably enhances name learning beyond that of additional study when a face is the only cue for retrieval.

Landauer and Bjork (1978, Experiment 2) used pictures of faces that were paired with first and last names and compared the effects of repeated test trials and study trials at two spacing intervals: a uniform schedule, with equal intervals between repetitions, and an expanded schedule, with increasing intervals between repetitions. Following presentation of the face and both names together, the face was presented again along with either the first or last name (the study condition), and the participant was required to retrieve the non-presented name (the test condition). On a final cued-recall test of all face-name stimulus-response pairs, retention was better for names that were tested than for names that were studied. There was also an interaction between the effects of testing and spacing such that tested items were better retained with the expanded interval schedule, but studied items were better retained with the uniform schedule. The advantage in retention for tested items that were repeated according to an expanded schedule was termed the *expanded retrieval effect*. Note, however, that the retrieval of tested items may have been supported by face-name associations and/or name-name associations, so it is not clear whether the testing and expanded retrieval effects hold for circumstances in which a face is the only retrieval cue.

In subsequent research, Morris and Fritz (2000, 2002; see also Morris, Fritz, & Buck, 2004) compared test versus study conditions for a name-learning task presented in a real classroom setting. Students learned the names of their classmates by playing the *name game*, in which one student stated his or her name, then a second student stated this name and also added his or her own name. A third student then stated the first two names and added his or her own name, and so forth. In this manner, each student reviewed the names of all previous students before adding his or her name. In a test condition, students attempted to retrieve previous names, but in a study condition, students were given a list of the previous names and read them aloud. Final retention of names was superior for tested items than studied items. However, the test condition required participants to not only retrieve names, but also to listen to the retrieval of names by other students. Thus, students who were tested also received additional study opportunities by virtue of listening to fellow students.

Three issues have yet to be resolved in the literature on testing and spacing effects in name learning. First, it is not clear whether tests benefit retention of face-name stimulus-response pairs relative to additional study, given the methodological issues described above. Second, it is not clear whether spaced presentations benefit retention of face-name stimulus-response pairs relative to massed presentation. Finally, the combined effects of testing and spacing on name learning are not well established because Landauer and Bjork's (1978) study is the only one to explore the effects of test versus study of face-name stimulus-response pairs at different spacing intervals. This lack of research is of particular concern because studies that have examined the combined effects of testing and spacing in the learning of verbal material (rather than names) have yielded mixed results. Cull (2000) compared test and study trials at massed, uniform and expanded intervals, and found that retention for both test and study trials was highest in the spaced conditions, with no difference between uniform and expanded intervals. Cull, Shaughnessy, and Zechmeister (1996), on the other hand, did observe the expanded retrieval effect when comparing massed, uniform and expanded test schedules, but they did not include study trials for comparison. Given the limited number of studies that have examined different spacing intervals and the mixed findings obtained in those studies, additional research on the combined effects of testing and spacing is warranted.

The current study addresses the three issues outlined above by using face-name stimulus-response pairs under well-controlled conditions that are characteristic of research on testing and spacing effects with verbal material. Specifically, the goals of the study were to: (1) determine whether the testing effect holds for face-name learning by comparing retention associated with a single test without feedback to retention associated with an additional study opportunity (Experiments 1a and 1b), (2) determine whether the spacing effect holds for face-name learning by comparing retention associated with spaced presentations to retention associated with massed presentations (Experiments 2 and 3), and (3) explore the potential combined effects of testing and spacing by comparing retention for repeated test trials versus repeated study trials administered at varying spacing intervals (Experiments 2 and 3). If the testing and spacing effects do not hold for unfamiliar face-name pairs, this would suggest that semantic processing plays an important role in these effects. If the testing and spacing effects generalize to face-name pairs, this would implicate other mechanisms instead of or in addition to semantic processes.

EXPERIMENT 1A

The primary purpose of Experiment 1a was to determine whether or not the advantage for testing over studying, as observed with verbal material, extends to face-name stimulus-response pairs. Participants learned to associate individual names to faces by either attempting to retrieve the name when presented with only the face (test condition) or by viewing an additional presentation of the face and name together (study condition). In order to approximate the type of everyday name-learning situation in which a person's name is not immediately available to verify the accuracy of retrieval, the test conditions in all four experiments did not involve feedback. If name learning is enhanced after a single test opportunity without feedback, relative to an additional study opportunity, then more tested items than studied items should be retrieved on the final cued-recall test in Experiment 1a.

Method

Participants

Sixty-two undergraduate students volunteered for the experiment in order to fulfil partial requirements for an introductory psychology course at Colorado State University. Participants were tested individually on personal computers.

Materials

A total of 48 colour photographs of non-famous faces (24 males and 24 females) were sampled from the online database at the University of Stirling (2003). All of the faces in the photographs were adult Caucasians, but faces were excluded if judged by the experimenters to have highly emotional expressions, distinctive accessories (i.e. glasses or jewellery), unusual hair colour or style, or unusual facial features. Each photograph was taken in front of a white background and was approximately 6 cm by 9 cm in size.

Forty-eight surnames were sampled from the norms published by Zechmeister, King, Gude, and Opera-Nadi (1975). Selected names were within one standard deviation of the mean ratings for frequency, familiarity, orthographic distinctiveness and pronounceability. Further sampling constraints were used so that all names were similar in length (between

five and eight letters), and dissimilar in phonology and orthography. The number of names sharing the same first letter was limited to three, and these names were comprised of unique subsequent letters to reduce confusion.

Design

The 48 faces were used to construct 12 different lists of four faces, with two males and two females in each list. The 12 lists of faces were the same for all participants and were given in the same order. Two of the lists were used as practice trials, one to illustrate study trials and the other to illustrate test trials. For practice lists, the pairing of names to faces was the same for all participants. For the experimental lists, five different random pairings of names to faces were used, counterbalanced across participants.

Out of the ten experimental lists, five were presented as study trials (ST) and five as test trials (TT). Both ST and TT consisted of a re-presentation of the face, along with the instructions, 'Please type the name of this person, followed by the ENTER key.' For ST, however, the name was re-presented below the face, requiring the participant to simply copy the name instead of retrieving it from memory. Thus, ST and TT were as similar as possible in all respects except for the act of retrieval in the TT. Lists were randomly assigned to the ST and TT conditions, then an alternate version was created such that all lists appearing as ST in the original version appeared as TT in the alternate version, and vice versa. These original and alternate versions were counterbalanced across participants. At the end of the experiment, all participants completed a cued-recall test to test memory for the names of the 40 experimental faces.

Procedure

Participants first read instructions on the computer monitor, then viewed the two practice lists followed by the ten experimental lists. Face-name pairs were presented sequentially for 6 s each, with the face in the centre of the computer screen and the name directly below it. The first letter of each name was capitalized and preceded by the appropriate title 'Mr.' or 'Ms.', depending on whether the face was male or female. Items within a list were presented in a new random order for each list and each participant. Following presentation of the last list item, participants completed a 15 s distracter activity in which they were required to add together single digit numbers presented at 1 s intervals. Following the distracter task, all of the items from the most recent list were presented as a ST or TT, with the items shown in a new random order for each list and each participant. Participants were instructed to type in the name of each person, followed by the Enter key, or the Enter key alone if they were unable to retrieve the correct name during the TT. The names that participants typed appeared on the screen directly below the face for TT, and directly below the face and re-presented name for ST. There was no time limit for entering names, and no feedback was provided to verify the accuracy of responses on TT.

After completion of the last experimental list, participants completed another distracter task in which they were asked to type in the names of as many US states as they could remember during a 5-min time interval. After the 5-min distracter task, participants read the instructions for the final memory test. Participants were informed that they would be presented all of the faces from the experiment and would be required to type in the names of as many of them as they could remember. All 40 faces were then presented one at a time in a new random order for each participant. Similar to intervening TT, participants were instructed to type in each person's name followed by the Enter key, or just the Enter key if they were unable to retrieve the person's name. The names that participants typed

appeared on the screen directly below the face. Unlimited time was allotted for retrieval of each name.

Results and discussion

We first computed the mean proportion of items correctly entered during the intervening test and study trials. Studied items were entered with 100% accuracy, whereas the mean proportion of tested items correctly entered was 0.58 ($SD = 0.17$). We next computed the proportion of tested versus studied items retrieved on the final test. These proportions were submitted to a 2×5 (Trial Type \times Name-Face Pairing) mixed analysis of variance (ANOVA), with trial type a within-participants factor and pairing a between-participants factor. This analysis revealed a significant main effect of trial type [$F(1, 57) = 40.73$, $p < 0.05$, $MSE = 0.007$] such that tested items ($M = 0.20$, $SD = 0.14$) were better remembered than studied items ($M = 0.10$, $SD = 0.09$). There was no main effect of face-name pairing [$F(4, 57) = 1.65$, $p > 0.05$, $MSE = 0.021$] nor a Trial Type \times Pairing Interaction [$F(4, 57) = 0.89$, $p > 0.05$, $MSE = 0.007$]. Thus, the retention of names was significantly better following TT than ST regardless of the particular pairing of names to faces.¹

To our knowledge, this is the first experiment to show that the testing effect generalizes to face-name stimulus-response pairs under carefully controlled laboratory conditions. There was, however, a potential confound in Experiment 1 that may have contributed to the benefit of TT over ST. On ST, each name was re-presented below the corresponding face and participants were required to type the name directly below the name shown. One participant reported that under these conditions, it was not necessary to look at the face, but only at the name below it, in order to correctly type in the name. Thus, the face-name association may not have received additional study on ST, thereby accounting for the advantage of testing over studying. This potential confound was addressed in Experiment 1b.

EXPERIMENT 1B

Given the paucity of research on testing effects in name learning, Experiment 1b was conducted as a replication of Experiment 1a, with one minor modification. In the study condition, the name was re-presented above the face rather than below it, and the name that participants entered still appeared below the face. As such, the face appeared between the correct name and the participant's entry of the name, making it more likely that participants would look at and attend to the face while entering the name. In all other respects, the method used in Experiment 1b was identical to that used in the above experiment. On the intervening test and study trials for the 59 new participants in Experiment 1b, studied items were entered with 100% accuracy, whereas the mean proportion of tested items correctly entered was 0.55 ($SD = 0.17$). A 2×5 (Trial Type \times Name-Face Pairing) mixed ANOVA showed that memory was significantly better for

¹We also examined final test performance conditionalized upon successful performance during intervening test and study trials. In Experiments 1a, 1b, and 2, analysis of these conditionalized scores revealed the same pattern of significant effects as those of the unconditionalized scores. In Experiment 3, analysis of conditionalized and unconditionalized scores also revealed the same pattern of significant main effects, however a significant Trial \times Interval Interaction for conditionalized scores indicated that final retention was significantly better for test over study trials at uniform and expanded intervals but not at massed intervals.

tested items ($M = 0.20$, $SD = 0.12$) than studied items ($M = 0.11$, $SD = 0.11$), $F(1, 54) = 43.18$, $p < 0.05$, $MSE = 0.006$. There was neither a main effect of face-name pairing [$F(4, 54) = 1.24$, $p > 0.05$, $MSE = 0.027$] nor a Trial Type \times Pairing Interaction [$F(4, 54) = 0.38$, $p > 0.05$, $MSE = 0.006$]. Thus, in all respects, Experiment 1b replicated the results of Experiment 1a.

The testing advantages observed in Experiments 1a and 1b are consistent with the results of previous studies on testing effects in name learning (see Landauer & Bjork, 1978; Morris & Fritz, 2000, 2002; Morris et al., 2004). The present findings also extend on past results in showing that testing effects hold (a) for the learning of face-name stimulus-response pairs when faces serve as the only cues for corresponding names, and (b) under carefully controlled laboratory conditions that disentangled study and test conditions. Thus, the first issue outlined in the Introduction, whether or not the testing effect holds for the learning of face-name stimulus-response pairs, was answered affirmatively. The second and third issues, whether or not the spacing effect holds for face-name stimulus-response pairs and whether or not the effects of testing and spacing interact, were addressed in Experiments 2 and 3.

EXPERIMENT 2

The purpose of Experiment 2 was to examine the spacing effect, as well as possible combined effects of testing and spacing, on name learning. Landauer and Bjork (1978) conducted the only known study using face-name stimulus-response pairs in which test and study trials were administered at uniform versus expanded spacing intervals, but they did not include a massed interval to investigate the spacing effect. In their study, 15 faces were paired with first and last names. During initial presentation, the faces were simultaneously projected onto a screen and each name was referenced on one of 15 cards given to participants. A fixed amount of time was allotted for participants to look at the name on each card and associate it with the face on the screen, and participants were required to change cards in response to a timed cue. For the intervening study and test conditions, all 15 faces were presented simultaneously, but only one name appeared on the card that corresponded to each face, and participants were required to recall the non-presented name that was associated with the presented name. The names that appeared on the card therefore served as study items, whereas the non-presented names that participants retrieved served as test items. There were four repetitions of each item, given at uniform intervals (with three intervening items between repetitions, expressed as 3–3–3–3) or expanded intervals (with no intervening items between presentation and first repetition, one between first and second repetition, three between second and third repetition, and eight between third and fourth repetition, expressed as 0–1–3–8).

Our Experiment 2 was modified in a number of ways to enhance experimental control and achieve greater consistency with more recent research. Study and test items were not associated with one another (as was the case for the first and last names in Landauer & Bjork's, 1978 study), instead each face was paired with just one name, and stimuli were presented sequentially instead of simultaneously. Sequential presentation provided better control of both processing time and the number of intervening items that defined spacing intervals. Similar to the more recent designs of Cull and colleagues (Cull, 2000; Cull et al., 1996; see also Landauer & Bjork, 1978, Experiment 1), we also included a massed condition, used three repetitions instead of four, and increased the number of items to 30.

Based on existing research, we expected better retention for tested items than for studied items, better retention for spaced items than for massed items, and a combination between the effects of testing and spacing such that final retention would be best for tested items given at spaced intervals.

A secondary purpose of Experiment 2 was to examine uniform versus expanded spacing intervals, given that prior research has been mixed as to whether an expanded schedule produces better memory than a uniform schedule for tests without feedback. Landauer and Bjork (1978) obtained an advantage for expanded over uniform tests using face-name pairs, and this pattern was replicated by Cull et al. (1996, Experiments 1 through 4) using first- and last-name pairs and general knowledge facts. However, subsequent research using verbal material has failed to replicate this finding (Cull, 2000).

Method

Participants

Sixty-five undergraduate students were sampled from the same participant pool used in Experiments 1a and 1b. Participants were tested individually on personal computers.

Design

Experiment 2 included 30 of the face-name stimulus-response pairs that were used in Experiments 1a and 1b. A complete within-subjects design was used to manipulate the type of trial (TT or ST) and spacing interval (massed, uniform, or expanded), with spacing defined by the number of intervening items between repetitions. Intervening items consisted of experimental items (test and study items) and filler items (a group of once-presented, untested control items). A total of 18 filler items were required to achieve the appropriate spacing schedules. Each of the 30 experimental items was encountered four times. The first encounter occurred during the initial presentation of all items, and the next three encounters were repetitions of test and study items that were either massed, uniform, or expanded. In the massed condition, no intervening items occurred between the initial presentation and three repetitions (0–0–0). In the uniform condition, three intervening items occurred between initial presentation and the three repetitions (3–3–3). In the expanded condition, one intervening item occurred between initial presentation and first repetition, three intervening items between first and second repetition, and five intervening items between second and third repetition (1–3–5). The uniform and expanded schedules were selected so that the total number of intervening items from initial presentation to final repetition would be the same.

The 30 experimental items were divided into three groups of ten (five males and five females) for the massed, uniform and expanded schedules. Each group of ten was further divided into five items each that served as study items and test items. Three different versions of the experiment were created so that each group of ten items appeared equally often in massed, uniform, and expanded intervals. For each of these three versions, an alternate version was also created so that the group of items appearing as test items in the original version appeared as study items in the alternate version, and vice versa.

Procedure

Participants first read instructions on the computer monitor, then all items were presented in a sequential fashion for 6 s each. Test and study items were repeated three additional

times, paced by participants. The order of presentation of each type of item (test or study) and each spacing condition (massed, uniform, or expanded) was random and unpredictable. Both TT and ST consisted of a re-presentation of the face, along with the instructions, 'Please type the name of this person, followed by the ENTER key,' with the name represented above the face on ST but not TT. No feedback was provided to verify the accuracy of retrieval on TT.

Following the presentation of all items and repetitions, participants engaged in the same distracter activity as used in Experiments 1a and 1b, in which they attempted to remember as many US states as possible during a 5-min period. After the distracter activity, participants were given instructions for the final memory test, then all 48 faces (the 30 experimental faces plus the 18 filler items) were presented in random order and participants were asked to type each person's name followed by the Enter key, or the Enter key alone if they were unable to retrieve the name. Participants were given as much time as needed on the final test.

Results and discussion

The mean proportion of names correctly entered on the intervening test was computed for each trial type (TT or ST) and spacing interval (massed, uniform, or expanded). Studied items were entered with 100% accuracy for each of the three spacing intervals, whereas the mean proportion of names correctly entered on test trials was 0.94 ($SD = 0.11$) for massed intervals, 0.54 ($SD = 0.22$) for uniform intervals, and 0.74 ($SD = 0.21$) for expanded intervals. A one-way repeated measures ANOVA revealed a significant difference in intervening test accuracy across the three spacing intervals [$F(2, 128) = 85.25, p < 0.05, MSE = 0.03$], and Bonferroni's post-hoc tests indicated that intervening test accuracy was significantly higher for the massed interval compared to the expanded interval, and significantly higher for the expanded interval compared to the uniform interval.

The mean proportion of names correctly retrieved on the final test was computed for each trial type and spacing interval (see Table 1). These proportions were submitted to a 2×3 (Trial Type \times Interval) repeated-measures ANOVA, which yielded significant main effects of trial type [$F(1, 64) = 32.29, p < 0.05, MSE = 0.02$] and interval [$F(2, 128) = 33.31, p < 0.05, MSE = 0.03$], and a significant Trial Type \times Interval Interaction [$F(2, 128) = 6.18, p < 0.05, MSE = 0.03$]. Tested items were remembered significantly better than studied items. Bonferroni's post-hoc tests showed that retention was significantly better for the uniform interval compared to the expanded interval, and for the expanded interval compared to the massed interval. Although an advantage for tested as opposed to studied items occurred at both uniform and expanded intervals, there was no advantage for tested as opposed to studied items at massed intervals. Only items from the uniform and expanded conditions were recalled at significantly higher rates than the non-repeated filler items, which were recalled at a mean rate of 0.09 ($SD = 0.07$).

All predictions for Experiment 2 were confirmed and the results fully replicated previous research by Cull (2000), who also observed combined advantages for testing and spacing, along with the absence of a testing effect under massed conditions. Contrary to the findings of Landauer and Bjork (1978), however, expanded test trials did not yield greater retention than uniform test trials. The next experiment considers one possible reason why the expanded retrieval effect was not observed in the present experiment.

EXPERIMENT 3

Bjork (1988) suggests that the expanded retrieval effect occurs to the extent that expanded tests are progressively more difficult relative to uniform tests, provided that the initial retrieval attempt is successful. The expanded test schedule used in our Experiment 2 may not have been progressively more difficult than the uniform schedule, however. In fact, the accuracy rates on the intervening tests suggested that the uniform schedule may have been more difficult than the expanded schedule. With uniform and expanded schedules of 3–3–3 and 1–3–5, respectively, only the last test may have been more difficult in the expanded condition than the uniform condition. Thus, Experiment 3 was conducted to test the assumption that progressive retrieval difficulty enhances retention by comparing the uniform interval of 3–3–3 to a new expanded interval of 3–5–7. This effectively controlled the difficulty of the first retrieval attempt across conditions, while increasing the difficulty of the two subsequent retrieval attempts for the expanded condition.

Method

Sixty-nine new participants were sampled from the same participant pool used in Experiments 1a, 1b and 2. Participants were tested individually on personal computers. The design and procedure were identical to those of Experiment 2, the only difference being the inclusion of the 3–5–7 expanded interval instead of a 1–3–5 expanded interval.

Results and discussion

The mean proportion of names correctly entered on the intervening test was computed for each trial type (TT or ST) and spacing interval (massed, uniform, or expanded). Studied items were entered with 100% accuracy for each of the three spacing intervals, whereas the mean proportion of names correctly entered on test trials was 0.97 ($SD = 0.09$) for massed intervals, 0.54 ($SD = 0.24$) for uniform intervals, and 0.57 ($SD = 0.23$) for expanded intervals. A one-way repeated measures ANOVA revealed a significant difference in intervening test accuracy across the three spacing intervals [$F(2, 136) = 141.7, p < 0.05, MSE = 0.03$], and Bonferroni's post-hoc tests indicated that intervening test accuracy was significantly higher for the massed intervals compared to the uniform and expanded intervals, but was not significantly different for uniform and expanded intervals.

The proportion of names correctly retrieved on the final test was computed for each trial type and spacing interval (see Table 1). These proportions were submitted to a 2×3 (Trial Type \times Interval) repeated measures ANOVA, which revealed significant main effects of trial type [$F(1, 68) = 3.90, p < 0.05, MSE = 0.04$] and interval [$F(2, 136) = 55.29, p < 0.05, MSE = 0.03$], and a non-significant Trial Type \times Interval interaction, $F(2, 136) = 1.14, p < 0.05, MSE = 0.03$. Tested items were better remembered than studied items. Bonferroni's post-hoc tests revealed that retention was significantly better for the expanded and uniform intervals than the massed interval, with no difference between expanded and uniform intervals. Memory was nominally (but not significantly) better for TT over ST in the uniform and expanded conditions, but not in the massed condition. Only items from the uniform and expanded conditions were recalled at significantly higher rates than the non-repeated filler items, which were recalled at a mean rate of 0.11 ($SD = 0.07$).

Table 1. Mean proportion of names recalled on the final test as a function of item type and spacing schedule

Spacing schedule	Type of item		
	Study	Test	Total
Experiment 2 (3–3–3 vs. 1–3–5)			
Massed	0.08 (0.14)	0.08 (0.12)	0.08 (0.13)
Uniform	0.18 (0.22)	0.30 (0.21)	0.24 (0.22)
Expanded	0.12 (0.15)	0.26 (0.23)	0.19 (0.20)
Total	0.13 (0.18)	0.21 (0.21)	
Experiment 3 (3–3–3 vs. 3–5–7)			
Massed	0.11 (0.16)	0.11 (0.14)	0.11 (0.15)
Uniform	0.24 (0.23)	0.32 (0.20)	0.28 (0.21)
Expanded	0.28 (0.23)	0.33 (0.21)	0.30 (0.22)
Total	0.21 (0.21)	0.25 (0.21)	

Note: Standard deviations are given in parentheses.

Like Experiment 2, an expanded retrieval effect was not observed, despite our use of an expanded interval (3–5–7) that would be expected to be progressively more difficult than the uniform interval (3–3–3). This result is contrary to those of Cull et al. (1996) and Landauer and Bjork (1978), who obtained the expanded retrieval effect using expanded schedules of overall spacing that were very similar to the one used here. Note, however, that Cull (2000) used the same schedule as Cull et al. and also failed to obtain the expanded retrieval effect. Possible explanations for these mixed results are discussed below.

GENERAL DISCUSSION

The testing effect and name learning

The current study addressed several questions concerning the applicability of the testing and spacing effects to the retention of names when faces serve as memory cues. First consider the testing effect. Landauer and Bjork's (1978) research on the testing effect in name learning was not constrained to face-name stimulus-response pairs, such that retrieval could have been supported by face-name associations or name-name associations. Other studies by Morris and colleagues used a protocol in which tested items were given additional study opportunities in addition to being tested (Morris & Fritz, 2000, 2002; Morris et al., 2004). The present experiments, on the other hand, focused on the learning of face-name stimulus-response pairs while disentangling testing from additional study opportunities. All four experiments revealed a significant advantage in retention for tested items relative to studied items, extending on past research in showing that tests benefit memory for names when faces serve as the only cues. Also note that in past name-learning studies, items were tested multiple times and/or feedback was provided. The present study demonstrated that the testing advantage is reliable for a single test without feedback (Experiments 1a and 1b) and for multiple tests without feedback (Experiments 2 and 3).

Such a pattern of results using unfamiliar face-name pairs suggests that semantic processing is unlikely to account for the testing effect under all circumstances. A number of theoretical explanations for the testing effect have been proposed (for a review, see

Dempster, 1996), however, the role of semantic processing is presently unspecified, as many theories of the testing effect have not been developed to the same degree as those of the spacing effect. Nonetheless, the current results help to guide future theorizing by demonstrating that significant testing effects do emerge using non-semantic material, which rules out a general explanation based on semantic processing.

The spacing effect and name learning

Next consider the spacing effect as it applies to the learning and retention of face-name stimulus-response pairs. In Experiments 2 and 3, the distribution of intervening tests and intervening study opportunities was either massed or spaced. Both experiments showed that final retention was significantly better for the spaced conditions than the massed conditions, and this held true for different spacing intervals and for both studied and tested items. These findings are consistent with a large body of literature showing that memory is better for spaced than massed study (e.g. Dempster, 1987; Hintzman, 1974; Melton, 1970) and better for spaced than massed testing (e.g. Cull et al., 1996; Glover, 1989; Izawa, 1992; Landauer & Eldridge, 1967; Modigliani & Hedges, 1987). The results also extend on Landauer and Bjork's (1978) findings by adding a massed condition and showing that spaced tests yield better memory than massed tests in a name-learning paradigm, and that this spacing advantage holds when only faces serve as memory cues.

According to the deficient semantic processing hypothesis (Challis, 1993), a significant spacing effect would not be expected for non-semantic material such as unfamiliar face-name pairs. The results of Experiments 2 and 3 demonstrate that the spacing effect is still intact for these stimuli, suggesting that the deficient semantic processing hypothesis does not account for the spacing effect under all circumstances. A potential alternative explanation may be in line with the multifaceted account of spacing effects offered by Russo, Parkin, Taylor, and Wilks (1998), which proposes that semantic processing may underlie the spacing effect for familiar, meaning-based stimuli (such as words), whereas perceptual processing may underlie the spacing effect for new and unfamiliar stimuli (such as faces and names). If enhanced perceptual processing is allocated to spaced items at the expense of massed items, then changing the appearance of the unfamiliar face across each presentation might reduce the spacing effect in future studies.

Testing and spacing combinations

Experiments 2 and 3 also examined the possible combined effects of testing and spacing. In both experiments, an advantage for tested items over studied items was observed for spaced intervals but not massed intervals. In a verbal-learning paradigm, Cull (2000) also obtained a testing advantage for spaced intervals but not massed intervals, but the present results extend this finding to name learning. The current findings are also consistent with studies on the testing effect that revealed no difference in retention for tested versus studied items under conditions of immediate testing (Bregman & Wiener, 1970; Donaldson, 1971; Tulving, 1967), but an advantage for tested items under conditions of delayed testing (Landauer & Eldridge, 1967; Madigan, 1969; Modigliani, 1976; Rothkopf & Coke, 1963, 1966; Wenger, Thompson, & Bartling, 1980; Whitten & Bjork, 1977). An important implication of this finding is that testing may not benefit memory beyond that of additional study unless the tests occur at non-immediate intervals. The best overall retention was observed for tests that were administered at spaced intervals. This combined

effect of testing and spacing has been reported in a number of prior studies using verbal materials (Cull, 2000; Cull et al., 1996; Glover, 1989; Izawa, 1992; Modigliani & Hedges, 1987). The present study is the first, however, to show that this pattern extends to name learning under conditions in which faces serve as the only memory cues.

The expanded retrieval effect

Experiments 2 and 3 included both uniform and expanding spacing intervals to further examine the expanded retrieval effect, given the small number of studies on this effect and the mixed results obtained in those studies. In Experiment 2, retention was best in the uniform (3–3–3) condition, followed by the expanded (1–3–5) condition, then the massed condition. This pattern held for both study and test conditions. In Experiment 3, a longer lag was used for the expanded interval, and results showed that retention was better for the spaced conditions than the massed condition, with no significant difference between uniform (3–3–3) and expanded (3–5–7) intervals. This pattern also held for both study and test conditions. As such, across different experiments and different spacing intervals, we did not observe an advantage of expanded intervals over uniform intervals.

These results contradict those of Landauer and Bjork (1978), who found that expanded intervals were better than uniform intervals for tested items (also see Cull et al., 1996), but uniform intervals were better for studied items. The results are consistent, however, with those of Cull (2000), who used a design similar to that used in our Experiments 2 and 3, but with words rather than names and faces as paired-associates. Cull found no difference in retention for uniform versus expanded intervals, and this held for both test and study trials.

One might argue that the difficulty of successive retrieval attempts may not have been greater for the expanded (1–3–5) interval than the uniform (3–3–3) interval in Experiment 2, and this may explain why we failed to observe an expanded retrieval effect. When we increased the difficulty of subsequent retrieval for the expanded relative to the uniform condition (3–5–7 vs. 3–3–3), we eliminated the advantage for the uniform schedule that was observed in Experiment 2. Indeed, retrieval on a final test is often facilitated by more difficult retrieval conditions at the time of intervening tests. The benefits of testing are larger for free recall than for recognition intervening tests (Bjork & Whitten, 1974), for interfering than for non-interfering intervening test conditions (Cuddy & Jacoby, 1982; Whitten & Leonard, 1980), and for delayed tests than for immediate tests (Madigan, 1969; Modigliani, 1976; Rothkopf & Coke, 1963, 1966). Glover (1989) argues that difficult intervening test conditions facilitate elaborate retrieval processing, which yields a more durable memory trace relative to easier test conditions.

By gradually increasing the lag between repeated tests, the expanded retrieval technique creates conditions that maximize the potential for preventive maintenance in memory. Under an expanded schedule, retrieval is likely to be accurate yet increasingly difficult, which prevents information from being forgotten and at the same time allows for elaborate retrieval processing. Bjork (1988) suggests that the most optimal expanded interval for long-term retention would be one in which the tests are administered just prior to the point at which the item would have been forgotten. Thus, final retention should be enhanced for items tested at more difficult expanded intervals relative to less difficult uniform intervals, provided that the intervening retrieval rates are comparable.

This reasoning does not account for all of the inconsistencies in the data, however. Even when we used an expanding schedule that should have been progressively more difficult

than the uniform schedule, under conditions of comparable intervening test retrieval rates (Experiment 3), the expanded schedule did not yield better memory than the uniform schedule. Also note that Cull et al. (1996) and Cull (2000) used the same expanded and uniform intervals, yet they obtained an expanded retrieval effect in the former study but not in the latter study.

Such inconsistencies may be a result of the degree to which the expanded schedule promotes preventive maintenance. An advantage for expanded over uniform schedules seems to emerge when preventive maintenance is relied upon, but this advantage is eliminated or reversed when preventive maintenance is suppressed and corrective maintenance is relied upon instead. For example, Cull et al. (1996) demonstrated an advantage for expanded over uniform schedules using tests without feedback (which promote preventive maintenance), but no such advantage using tests with feedback (which promote corrective maintenance). Cull (2000) observed a small, non-significant advantage of expanded over uniform schedules using tests without feedback, and a reversal of this trend using tests with feedback. Although our study used tests without feedback and still failed to replicate the expanded retrieval effect, other factors (such as the specific intervals used) may have failed to maximize the potential for preventive maintenance to a necessary degree that would allow the advantage for expanded over uniform schedules to emerge. Further research is clearly needed to determine the boundary conditions for the expanded retrieval effect.

Practical implications for name learning

The present results have several practical implications for everyday name learning. One suggestion is based on the act of retrieval compared to additional exposure. Experiments 1a and 1b demonstrated that retrieving a name only once, relative to receiving an additional presentation, approximately doubled the rate at which names were recalled on a final test. This finding suggests that when attempting to learn the names of new acquaintances, one should actively attempt to retrieve the newly learned names rather than request that the individual repeat their name.

A second suggestion is based on the combined effects of retrieval and spaced practice. Experiments 2 and 3 demonstrated that names repeatedly retrieved were better retained than names repeatedly presented, and names repeatedly retrieved at spaced intervals were retained almost three times as well as names repeatedly retrieved at massed intervals. These findings suggest that in order to maximize the benefits of repeated retrieval, one should attempt to retrieve newly learned names at spaced time intervals rather than in immediate succession. For example, mentally retrieving an acquaintance's name every few minutes during a conversation would be expected to be more effective than mentally retrieving the name in succession at the beginning of the conversation.

Finally, the results of the present study caution against the use of a non-distributed retrieval schedule when using retrieval as a means to improve name retention. Experiments 2 and 3 not only showed that massed retrieval was the least effective repetition schedule, it also rendered the otherwise beneficial effects of retrieval no more effective than that of an additional study opportunity. Furthermore, massed items were not retained any better than items presented only once. Distributed practice not only benefits name retention beyond that of massed practice, but also seems to be necessary in order to reap the benefits of retrieval and improve retention beyond that which would occur for just a single exposure. It is the combination of retrieval and

spaced practice that offers a simple, yet highly effective strategy in improving one's memory for names.

ACKNOWLEDGEMENTS

Shana K. Carpenter, Department of Psychology, Colorado State University; Edward L. DeLosh, Department of Psychology, Colorado State University.

Portions of this research were presented at the Ellis-Battig Symposium of the 2004 Annual Meeting of the Rocky Mountain Psychological Association, Reno, April 15–17. We acknowledge several undergraduate research assistants for their help in data collection and scoring for these experiments: Karena Alons, Melissa Athey, Corey Condo, Jaclyn Gentry, David Gregory, Rachel Perez, Erin Stewart, Tania Valdez, and Aaron White.

REFERENCES

- Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 8, 463–470.
- Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 344–349.
- Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 397–401). New York: Academic Press.
- Bjork, R. A., & Whitten, W. B. (1974). Recency sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, 6, 173–189.
- Bregman, A. S., & Wiener, J. R. (1970). Effects of test trials in paired-associate and free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 9, 689–698.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory and Cognition*, 20, 633–642.
- Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 389–396.
- Cohen, G. (1990). Why is it difficult to put names to faces? *British Journal of Psychology*, 81, 287–297.
- Cohen, G., & Burke, D. M. (1993). Memory for proper names: a review. *Memory*, 1, 249–263.
- Cohen, G., & Faulkner, D. (1986). Memory for proper names: age differences in retrieval. *British Journal of Developmental Psychology*, 4, 187–197.
- Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: an analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior*, 21, 451–467.
- Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, 14, 215–235.
- Cull, W. L., Shaughnessy, J. J., & Zechmeister, E. B. (1996). Expanding the understanding of the expanding-pattern-of-retrieval mnemonic: toward confidence in applicability. *Journal of Experimental Psychology: Applied*, 2, 365–378.
- Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology*, 79, 162–170.
- Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork, & R. A. Bjork (Eds.), *Handbook of perception and cognition: Memory* (pp. 317–344). San Diego, CA: Academic Press.
- Donaldson, W. (1971). Output effects in multitrial free-recall. *Journal of Verbal Learning and Verbal Behavior*, 10, 577–585.
- Glover, J. A. (1989). The ‘testing’ phenomenon: not gone but nearly forgotten. *Journal of Educational Psychology*, 81, 392–399.

- Greene, R. L. (1989). Spacing effects in memory: evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 371–377.
- Greene, R. L. (1990). Spacing effects on implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1004–1011.
- Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 77–99). Potomac, MD: Erlbaum.
- Hintzman, D. L., & Rogers, M. K. (1973). Spacing effects in picture memory. *Memory and Cognition, 1*, 430–434.
- Izawa, C. (1992). Test trial contributions to optimization of learning processes: study/test trials interactions. In A. F. Healy, & S. M. Kosslyn (Eds.), *Essays in honor of William K. Estes: From learning theory to connectionist theory* (Vol. 1, pp. 1–33). Hillsdale, NJ: Erlbaum.
- Kuo, T., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology, 109*, 451–464.
- Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). New York: Academic Press.
- Landauer, T. K., & Eldridge, L. (1967). Effects of tests without feedback and presentation-test interval in paired-associate learning. *Journal of Experimental Psychology, 75*, 290–298.
- Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior, 8*, 828–835.
- McCarty, D. L. (1980). Investigation of a visual imagery mnemonic device for acquiring face-name associations. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 145–155.
- McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology, 16*, 192–201.
- McWeeny, K. H., Young, A. W., Hay, D. C., & Ellis, A. W. (1987). Putting names to faces. *British Journal of Psychology, 78*, 143–149.
- Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9*, 596–606.
- Modigliani, V. (1976). Effects on a later recall by delaying initial recall. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 609–622.
- Modigliani, V., & Hedges, D. G. (1987). Distributed rehearsals and the primacy effect in single-trial free recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13*, 426–436.
- Morris, P. E., & Fritz, C. O. (2000). The name game: using retrieval practice to improve the learning of names. *Journal of Experimental Psychology: Applied, 6*, 124–129.
- Morris, P. E., & Fritz, C. O. (2002). The improved name game: better use of expanding retrieval practice. *Memory, 10*, 259–266.
- Morris, P. E., Fritz, C. O., & Buck, S. (2004). The name game: acceptability, bonus information and group size. *Applied Cognitive Psychology, 18*, 89–104. DOI: 10.1002/acp.948
- Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: effects on retention. *Journal of Educational Psychology, 74*, 18–22.
- Rothkopf, E. Z., & Coke, E. U. (1963). Repetition interval and rehearsal method in learning equivalences from written sequences. *Journal of Verbal Learning and Verbal Behavior, 2*, 406–416.
- Rothkopf, E. Z., & Coke, E. U. (1966). The effect of some variations in procedure on response repetition following verbal outcomes. *Journal of Verbal Learning and Verbal Behavior, 5*, 86–91.
- Russo, R., Parkin, A. J., Taylor, S. R., & Wilks, J. (1998). Revising current two-process accounts of spacing effects in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 161–172.
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207–217.
- Stanhope, N., & Cohen, G. (1993). Retrieval of proper names: testing the models. *British Journal of Psychology, 84*, 51–65.
- Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6*, 175–184.
- University of Stirling. (2003). *The psychological image collection at stirling (PICS)*. Retrieved August 8, 2003, from <http://lionel.stir.ac.uk>

- Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 135–144.
- Wheeler, M. A., & Roediger, H. L., III. (1992). Disparate effects of repeated testing: reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, 3, 240–245.
- Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: effects of spacing. *Journal of Verbal Learning and Verbal Behavior*, 16, 465–478.
- Whitten, W. B., & Leonard, J. M. (1980). Learning from tests: facilitation of delayed recall by initial recognition alternatives. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 127–134.
- Zechmeister, E. B., King, J., Gude, C., & Opera-Nadi, B. (1975). Ratings of frequency, familiarity, orthographic distinctiveness and pronunciability for 192 surnames. *Behavior Research Methods and Instrumentation*, 7, 531–533.