Psychonomic Bulletin & Review
2006, 13 (5), 826-830

What types of learning are enhanced by a cued recall test?

SHANA K. CARPENTER, HAROLD PASHLER, and EDWARD VUL
University of California, San Diego, La Jolla, California

In two experiments, we investigated what types of learning benefit from a cued recall test. After initial exposure to a word pair (A+B), subjects experienced either an intervening cued recall test (A→?) with feedback, or a restudy presentation (A→B). The final test could be cued recall in the same (A→?) or opposite (?→B) direction, or free recall of just the cues (Recall As) or just the targets (Recall Bs). All final tests revealed a benefit for testing as opposed to restudying. Tests produced a direct benefit for information that was retrieved on the intervening test (B). This benefit also "spilled over" to facilitate recall of information that was present on the test but not retrieved (A). Both theoretical and practical implications are discussed.

Memory tests are commonly used to measure the accuracy or speed of memory. They can also be used to *modify* memory—sometimes in a beneficial way. Duchastel (1981), for example, showed that students remembered textbook information better if they completed test questions on the material instead of engaging in an unrelated activity. Furthermore, a number of studies have shown that testing is even more beneficial than additional study presentations (Carpenter & DeLosh, 2006; Carrier & Pashler, 1992; Kuo & Hirshman, 1996, 1997; Wheeler, Ewers, & Buonanno, 2003). This benefit for tested as opposed to restudied information is often referred to as the *testing effect* (see Dempster, 1996, for a review).

We can shed light on why the testing effect occurs by asking what types of learning can benefit from testing. Are testing benefits confined to the very items that were retrieved on the test? Or do they also occur for items that were on the test but not retrieved? If the benefits are confined to the retrieved items, do they manifest only when the final and intervening tests are the same? We examined these questions using cued recall (A→B). Previous research indicates that a cued recall test (A→?) is more beneficial than restudy (A+B) when the final test is cued recall in the same direction (A→?) (Carpenter & DeLosh, 2005; Carrier & Pashler, 1992; Cull, 2000; Izawa, 1969, 1992). Do these benefits also occur when the final test is cued recall in the opposite direction (?→B), or free recall of just the targets (Recall Bs) or cues (Recall As)?

This question has clear practical implications. Many researchers have argued that the testing effect may have important and unexploited educational potential (e.g., Chan, McDermott, & Roediger, 2006; Dempster, 1989, 1996; Glover, 1989; McDaniel & Fisher, 1991; Roediger & Karpicke, 2006). Before accepting this assertion, however, we must know whether these benefits occur for all sorts of memory, or solely for one. For example, one's enthusiasm for using testing to enhance the learning of the German–English correspondence *Hund*↔*Dog* would be tempered if a test (*Hund*→?) enhanced forward recall but not backward recall (?→*Dog*). In the present study, we explored the breadth of the testing effect to determine when testing might be beneficial or harmful in comparison with restudy opportunities.

This work was supported by the Institute of Education Sciences (U.S. Department of Education, Grant R305H040108). We thank Matt Bielich for his programming expertise in both experiments. Correspondence concerning this article should be addressed to S. K. Carpenter, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109 (e-mail: scarpenter@ucsd.edu).

EXPERIMENT 1

In Session 1, subjects were presented with 40 weakly related cue–target pairs. After a study presentation, subjects were given an additional chance to learn each pair. This took the form of either restudying the pair (A+B) or taking a cued recall test (A→?) immediately followed by a presentation of the pair (A+B). These two types of learning opportunities are referred to here as *study trials* and *test/study trials*, respectively, and their duration was always equal. The following day (Session 2), subjects completed one of four different types of final tests: cued recall in the same (A→?) or opposite (?→B) direction relative to the test/study trial of Day 1, or free recall over just the cues (Recall As) or just the targets (Recall Bs).

Method

Subjects. We recruited subjects from an online pool of individuals who volunteered to complete the experiment in exchange for enrollment in a cash prize drawing. Out of the 365 subjects who began the experiment, 90 dropped out during Session 1, 49 during Session 2, and 50 failed to follow instructions (e.g., they waited longer than 48 h to complete Session 2). The remaining 176 subjects were randomly distributed across the four final test conditions: A→? (n = 43), ?→B (n = 53), Recall As (n = 45), and Recall Bs (n = 35).

Internet testing allowed us to collect data from a larger and more demographically diverse group of subjects than would have been possible with standard laboratory testing. Although Web-based data collection has only recently become common, parallel patterns of results have been obtained in numerous laboratory- and Web-based experiments both in our own research and in others' (e.g., Birnbaum, 1999; Krantz & Dalal, 2000; McGraw, Tew, & Williams, 2000; Reips, 2002). An analysis of subjects' reports on their participation environment provided further reassurance that even those who participated in a semipublic environment did not show any detectable decrements or changes in performance (see the Results section). Because Web-based studies have higher dropout rates than laboratory studies do, the critical manipulation of study versus test/study was carried out within subjects to avoid the possibility of differential dropout effects.

Materials. From Wilson's (1988) database, we obtained 80 nouns that were 5–7 letters and 1–3 syllables in length, and high in concreteness (400–700) and frequency (at least 30 per million). Free-association norms (Nelson, McEvoy, & Schreiber, 1998) were used to create 40 weakly associated pairs of similar forward and backward strength (see the Appendix). Each word in a pair was randomly assigned as a cue or target for each subject.

Design and Procedure. The subjects read instructions, answered several demographic questions, and indicated what type of environment they were in while doing the experiment (e.g., at home, in an office, in an Internet café, in a library). The experiment began with the presentation of the 40 word pairs, shown one at a time, for 6 sec each. The cue appeared on the left and the target on the right, each in separate boxes with the labels *cue* and *target* above them.

We used a 2 × 2 × 2 mixed design. The within-subjects factor (test condition: test/study vs. study) was manipulated across items during Session 1. First, all 40 word pairs were presented; then subjects completed a test/study trial on 20 of the word pairs. During a test/study trial, subjects were instructed to covertly retrieve the target within 4 sec while the *cue* box displayed the cue and the *target* box was blank. After 4 sec, the target appeared and both items remained present in their respective boxes for an additional 2 sec. For the other 20 word pairs, subjects completed a study trial in which they were given an additional opportunity to view the cue and target in their respective boxes for 6 sec. Session 1 was complete after all 40 word pairs were presented in either a test/study trial or a study trial. The assignment of items to test condition and their order of presentation were random for each subject.

The two between-subjects factors (item retrieved on final test, cue vs. target; type of final test, cued recall vs. free recall) were manipulated during Session 2, which subjects could complete online from 18 to 48 hours following Session 1. A combination of the two factors yielded four final test conditions, and subjects were randomly assigned to one of them. They were instructed to do one of the following: (1) type the correct target when given the cue (A→?), (2) type the correct cue when given the target (?→B), (3) type all of the targets they could remember (Recall Bs), or (4) type all of the cues they could remember (Recall As). No time limit was imposed, and no feedback was provided. Session 2 was completed when subjects typed an answer to all 40 items for the cued recall tests, or when they clicked a button marked *finish* to indicate that they could no longer remember any items for the free recall tests.

Results and Discussion

Most subjects (72% in Session 1, and 76% in Session 2) reported that they performed the experiment while in a room alone. The rest were more or less evenly distributed among other environments. Environment during the final test did not significantly affect accuracy, nor did it interact with any variables.

Test/study trials produced higher final test accuracy (40% overall) than did study trials (30% overall). The testing benefit occurred regardless of the nature of the final test (see Figure 1). When the final test required cued recall in the same (A→?) or opposite (?→B) direction, the testing benefit was 14%. When the final test required free recall of the targets (Recall Bs) or the cues (Recall As), the testing benefits were 8% and 6%, respectively. The

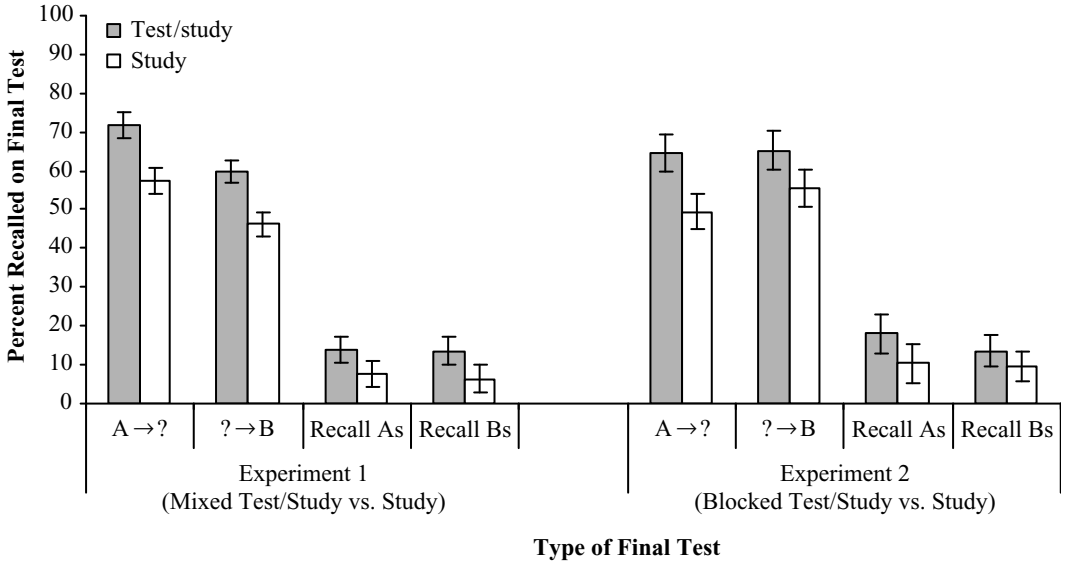

Figure 1. Percent of items recalled on the final tests. Items experienced an intervening cued recall test with feedback (test/study) versus a restudy opportunity (study). Test/study items were recalled better than study items whether the final test required cued recall in the same (A→?) or opposite (?→B) direction relative to that in the intervening test, or free recall of just the cues (Recall As) or just the targets (Recall Bs). Error bars represent standard errors.

significance of these effects was examined in a 2 (test condition) × 2 (item retrieved on final test) × 2 (type of final test) ANOVA. There was a main effect of test condition [$F(1,172) = 101.58$, $MS_e = 0.009$, $p < .001$], but no main effect for item retrieved on the final test ($F = 3.10$) or interaction ($F = 0.15$).

There were three additional significant effects. The first was a main effect for type of final test, indicating that, unsurprisingly, cued recall was easier than free recall [$F(1,172) = 244.76$, $MS_e = 0.083$, $p < .001$]. The second was an interaction between test condition and type of final test, indicating that the testing effect was larger for cued recall than for free recall [$F(1,172) = 11.99$, $MS_e = 0.009$, $p < .005$], probably because free recall was subject to floor effects. The third was an interaction between type of final test and item retrieved on the final test, indicating that cued recall showed an advantage of targets (A→?) over cues (?→B), whereas free recall did not (Recall As = Recall Bs) [$F(1,172) = 3.975$, $MS_e = 0.083$, $p < .05$].[1]

In sum, the key results from Experiment 1 are: The benefits of the intervening cued recall test occurred regardless of whether the final test required cued recall in the same or opposite direction as the intervening test. Furthermore, these benefits occurred even when the final test mandated free recall of the items for which retrieval was required on the intervening test (targets) and free recall of the items for which retrieval was not required (cues).

EXPERIMENT 2

The methodology of Experiment 1 had one possible disadvantage. During a study trial, subjects were given 6 sec to just read the word pair again. During that time, they might have thought about other, previously presented word pairs. Subjects sometimes use the time available during presentation of one item to think about a previous difficult-to-learn item (Slamecka & Katsaiti, 1987). Therefore, the test/study items could have been considered more difficult, because it is harder to access information through retrieval than mere presentation.

When test/study and study pairs are presented in random order, as in Experiment 1 (. . . test/study → study → study → test/study → study → test/study . . .), a previous test/study pair can be easily retrieved during presentation of a study pair. This is harder to do when the order is blocked so that all test/study pairs come before study pairs (. . . test/study → test/study → test/study → study → study → study . . .), and it is impossible to do when the order is blocked so that all study pairs come before all test/study pairs (. . . study → study → study → test/study → test/study → test/study . . .).

Method

In Experiment 2, subjects received one block of 20 study pairs followed by one block of 20 test/study pairs, or vice versa. The order of the blocks and the items within them were random for each subject. In all other respects, Experiment 2 was identical to Experiment 1. We recruited new subjects from the same pool as before. Out of the 177 who began the experiment, 43 dropped out during Session 1, 24 during Session 2, and 28 failed to follow instructions. The remaining 82 subjects were randomly distributed across the final test conditions: A→? ($n = 19$), ?→B ($n = 19$), Recall As ($n = 18$), and Recall Bs ($n = 26$).

Results and Discussion

Most subjects (74% in both sessions) completed the experiment while in a room alone, and the rest were more or less evenly distributed among the other environments. As in Experiment 1, environment during the final test did not significantly affect accuracy, nor did it interact with any variables.

Test/study produced higher final test accuracy (40% overall) than did study (31% overall). The testing benefit appeared regardless of the nature of the final test (see Figure 1). When the final test required cued recall in the same (A→?) or opposite (?→B) direction, the testing benefits were 14% and 9%, respectively. When the final test required free recall of the targets (Recall Bs) or the cues (Recall As), the benefits were 4% and 8%, respectively.

Experiment 2 replicated the same basic pattern of ANOVA results from Experiment 1: a significant main effect for test condition [$F(1,78) = 30.80$, $MS_e = 0.011$, $p < .001$], but no main effect for item retrieved on the final test ($F = 0.48$), and no interaction ($F = 0.04$). Two other significant effects were found: a main effect for type of final test, reflecting the fact that cued recall was easier than free recall [$F(1,78) = 111.25$, $MS_e = 0.076$, $p < .001$], and an interaction between type of final test and test condition, indicating that the testing effect was larger for cued recall than for free recall [$F(1,78) = 4.01$, $MS_e = 0.011$, $p < .05$]. This effect probably occurred because free recall was subject to floor effects.

The blocked order of test/study versus study made it unlikely that cues were retrieved during the intervening test. Nonetheless, Experiment 2 still showed a testing effect for final tests that were in either the same direction as that of the intervening test, or in the opposite direction, and for items that were required to be retrieved (targets) and items that were not required to be retrieved (cues).

GENERAL DISCUSSION

In two similar experiments, what we refer to as a test/study trial—an intervening cued recall test (A→?) followed by re-presentation of the word pair (A+B)—enhanced retention more than a comparable amount of time for pure study (A+B). This result held true, whether retention was tested for cued recall in the same (A→?) or in the opposite (?→B) direction in comparison with the intervening test, as well as for free recall of either the targets (Recall Bs) or the cues (Recall As). The significant testing effect in the same direction replicates previous reports (Carpenter & DeLosh, 2005; Carrier & Pashler, 1992; Cull, 2000; Izawa, 1969, 1992). The present study, however, extends these findings by showing that the testing effect is not specific to the items for which retrieval was required on the intervening test or to the type of testing employed.

Theoretical Implications

Additional study time. We obtained no evidence that tested items benefit simply because they receive more study time than nontested items. First, an intervening test/study trial was more beneficial than a study trial, even though the cue and target were presented together for more time in a study trial (6 sec) than in a test/study trial (2 sec). Second, we obtained a significant testing effect whether test/study and study trials were mixed (Experiment 1) or blocked (Experiment 2). It therefore seems unlikely that a test/study trial produces superior learning because it "steals" study time away from a study trial. Our results are consistent with those of past studies that obtained the testing effect using blocked lists (e.g., Carrier & Pashler, 1992) and between-subjects manipulations of test versus restudy (Wenger, Thompson, & Bartling, 1980).

Transfer-appropriate processing. The processes required by an intervening test and final test are more similar, as compared with the processes required by an intervening study opportunity and final test. According to a transfer-appropriate processing (TAP) view (see, e.g., Morris, Bransford, & Franks, 1977), tests could benefit learning simply because they provide practice at the relevant aspects of the task that are needed for the final test. Some studies have supported this notion by showing that intervening tests are more effective if they are more similar to the final test (e.g., McDaniel & Fisher, 1991; McDaniel, Kowitz, & Dunay, 1989). In the present experiments, however, the intervening test always required recall in one direction (A→?). Moreover, we observed a testing effect whether the final test required recall in the same (A→?) or opposite (?→B) direction, or recall of just the targets (Recall Bs) or the cues (Recall As). In agreement with past studies and contrary to the TAP view, we found that an intervening test was beneficial to retention even if the final test was of a different type (Carpenter & DeLosh, 2006; Glover, 1989; Kang, McDermott, & Roediger, 2005).

Error correction learning. Carrier and Pashler (1992) proposed an explanation for the testing effect on the basis of error-correction learning models (e.g., McClelland & Rumelhart, 1986). According to this view, the association between two items (A→B) is learned by adjusting connections in a network in a way that minimizes the error in producing B from A. If the retrieval of B is required (A→?), as in a test/study trial, then learning occurs by comparing one's actual response (B′) with the desired response (B) to determine how much adjustment is necessary (see also Mozer, Howe, & Pashler, 2004). When both items are present (A+B), as in a pure study trial, learning is impoverished. This is because the availability of B makes it harder for the system to ascertain what response it would produce on its own, thus interfering with the calculation of appropriate weight changes. It is not clear how this hypothesis would account for the advantage of test/study trials when the final test runs in the opposite direction (?→B), since the to-be-retrieved item A was never produced on the intervening test.

Practical Implications

The generality of the testing effect suggests that tests have great potential to enhance learning in practical domains. Specifically, the use of flashcards likely improves recallability not only in the direction that was practiced (e.g., German–English vocabulary *Hund→Dog*), but also in the direction that was not practiced (*Dog→Hund*). Tests might also be useful in improving patients' recall of medical information, which is frequently misremembered (see, e.g., Kessels, 2003). For example, a patient's memory for symptoms and medications may be improved by attempting to recall what medication to take when experiencing specific symptoms.

By regularly using tests with feedback in lieu of restudying the same material over again, it appears that there is much to be gained and little, if anything, to be lost. A promising direction for further research would be to explore how the testing effect might be obtained for knowledge that is more complex and structured than the paired associate information examined in this and other studies of testing effects.

REFERENCES

Birnbaum, M. (1999). Testing critical properties of decision making on the Internet. *Psychological Science*, **10**, 399-407.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, **19**, 619-636.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, **34**, 268-279.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, **20**, 633-642.

Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, **135**, 553-571.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, **14**, 215-235.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, **1**, 309-330.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. C. Carterette & M. P. Friedman (Series Eds.) & E. L. Bjork & R. A. Bjork (Vol. Eds.), *Handbook of perception and cognition: Vol. 10. Memory* (pp. 317-344). San Diego: Academic Press.

Duchastel, P. C. (1981). Retention of prose following testing with different types of tests. *Contemporary Educational Psychology*, **6**, 217-226.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, **81**, 392-399.

Izawa, C. (1969). Comparison of reinforcement and test trials in paired-associate learning. *Journal of Experimental Psychology*, **81**, 600-603.

Izawa, C. (1992). Test trials 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: Vol. 1. From learning theory to connectionist theory* (pp. 1-33). Hillsdale, NJ: Erlbaum.

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition*, **30**, 823-840.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III (2005, May). *Testing enhances memory retention, but which test format is better?* Poster session presented at the annual meeting of the American Psychological Society, Los Angeles.

Kessels, R. P. C. (2003). Patients' memory for medical information. *Journal of the Royal Society of Medicine*, **96**, 219-222.
Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 35-60). San Diego: Academic Press.
Kuo, T.-M., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology*, **109**, 451-464.
Kuo, T.-M., & Hirshman, E. (1997). The role of distinctive perceptual information in memory: Studies of the testing effect. *Journal of Memory & Language*, **36**, 188-201.
McClelland, J. L., & Rumelhart, D. E. (1986). A distributed model of human learning and memory. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2: Psychological and biological models* (pp. 170-215). Cambridge, MA: MIT Press.
McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, **16**, 192-201.
McDaniel, M. A., Kowitz, M. D., & Dunay, P. K. (1989). Altering memory through recall: The effects of cue-guided retrieval processing. *Memory & Cognition*, **17**, 423-434.
McGraw, T. O., Tew, M. D., & Williams, J. E. (2000). The integrity of Web-delivered experiments: Can you trust the data? *Psychological Science*, **11**, 502-506.
Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior*, **16**, 519-533.
Mozer, M. C., Howe, M., & Pashler, H. (2004). Using testing to enhance learning: A comparison of two hypotheses. In *Proceedings of the Twenty Sixth Annual Conference of the Cognitive Science Society* (pp. 975-980). Hillsdale, NJ: Erlbaum.
Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Available at w3.usf.edu/FreeAssociation.
Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, **49**, 243-256.
Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, **17**, 249-255.
Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory & Language*, **26**, 589-607.
Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 135-144.
Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, **11**, 571-580.
Wilson, M. (1988). The MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, **20**, 6-10.

NOTE

1. The advantage of targets over cues does not seem inconsistent with the principle of associative symmetry (for an excellent review, see Kahana, 2002); this advantage was modest in Experiment 1 ($p = .048$) and nonexistent in Experiment 2. Rather, the advantage seems to have been influenced by random error combined with the fact that the subjects were not aware of what type of final test they would receive. Thus, they could have reasonably expected another test in the same direction instead of the opposite direction.

APPENDIX

Item Pairs		Forward Strength	Backward Strength	Absolute Value Difference
angle	corner	.020	.029	.009
author	poet	.028	.035	.007
beach	blanket	.012	.016	.004
block	street	.040	.019	.021
chain	fence	.022	.031	.009
child	mother	.030	.010	.020
cloth	table	.012	.026	.014
coffee	morning	.025	.034	.009
college	student	.035	.046	.011
curve	shape	.018	.011	.007
engine	machine	.033	.027	.006
factory	product	.020	.028	.008
frame	window	.014	.013	.001
group	meeting	.027	.041	.014
guard	prison	.024	.020	.004
lunch	supper	.019	.028	.009
master	owner	.010	.028	.018
nation	state	.042	.055	.013
native	foreign	.056	.031	.025
nature	trail	.023	.012	.011
novel	story	.034	.034	0
object	symbol	.014	.021	.007
office	doctor	.014	.010	.004
paint	picture	.036	.031	.005
pencil	point	.021	.073	.052
people	world	.014	.030	.016
quarter	dollar	.061	.027	.034
range	rifle	.015	.028	.013
report	weather	.015	.024	.009
sheet	cover	.021	.053	.032
slave	worker	.069	.062	.007
smile	teeth	.061	.042	.019
sound	speaker	.024	.027	.003
station	radio	.067	.095	.028
stick	branch	.067	.047	.020
store	general	.016	.028	.012
taste	touch	.016	.012	.004
throat	voice	.039	.020	.019
train	plane	.051	.049	.002
vehicle	truck	.013	.014	.001
Mean		.029	.032	.002

(Manuscript received October 18, 2005;
revision accepted for publication January 10, 2006.)