The Effects of Test Trial and Processing Level on Immediate and Delayed Retention

Sau Hou Chang

Indiana University Southeast

Completed in

July 2013

Abstract

The purpose of the present study was to investigate the effects of test trial and processing level on immediate and delayed retention. Seventy-six college students were randomly assigned first to the single test and the repeated test trials, and then to the shallow processing level and the deep processing level to study forty stimulus words. Results showed that single test trial enhanced immediate retention and deep processing enhanced immediate and delayed retention. Findings also showed an interaction between test trial and processing level in delayed recall. When information was processed at a shallow level, single test trial enhanced delayed retention. When information was processed at a deep level, single test trial and repeated test trial performed similarly in delayed retention. The processing level at encoding seems to mediate the benefits of test trials at retrieval practice in delayed retention.

Keywords: test trial, processing level, retention

The Effects of Test Trial and Processing Level on Immediate and Delayed Retention

Testing is usually viewed as a way of assessing how much students know, but is seldom seen as a way of enhancing students' learning. However, Roediger and Karpicke (2006a, 2006b) argued that taking a test had a greater positive effect than studying the material on future retention. Such an improved performance from taking a test is known as the *testing effect*.

The research design of testing effect usually includes a study phase, an intervening phase, and a test phase. During the study phase, participants take study trials to study some set of material varying from word lists to prose passages. During the intervening phase, some participants take study trials again to study the material (or sometimes study the material several times) and some take test trials of the material (or sometimes take the test trials several times). During the test phase, participants are given a final retention test of the material. The typical finding is that those participants who take test trials outperform those who take study trials during the intervening phase.

Evidence for the testing effect in promoting learning comes from laboratory studies, educationally related studies, and classroom studies. Laboratory studies typically use word lists as material, and free recall as test. For example, Wheeler, Ewers, and Buonanno (2003) asked participants to study a 40-word list in either one study trial or four study trials and then take four consecutive test trials or one test trial respectively. Final free-recall tests were given to participants either 5 minutes or 1 week later. Results revealed a huge advantage for repeated study trials on the immediate free-recall test, but repeated test trials were found to be favorable on the final free-recall test given a week later.

Dempster (1997) identified two hypotheses to account for the positive effects of test trials on learning. The first hypothesis stated that the testing effect was a result of additional exposure to material and overlearning of the material during the test trials. The idea that encoding accounted for testing effect was rejected because the testing effect occurred even when exposure to the material was equated in the study trials when participants were asked to study the material several times, and in the test trials when participants were given a test several times (Roediger & Karpicke, 2006a; Wheeler, Ewers & Buonanno, 2003). If participants encode the material at different processing levels, would there be testing effect?

The second hypothesis stated that the testing effect was a result of retrieval processes that reactivated and operated on memory traces. Bjork (1975) pointed that retrieval might increase the elaboration of a memory trace and multiply retrieval routes, and argued that depth of retrieval might operate similarly to depth of processing at encoding (e.g., Craik & Tulving, 1975). In addition, Bjork and Bjork (1992) further explained that more effortful retrieval practice did enhance storage strength and promote more permanent and long-term learning. The elaboration of memory trace and creation of an effortful retrieval are better able to explain the testing effect. If memory trace is strengthened by retrieval, would three test trials be better than one test trial in retention?

With a sizable research on the testing effect, several variables have been investigated: the material to be learned, the format of the test trial and final retention test, the feedback received on the test trial, the time interval between study and test trials, and the interval before the final retention test. However, little is known about the encoding and retrieval of the material.

To examine whether elaboration at encoding contributed to long-term retention, Karpicke and Smith (2012) asked participants to learn word pairs across alternating study and test trials. In elaborative study conditions, participants used an imagery-based keyword method or a verbal elaboration method to encode items during study trials. On a criterial test 1 week after the learning phase, repeated test trials produced better long-term retention than repeated study trials even under the elaborative study conditions. Test trials produced learning by virtue of mechanisms other than elaboration at encoding.

In addition to elaboration at encoding, processing level at encoding was also investigated. Jacoby, Shimizu, Daniels and Rhodes (2005) had participants encoded material under shallow or deep encoding conditions. During a first recognition test, participants discriminated between old words that were studied under either the shallow or the deep conditions and new items (foils). A second recognition test was administered to assess memory for the new items on the first test. Results showed that having taken the first recognition test with the deeply studied items enhanced recognition of new items on the later test than with the shallowly studied items. The manipulation of the processing level at encoding on the first test produced a large effect on recognition of the new items on the later test.

In a separate study, Kuo and Hirshman (1997) manipulated the processing level by asking participants to say aloud a word that was either related in meaning to the initial word (deep processing) or to share the first letter of the initial word (shallow processing). Results showed that the mean proportions of regular words correctly recalled were significantly higher in the deep processing condition than those in the shallow processing condition. Similarly, Morris, Bransford, and Franks (1977) found the same effect of the processing level on memory. They had participants encode words phonemically or semantically, and found that semantic encoding led to greater recognition than phonemic encoding in standard recognition test.

Not only did the processing level at encoding affect retention, the number of test trials at retrieval also affect retention. Roediger and Karpicke (2006a) had participants either study a passage three times and take one test or study a passage once and take three tests. Results showed that those who had one test trial recalled more than those who had three test trials in immediate retention, but the opposite happened in delayed retention. Wheeler and Roediger (1992) also showed that taking three tests immediately after studying a list of pictures greatly improved retention on a final test relative to taking a single test. The higher the number of test trials is, the more robust the testing effect is.

The purpose of the present study was to investigate the effects of test trial and processing level on immediate and delayed retention. Research questions included (a) Was there any difference between single test trial and repeated test trial on immediate and delayed retention? The testing effect expected that single test trial enhanced immediate retention but repeated test trial enhanced delayed retention. (b) Was there any difference between shallow and deep processing on immediate and delayed retention? The level of processing effect expected that deep processing enhanced immediate and delayed retention. (c) Was there any interaction between test trial and processing level on immediate and delayed retention? It was expected that there was an interaction between test trial and processing level on immediate and delayed retention.

Method

Participants

Ninety-one college students were invited to participate in the present study. Data of fifteen participants were discarded because nine of them were over 27 years-old, two of them did not show up for the delayed free-recall test, and four of them did not follow instruction to provide complete data. At the end, seventy-six college students (mean age = 21.3 years old; range = 19 – 27 years old; Male = 8; Female = 68) completed the immediate and delayed tests in partial fulfillment of a psychology course requirement. The procedures met all American Psychological Association (APA) ethical principles for use of human subjects (APA, 2002), and participants were provided informed consent in accordance with guidelines set by the Institutional Review Board of the university.

Materials

Forty stimulus words were taken from the words used by Craik and Tulving (1975, Experiment 9, see Table 1). From the MRC Psycholinguistic Database (Wilson, 1998), several properties of the stimulus words were obtained. The average number of letters was 4.75 (*SD* = .74), the average number of syllables was 1.23 (*SD* = .42), the average printed Kucera- Francis word frequency was 16.3 per million (*SD* = 14.45), the average concreteness rating was 571.91 (*S.D.* = 40.7), and the average familiarity rating was 507.73 (*SD* = 54.06).

Design

A 2 × 2 × 2 mixed ANOVAs was used with two between-subject factors of test trial (single test, repeated test) and processing level (shallow, deep), and one within-subject factor of final recall (immediate, delayed). Participants were randomly assigned to the single test and the repeated test trials. They were then randomly assigned again to the shallow processing level and the deep processing level. Therefore, there were 38 participants in each test trial (single test and repeated test) and each processing level (shallow and deep).

In the single test trial, participants studied the stimulus words three times and took one free-recall test in each cycle. In the repeated test trial, participants studied the stimulus words once and took three consecutive free-recall tests in each cycle. There were three cycles of study/test trials (either SSST or STTT) for 12 trials total. There were nine study and three test trials in the single test trial, and there were three study and nine test trials in the repeated test trial.

In the shallow processing level, participants were asked whether each stimulus word was presented in capital letter or in small letter. In the deep processing level, participants were asked whether each stimulus word belonged to a particular category (see Table 1). The final immediate free-recall test was administered 5 minutes after the 12 study and test trials, whereas the delayed free-recall test was administered one week later.

Procedure

Participants were tested in groups of five or fewer. They were told to study and recall a list of words, and answer some questions to help them remember the words. The task was programmed by E-prime experimental software (Version 1.1; Schneider, Eschman, & Zuccolotto, 2002). Before the word list was presented, participants were given a practice list of two words to familiarize themselves with the task and the presentation rate, and a practice recall test to familiarize themselves with the testing procedure.

There were a learning phase and a testing phase after the practice. The learning phase consisted of 12 study and test trials and took about 30 minutes. At the beginning of each study trial, participants were asked to rest their hands on a key labeled “yes” and the other on a key labeled “no” on the computer keyboard. First, a “Ready” prompt was shown on the computer screen for 1 s. The typescript question or category question was then shown for 1 s, and participants were asked to answer the question by pressing the appropriate key. The typescript question was asked in the form, “Is the word in capital letter?” or “Is the word in small letter?” The category question was asked in the form, “Is the word (a category)?” Both typescript and category questions were counterbalanced, so that half of the answers to the questions was “yes” and half was “no.”

The purpose of the question was to induce the participant to process the word at a relatively shallow level (typescript questions) or at a relatively deep level (category questions). No matter if participants answered the typescript or category questions, stimuli words were presented on a computer at 2 s per word and the screen proceeded to the next word after 2 s. To present 40 stimuli words, the total time for one study trial was 80 s. Participants who were not able to answer the questions correctly over 80% were discarded from the analysis.

The beginning of each test trial was indicated by a tone (presented over headphones for 0.5 s) and a “Recall” prompt that remained on the computer screen throughout the test. During each test trial, participants were given 80 s to write down as many of the words as possible, in any order, on a response booklet. Therefore, the time of exposure to materials on study trials and test trials was equated (both are 80 s). The transition from one test trial to another (in the repeated test condition) was indicated by a tone as well as a change in the background color on the computer screen: The background was blue during the first test, green during the second test, and red during the third test. At the end of each test trial, participants were instructed to turn to the next page on their response booklets and not to look back at any of their previous responses at any time during the learning phase.

After the learning phase of three cycles of 12 study and test trials, participants proceeded to the testing phase and were asked to complete mazes for five minutes. Participants were then given an immediate free-recall test to write down as many of the words as they could recall in 10 minutes, and were instructed to draw a line on their recall sheet to mark their progress at one minute intervals. This procedure ensured that participants had exhausted their knowledge by the end of the 10 minutes recall test and allowed the researcher to measure the number of words recalled.

All participants, except two, returned for the delayed free-recall test one week later. They were given 10 minutes to write down as many of the words as they could recall, and were instructed to draw a line on their recall sheet to mark their progress at one minute intervals. Finally, participants were asked whether they expected to be given a test in the second session and whether they consciously rehearsed the test items after the first session. At the end of the delayed free-recall test, participants were debriefed and thanked for their participation.

Results

The mean number of correct words recalled out of 40 words on the immediate and delayed free-recall tests is presented in Table 2 and Figure 1, as a function of test trial (single test, repeated test) and processing level (shallow, deep). A significance level of .05 is used for all analyses in this study. A 2 test trial (single test vs. repeated test) × 2 processing levels (shallow vs. deep) × 2 final recall (immediate vs. delay) mixed Analysis of Variance (ANOVA) revealed a main effect of final recall, $F(1, 72) = 400.446$, $p = .000$, partial $\eta^2 = .848$. Effect size indicated a high proportion of variance accounted for by final recall. Further pairwise comparisons using a Bonferroni correction showed that the mean number of words recalled in five minutes (23.868) was significantly higher than words recalled in one week (14.395), $p = .000$.

Results showed a main effect of test trial, $F(1, 72) = 13.7$, $p = .000$, partial $\eta^2 = .160$. Effect size revealed low strength in associations. Further pairwise comparisons using a Bonferroni correction showed that the mean number of words recalled in single test trial (21.737) was significantly higher than those recalled in repeated test trial (16.526), $p = .000$. There was also a main effect of processing level, $F(1, 72) = 34.676$, $p = .000$, partial $\eta^2 = .325$. Further pairwise comparisons using a Bonferroni correction showed that the mean number of words recalled in deep processing level (23.276) was significantly higher than those recalled in shallow processing level (14.987), $p = .000$.

There was an interaction between final recall, test trial and processing level, $F(1, 72) = 9347$, $p = .003$, partial $\eta^2 = .115$. A two-way ANOVA was then conducted to investigate the interaction between final recall and test trial. There was an interaction between final recall and test trial, $F(1, 74) = 4.361$, $p = .04$, partial $\eta^2 = .056$. Further analyses of interaction showed that the mean number of words recalled in 5 minutes were significantly higher than those recalled in 1 week at the single test trial (*M*=27.11, *SD*=8.611 vs. *M*=16.37, *SD*=7.134) and repeated test trial (*M*=20.63, *SD*=8.274 vs. *M*=12.42, *SD*=7.417). The mean number of words recalled in 5 minutes at the single test trial (*M*=27.11, *SD*=8.611) were significantly higher than those recalled at the repeated test trial (*M*=20.63, *SD*=8.274). However, there was no difference between the mean number of words recalled in 1 week at the single test trial (*M*=16.37, *SD*=7.134) and those recalled at the repeated test trial (*M*=12.42, *SD*=7.417).

Another two-way ANOVA was then conducted to investigate the interaction between final recall and processing level. There was an interaction between final recall and processing level, $F(1, 74) = 33.003$, $p = .000$, partial $\eta^2 = .308$. Further analyses of interaction showed that the mean number of words recalled in 5 minutes were significantly higher than those recalled in 1 week at the shallow processing level (*M*=18.24, *SD*=7.793 vs. *M*=11.74, *SD*=7.675) and deep processing level (*M*=29.5, *SD*=6.185 vs. *M*=17.05, *SD*=6.363). The mean number of words recalled in deep processing level were significantly higher than those recalled in shallow processing level in 5 minutes (*M*=29.5, *SD*=6.185 vs. *M*=18.24, *SD*=7.793) and in 1 week (*M*=17.05, *SD*=6.363 vs. *M*=11.74, *SD*=7.675).

One more multivariate ANOVA was then conducted to investigate the interaction between test trial and processing level in 5 minutes and I week. No interaction was found between test trial and processing level, $F(1, 72) = .085$, $p = .772$, partial $\eta^2 = .001$ at immediate final recall. However, there was interaction between test trial and processing level, $F(1, 72) = 4.741$, $p = .033$, partial $\eta^2 = .062$ at delayed recall. When participants were involved in repeated test trial, the mean number of words recalled at deep processing level (M=16.74, SD=7.086) was higher than those at shallow processing level (M=8.11, SD=4.852). When participants were involved in shallow processing level, the mean number of words recalled at single test trial (M=15.37, SD=8.348) was higher than those at repeated test trial (M=8.11, SD=4.852).

The free-recall tests in the learning phase were also analyzed to see if results were similar to the free-recall tests in the testing phase (immediate and delay). Another 2 test trial (single test vs. repeated test) × 2 processing levels (shallow vs. deep) × 3 recall (recall 1vs. recall 2 vs. recall 3) mixed Analysis of Variance (ANOVA) was conducted. Similar results were found with main effects in test trial, $F(1, 72) = 28.704$, $p = .000$, partial $\eta^2 = .285$; processing level, $F(1, 72) = 32.054$, $p = .000$, partial $\eta^2 = .308$; and recall, $F(2, 144) = 124.216$, $p = .000$, partial $\eta^2 = .633$. The mean number of words recalled in single test trial (15.588) was significantly higher than those recalled in repeated test trial (10.64); those in deep processing level (15.728) was significantly higher than those in shallow processing level (10.50, those in the third recall test (16.618) was significantly higher than those in the second (13.697) and first recall tests (9.026) in the learning phase. However, no interactions were found between recall and test trial, $F(2, 144) = .362$, $p = .697$, partial $\eta^2 = .005$, recall and processing level, $F(2, 144) = 1.593$, $p = .207$, partial $\eta^2 = .022$; test trial and processing level, $F(1, 72) = 28.704$, $p = .000$, partial $\eta^2 = .285$; and among recall, test trial and processing level, , $F(2, 144) = .159$, $p = .691$, partial $\eta^2 = .002$.

Discussion

The present study investigated the effects of test trial and processing level on immediate and delayed retention. Results showed that single test trial enhanced immediate retention but performed the same as repeated test trial in delayed retention. Deep processing was found to enhance immediate and delayed retention. Findings also showed an interaction between test trial and processing level in delayed retention. When information was processed at a shallow level, single test trial enhanced delayed retention. When information was processed at a deep level, single test trial and repeated test trial performed similarly in delayed retention.

The finding that participants at single test trial recalled more words than repeated test trial in immediate final free-recall test was consistent with previous studies that single test trial produced more short-term benefits than repeated test trial (Roediger & Karpicke, 2006a; Wheeler, Ewers & Buonanno, 2003). Participants in the single test trial were exposed to the words in 9 study trials and those in the repeated test trial were exposed to the words in 3 study trials. The additional exposure to the words may lead to overlearning of the words on immediate test.

The present study failed to find that repeated test trial produced greater benefits on the delayed final free-recall test. The free-recall test contained no cues to assist the student in answering the test and might therefore result in recall of only part of the contents, or a lesser testing effect (Duchastel, 1981). The failure of repeated test trial in producing greater benefits on the delayed recall test might come from the type of retention test adopted in the present study. In addition to the final delayed free-recall test, participants in the repeated test trials did not recall many items at the retention tests at test trials in the learning phase. Testing could be of little help when very few items were successfully retrieved on test trials (Kang, McDermott, & Roediger, 2007). There seemed to be no benefit of repeated test trials because retrieval practice was only beneficial to memory when retrieval was successful. It might be an item selection problem in which participants in the single test trial had much more exposure to the items than those in the repeated test trial.

Since the present study found that single test trial and repeated test trial performed the same in delayed retention, the number of tests trials needed to bring out the testing effect has to be further investigated. It may depend on the material to be studied, the format of test trials and final retention test, the interval between the immediate and delayed retention, etc.

Consistent with the levels of processing effect, participants at the deep processing level performed better than those who were at the shallow processing level in immediate and delayed recall. Craik and Tulving (1975) stated that memory performance depends on the elaborateness of the final encoding, and retention was enhanced when the encoding context was more fully descriptive. The effort participants put forth to differentiate if each stimulus word belonged to a particular category promoted a deep processing of the words whereas the effort to differentiate if the words were presented in capital letters encouraged a shallow processing. With deep processing as a powerful way to enhance immediate and delayed retention, encoding at a deep level should be encouraged.

The interaction of test trial and processing level in delayed free-recall test showed that the memory benefit of effortful retrieval practice at the single test trial was compromised by the levels of processing effect. When participants were asked to process the words at a shallow level, the single test trial enhanced long-term retention. The result was consistent with Karpicke and Roediger's (2007) findings that equally-spaced retrieval practice promoted long-term retention. With a delay of the first test trial, participants had to make a much greater effort to retrieve the information, and the effort involved enhanced long-term retention (delayed free-recall) but not short-term retention (immediate free-recall). However, when participants were asked to process the words at a deep level, the effortful retrieval practice at single test trial performed similarly to the repeated test trials. Lockhart and Craik (1990) explained the superiority of deep processing on memory performance with the idea of robust encoding in which deep processing yielded a trace that was accessible to a broader range of retrieval cues. Such multiple retrieval routes enhanced delayed retention better than effortful retrieval practices.

Conclusion

The deep processing level at encoding seems to mediate the benefits of effortful retrieval practice at test trials in delayed retention. The increase in the elaboration of a memory trace and the number of retrieval routes by retrieval practice promoted retention in shallow processing level, but the accessibility to a broader range of retrieval cues by deep processing level enhanced retention regardless of the effortful retrieval practice.

References

American Psychological Association (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57,* 1060-1073.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123-144). Hillsdale, NJ: Erlbaum.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35-67). Hillsdale, NJ: Erlbaum.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268-294.

Dempster, F. N. (1997). Using tests to promote classroom learning. In R. F. Dillon (Ed.), *Handbook on testing* (pp. 332-346). Westport, CT: Greenwood Press.

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

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin &*

Review, 12, 852-857.

Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 704-719.

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language, 67,* 17-29. Doi:10.1016/j.jml.2012.02.004

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. III (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19,* 528-558.

Kuo, T., & Hirshman, E. (1997). The role of distinctive perceptual information in memory: Studies of the testing effect. *Journal of Memory and Language, 36*, 188-201.

Lockhart, R. S., & Craik, F. I. M. (1990). Levels of processing: A retrospective analysis of a framework for memory research. *Canadian Journal of Psychology, 44*, 87-112.

Morris, C. D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519-533.

Roediger, H. L., III, & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249-255.

Roediger, H. L., III, & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181-210.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools.

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory, 11,* 571-580.

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.

Wilson, M. D. (1998). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers, 20,* 6-11.

Table 1

Stimulus Words and Category Questions

Word	Category Question	Word	Category Question
Bear	a wild animal	Lamp	a type of furniture
Brake	a part of a car	Lane	a type of road
Brush	used for cleaning	Lark	a type of bird
Cart	a type of vehicle	Mast	a part of a ship
Chapel	a type of building	Monk	a type of clergy
Cheek	a part of the body	Nurse	associated with medicine
Cherry	a type of fruit	Pail	a type of container
Clip	a type of office supply	Pond	a body of water
Copper	a type of metal	Rice	a type of grain
Drill	a type of tool	Roach	a type of insect
Earl	a type of nobility	Robber	a type of criminal
Fence	found in the garden	Sheep	a type of farm animal
Fiddle	a musical instrument	Soap	a type of toiletry
Flame	something hot	Sonnet	a written form of art
Flour	used for cooking	Speech	a form of communication
Glove	something to wear	Tire	a round object
Gram	a type of measurement	Tribe	a group of people
Grin	a human expression	Trout	a type of fish
Honey	a type of food	Witch	associated with magic
Juice	a type of beverage	Wool	a type of material

Table 2

Means and Standard Deviations of the Number of Words (Total=40) Recalled in Immediate and Delayed Final Recall by Test Trial and Processing Level (N = 76)

Test Trial	Processing Level	Immediate	Delay	Row Mean
		Mean *SD*	Mean *SD*	Mean *SD*
Single Test	Shallow (n = 19)	21.68 (8.226)	15.37 (8.348)	18.52 (8.287)
	Deep (n = 19)	32.53 (4.765)	17.37 (5.727)	24.95 (5.246)
	Mean	27.11 (8.611)	16.37 (7.134)	21.74 (7.872)
Repeated Test	Shallow (n = 19)	14.79 (5.663)	8.11 (4.852)	11.45 (5.258)
	Deep (n = 19)	26.47 (6.050)	16.74 (7.086)	21.60 (6.568)
	Mean	20.63 (8.274)	12.42 (7.417)	16.52 (7.845)
	Column Mean	23.87 (8.998)	14.39 (7.496)	19.13 (6.339)

Figure 1. Means and Standard Error of the Number of Words (Total = 40) Recalled in Immediate and Delayed Final Recall by Test Trial and Processing Level (N = 76)