

A test of the Battig procedure for controlling the level of individual item learning in paired-associate lists¹

ALEXANDER J. WEARING,² YALE UNIVERSITY, New Haven, Connecticut 06510, and WILLIAM E. MONTAGUE, UNIVERSITY OF ILLINOIS, Urbana, Illinois 61801

Battig (1965) published a correction adjusted-learning method for ensuring equal learning of all items in a paired-associate list. The present experiment tested the adequacy of his technique by using response latency as a measure of learning. The results showed that the Battig procedure was not entirely successful in equating the level of learning across items.

In paired-associate (PA) learning studies, it is often necessary to ensure that all the items in the PA list are learned equally well (Underwood, 1964). Battig (1965) has proposed a technique for achieving this goal that is based on a modification of the study-test procedure.

One begins by showing S both the stimulus and response members of a PA list (a study cycle) and then presenting him with only the stimulus members and asking him to recall the responses (a test cycle). The test cycle is followed by a second study cycle during which *only* those items that were incorrect on the test cycle are presented. After this second study cycle, the items are again tested and the correct items dropped from the succeeding study list. The study-test sequence continues until each item has been responded to correctly, whereupon the trial is concluded. Therefore, at the end of a trial, each pair has been recalled correctly exactly once, and so the memory traces of all the items should be approximately equal in strength. Battig (1965) calls this technique the correction-adjusted learning procedure.

A test of the adequacy of the Battig technique has been made by Montague and Kiess (1966). When Ss learned a list of eight CVC pairs, no reliable differences were found between item-recall frequencies at criterion, and so they concluded that the method achieved its purpose. However, their criterion of six out of eight items correct on an initial test requires that nearly all the items be correct. If trace strength is defined only by the correctness or incorrectness of an item, then their definition of criterion may result in homogeneous item strengths.

A better way of assessing item strength

is to use a measure that varies continuously and is not constrained by the definition of criterion that is used. An appropriate measure is response latency, since it is generally regarded as a sensitive index of trace strength (Leonard & Conrad, 1963; Millward, 1964; Montague & Wearing, 1967b; Osgood, 1953, pp. 325-328; Shapiro, 1968). If the Battig procedure does equate learning across items, then not only should items at criterion be correct (as by definition most of them must be), but their response latencies should be homogeneous. In addition, the number of errors associated with an item prior to criterion should be unrelated to its latency at criterion.

A possible complication of the acquisition process may arise from the learning strategy that S employs. If S encodes a pair with a mediating device, it is possible that response latencies may reflect code complexity rather than trace strength. A perfectly well retained item may be encoded in a complex fashion and so yield a long response latency. The influence of mediating devices or natural language mediators (NLMs) on learning is well established (Adams, 1967; Kiess, 1968; Montague & Wearing, 1967a, c; Wearing, Walker, & Montague, 1967), so consequently Ss were questioned about the learning methods they used.

METHOD

Fifty-nine undergraduates served as Ss in the experiment. Ss learned a 12-item PA list whose stimuli were CVCs of 24%-30% association value (Archer, 1960) and whose responses were high-frequency (AA) words selected from Thorndike and Lorge (1944).

Every item was presented automatically for 5 sec on a modified CRT display in front of each S (Bitzer, Hicks, Johnson, & Lyman, 1967). The procedure developed by Battig (1965) was used. All pairs were presented singly, and then each stimulus was presented in a different order to test recall. Items not correctly recalled were presented and tested again. This procedure was repeated until all items were correctly recalled, which ended a trial. Learning ended when S attained at least 10 out of 12 items correct on the initial test sequences of five trials. Initial test sequences (on which 10 or more items were presented) were termed criterion trials, and the latencies from these trials provided the response-speed data in the present experiment. S typed his responses with the index finger of his preferred hand

on a keyboard directly linked to a computer. Between responses, S rested his hand on the table just in front of the keyset. All responses were automatically timed and recorded. The first key pressed provided the measure of response latency.³

Immediately after attaining criterion, Ss were asked to respond in writing to a questionnaire. Stimuli were shown by themselves on the display, and Ss wrote on a data sheet both the response (if they recalled it) and any NLM that they used in learning the pair.

RESULTS

Because items, once recalled correctly, were virtually always correct thereafter, the number of errors associated with each item before its first correct recall was used to differentiate items in terms of difficulty. Because the number of items recalled incorrectly on the criterion trials was so small (less than 1% of the responses were incorrect), only the latencies of correct responses were considered. Items were also classified according to the method used to learn them: Learning methods were categorized as either rote or NLM, using a classification described elsewhere (Montague, Adams, & Kiess, 1966; Montague & Wearing, 1967a). A consequence of this bivariate classification is that the cases in each cell are only partly independent because different Ss may not contribute the same number of cases to each of the categories, e.g., one S may have eight rote-learned items, another four, and of these, one may be correct for the first S, but four for the second. For this reason, no statistical tests are reported that involve the learning classification.

Pairs learned with an NLM were associated with fewer errors than rote-learned pairs (Adams, 1967; Montague, Adams, & Kiess, 1966; Montague & Wearing, 1967a). Of the pairs with zero errors, 23% were learned by rote and 77% with NLMs, whereas for the pairs with three or more errors, 52% were learned by rote and 48% with NLMs. Whether or not an item was NLM or rote-learned made little difference to its response latency. The respective latencies (in seconds) of NLM and rote items on successive criterion trials were: 1.97 and 2.02, 1.85 and 1.91, 1.74 and 1.76, 1.68 and 1.73, 1.66 and 1.69. The absence of a firm relationship, despite the correlation between NLM use and errors, was because the rote items with zero errors tended to

Table 1
Mean Response Latencies as a Function of Criterion Trial Number and the Number of Errors Made Before the First Correct Response.
The standard error of each mean is shown in parentheses.

Number of Errors	Number of Cases	Criterion Trial Number					Means
		1	2	3	4	5	
0	154	1.90 (.16)	1.69 (.14)	1.65 (.14)	1.55 (.13)	1.53 (.13)	1.64
1	244	1.84 (.12)	1.83 (.12)	1.68 (.11)	1.62 (.11)	1.61 (.11)	1.72
2	132	2.10 (.19)	1.90 (.17)	1.86 (.17)	1.75 (.16)	1.73 (.15)	1.87
3	81	2.26 (.27)	2.01 (.23)	1.80 (.21)	1.76 (.20)	1.71 (.19)	1.91
4	97	2.15 (.23)	2.12 (.22)	1.90 (.20)	2.06 (.21)	1.95 (.20)	2.03
Means		1.99	1.87	1.75	1.70	1.67	

have slightly shorter latencies than NLM items in the same category.

Mean latencies (in seconds) declined during successive criterion trials: 1.99, 1.87, 1.75, 1.70, and 1.67. In virtually every case, a S's five criterion trials were successive. Mean latencies of responses classified by the number of errors made prior to the first correct recall are shown in Table 1 for all five criterion trials. The standard errors of the means are also shown. Although the individual scores are highly variable, there is a consistent tendency for items with few errors to have shorter mean latencies than do items with many errors. If the number of errors were unrelated to latency, then the probability of any item with X_1 preceding errors having a shorter mean latency than any other item with X_2 preceding errors on a particular criterion trial would be one-half. On each trial, the mean latencies of the groups of items with 0, 1, 2, 3, or 4 errors were compared with one another to see if latency depended on the number of preceding errors. In fact, of the 50 possible pairwise comparisons from Table 1 (10 on each trial), there are only 4 in which items with the greater number of errors have shorter latencies. By the normal approximation to the binomial, this distribution is highly unlikely under the hypothesis that $p = \frac{1}{2}$ ($Z > 5.93, < .0001$).

DISCUSSION

If the Battig procedure is effective, then the number of errors associated with an item before criterion is attained should be unrelated to response latency at criterion. That was not the case in these data. Although the differences were not great, there was a significant tendency for items with more precriterion errors to have longer latencies, and these differences persisted throughout the criterion trials. These results suggest that, contrary to Battig's (1965, p. 6) suggestion, a difficult item remains difficult, even after considerable practice. It could be that the

Battig procedure is able to equalize trace strength for items of approximately the same level of difficulty, but it would be unwise to rely on it to handle lists that contain items that vary much in difficulty. There may be differences in learnability between items that are not amenable to short-term practice.

The learning method used did not complicate the picture, since response latency did not seem reliably affected by whether the item was learned by rote or an NLM. This finding lends support to the view that the effect of a mediating device like an NLM is to increase associative strength in the form of an S-R bond rather than by providing an alternative code. The fact that rote latencies were a little shorter than the NLM latencies in the zero-error category contradicts this assertion, but the differences were small, being merely enough to leave the suspicion that perhaps, after all, NLM and rote learning do not operate in quite the same way.

However, for practical purposes, one may conclude that the Battig technique is relatively invulnerable to differences in learning method, but it does require that the distribution of errors before the first correct recall be small.

REFERENCES

- ADAMS, J. A. *Human memory*. New York: McGraw-Hill, 1967.
- ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, 74(10, Whole No. 497).
- BATTIG, W. F. Procedural problems in paired-associate learning research. *Psychonomic Science Monograph Supplements*, 1965, 1(1, Whole No. 1).
- BITZER, D. L., HICKS, B., JOHNSON, R., & LYMAN, E. R. The PLATO system: Current research and development. *IEEE Transactions on Human Factors in Electronics*, 1967, 8, 64-70.
- KIESS, H. O. Effects of natural language mediators on short-term memory. *Journal of Experimental Psychology*, 1968, 77, 7-13.
- LEONARD, J. A., & CONRAD, R. Maintenance of high accuracy without feedback. *Nature*, 1963, 199, 512-513.

MILLWARD, R. Latency in a modified paired associate learning experiment. *Journal of Verbal Learning & Verbal Behavior*, 1964, 3, 309-316.

MONTAGUE, W. E., ADAMS, J. A., & KIESS, H. O. Forgetting and natural language mediation. *Journal of Experimental Psychology*, 1966, 72, 829-833.

MONTAGUE, W. E., & KIESS, H. O. Test of a procedure to control inter-pair learning and the effect of repeated, post-criterion test trials on retention. *Psychonomic Science*, 1966, 6, 191-192.

MONTAGUE, W. E., & WEARING, A. J. The complexity of natural language mediators and its relation to learning. *Psychonomic Science*, 1967a, 7, 135-136.

MONTAGUE, W. E., & WEARING, A. J. The retention of responses to individual stimuli and stimulus classes. *Psychonomic Science*, 1967b, 9, 81-82.

MONTAGUE, W. E., & WEARING, A. J. Natural language mediation: A source of interference with extra-experimental interference. *Psychonomic Science*, 1967c, 9, 317-318.

OSGOOD, C. E. *Theory and method in experimental psychology*. New York: Oxford University Press, 1953.

SHAPIRO, S. I. Paired-associate response latencies as a function of free association strength. *Journal of Experimental Psychology*, 1968, 77, 223-231.

THORNDIKE, E. L., & LORGE, I. *The Teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

UNDERWOOD, B. J. Degree of learning and the measurement of forgetting. *Journal of Verbal Learning & Verbal Behavior*, 1964, 3, 112-129.

WEARING, A. J., WALKER, C. B., & MONTAGUE, W. E. Recall of paired associates as a function of their associability. *Psychonomic Science*, 1967, 9, 533-534.

NOTES

1. This research was supported in part by the Joint Services Electronics Program under Contract No. DA 28 043 AMC 00073 (E) and in part by the Computer-based Educational Research Laboratory at the University of Illinois.

2. The authors are grateful to Rosemary Wearing and Maryann Carp for assistance in carrying out this research.

3. In an unpublished study on this apparatus, the authors found that when Ss had to use only their index fingers, latency of the first key press was a function of word length but not of the first letter.