THEORETICAL AND EMPIRICAL PERFORMANCE DIFFERENCES BETWEEN ANTICIPATION AND STUDY-TEST METHODS: RETENTION INTERVAL HYPOTHESIS VS. SPACED PRACTICE EFFECTS¹

CHIZUKO IZAWA

Tulane University

The retention interval has become a critical factor in explaining inconsistent results in relative differences between anticipation and study-test methods. The retention interval in paired-associate (and verbal discrimination) learning consists of two qualitatively different components: Types I (rest, intercycle) and II (work, intervening events) intervals. To determine relative effects of the two components, six sets of comparisons were made of the two methods; massed practice (omitting Type I intervals) was employed with various paired-associate lists and presentation rates, using 140 college students. In the absence of the Type I interval, both significant and nonsignificant performance differences between the two methods occurred. Thus, performance inconsistency seems not easily explained by Type I interval effects, as e.g., done in the spaced practice view. Specifically, the results from five of the six experiments contradicted the spaced practice view. By contrast, the retention interval hypothesis, based on Type II interval effects, is supported more strongly—and therefore suggests developing a better theory.

Given the long-standing contradiction that the study(reinforcement)-test (RT) method produces superior performance at times, but equivalent performance at other times when compared to the anticipation method in verbal learning; any factor, e.g., such as immediate reinforcement or task separation, favoring one method or the other cannot constitute an adequate

explanation. A satisfactory theory must reconcile contradictory results. To this end, Izawa (1972) proposed the retention interval hypothesis on the basis of differential and equal intervening events in retention intervals between anticipation and study-test methods. However, Battig (1973) argued that this inconsistency could be better resolved by Izawa's (1971a) findings which involved the differential spaced practice effects between the two methods. His affinity for this view will henceforth be referred to as the spaced practice view.

Consider, for instance, a 4-pair list given under the anticipation method, as in the top panel of Fig. 1. During an A-presentation, the subject is expected to respond with the corresponding B term. These test (T) events are illustrated each by a square with broken lines (Fig. 1). Each T event is immediately followed by

¹ The present study was supported in part by a 1974 Faculty Fellowship from Tulane University. Portions of the research described here were presented at the Psychonomic Society Meeting in Boston, November, 1974. I am obligated to Steve Meyer, James DeVoll, Hurst Hessy, Mimi Vollstedt, Jesse Gonzales, and Gregory Boertje for their faithful assistance in collecting and/or tabulating data. Mark Lampert and Sydney Miller assisted in the preparation of figures.

Requests for reprints should be sent to Chizuko Izawa, Department of Psychology, Tulane University, New Orleans, Louisiana, 70118 U.S.A.

FIG. 1. An illustration for determining the number of intervening events (Type II intervals) of the retention intervals with a 4-pair list (list length $n=4$), under the anticipation method (top) and the RT (study-test) method (bottom). [Each number in parenthesis between a reinforcement (study) event (A_j-B_j presentation) and its test event (A_j -presentation) for a target Pair j indicates the number of intervening events one by one, starting from the first intervening event immediately following the study event.]

the presentation of both A and B terms, i.e., a reinforcement (R) event during which the subject may study the item (indicated by a square with solid lines in Fig. 1). Under the study-test (RT) method, however, the R and T events occur in separate cycles as seen in the bottom panel of Fig. 1. A cycle is one run through the list of n pairs. Izawa (1972) found that a critical factor differentiating the two methods is the retention interval of individual items, which is definable as the interval between a study (R) event of an item and its subsequent T event.

The retention interval is composed of two qualitatively different and operationally distinguishable components. One component, which henceforth is called a Type I interval, is an intercycle interval which is unfilled with paired-associate items and thus can be regarded as a rest period (a shaded bar in Fig. 1). The other component, a Type II interval, is the remainder of the retention interval; it is filled with other R and T events and is

a work period. Whereas rest intervals are likely to generate facilitating or positive effects, work intervals are likely to generate interfering or negative effects.² The lengths of Type II intervals differ greatly depending on the method employed. Despite random item presentation, the following regularity emerges (Fig. 1; Izawa, 1972): For a 4-pair list ($n=4$), under the anticipation method, the longest Type II interval has 12 intervening events ($4n-4$), the shortest, 0, with a mean of 6 events ($2n-2$). Under the study-test method, however, the longest Type II interval involves only 6 events ($2n-2$), the shortest, 0, with a mean of 3 intervening events ($n-1$).

² Daily experience tells us that hard work (i.e., Type II interval) and rest (i.e., Type I interval) produce different psychological effects. Hull, for example, postulated that during work, performance decrements occur because of reactive inhibition, whereas rest dissipates such inhibition and performance improves. In paired-associate learning situations, too, work and rest generate different effects as seen in Izawa (1968).

The mean absolute retention (R-T) intervals for anticipation and study-test (RT) methods are, respectively, as follows (Izawa, 1972, p. 18) in Equations 1 and 2:

$$(\text{Mean Abs. R-T Interval})_{\text{ant}} = 2(n-1)d + (\text{Intercycle Interval}) \quad (1)$$

$$(\text{Mean Abs. R-T Interval})_{\text{RT}} = (n-1)d + (\text{Intercycle Interval}) \quad (2)$$

$$\begin{array}{ccc} \vdots & & \vdots \\ (\text{Type II}) & & (\text{Type I}) \end{array}$$

in which d is duration of exposure. The first terms correspond to Type II, and the second, to Type I intervals. Thus, the mean Type II interval for the anticipation method is twice as long as that for the study-test method. This may be disadvantageous for the anticipation method (e.g., Tulving & Arbuckle, 1963; Bower, 1967; Izawa, 1972). In contrast, the length of the Type I interval does not differentiate the two methods since it usually is the same for both.

Because items are commonly presented randomly from cycle to cycle, the length of the Type II interval for a given item varies. The theoretical distribution curves show a relative advantage favoring the study-test method (Izawa, 1972, Fig. 1). However, since the two methods share the same lower limit, but have different upper limits, the distributions overlap. The overlap indicates that the probability that an item will have the same Type II interval under both methods is high. In this case, the retention interval hypothesis predicts equal performances for anticipation and study-test methods. However, without an overlap, the study-test method will average a shorter Type II interval: The assumption is that a shorter Type II interval permits less short-term forgetting, therefore, the study-test method should be superior to the anticipation method. Consequently, the retention interval hypothesis is better able to account for both equal and superior performances when the study-test method and the anticipa-

tion method are compared.

The foregoing rationale incorporates the effects of different item-types. Items may be of two kinds: (1) Critical items which are in the short-term memory (STM) at the end of a given R event, and (2) noncritical items which are not in STM. Noncritical items can be further subdivided into: (a) Unlearned items and (b) *well-learned* items that are in *stable* long-term memory (LTM). The noncritical items produce the same results for both methods; i.e., unlearned items will be incorrect for both methods and *well-learned* items will be correct for both methods. Thus, noncritical items do not differentiate between the methods.

In contrast, the critical items that enter STM on a study event are likely to survive only during a short Type II interval, but not during a long one. Under certain circumstances all n items can be critical, but generally the number of critical items is likely to be less than n . With respect to item distributions (Type II intervals, Izawa, 1972, Fig. 1) critical items are on the average more likely to fall into the non-overlapping area (always larger than the overlapping area) under each method. If so, and when critical items are sufficient in number, performance differences between the methods are expected to be large. However, there *also* exists a substantial possibility that more of the critical items may fall into the overlap area; thus, the two methods are likely to produce the same performance levels. It should also be noted that: When the number of critical items is very small, performance differences between the two methods are likely to be small no matter whether or not these critical items fall into the overlap area.

Given the fact that the retention interval consists of two qualitatively different components which may produce psychologically opposite effects, separate research on each component has been conducted. Izawa (1971a) investigated ef-

the study-test method. The findings have profound impact upon how a theory accounts for relative performances in the two methods. For one possibility, see the left panel of Fig. 3: Spaced practice effects are not significant for the anticipation method, but significant for the study-test method. Consequently, performance differences (shown as the vertical distance) between the two curves are small with massed practice, but large with spaced practice. This would imply that both significant and nonsignificant differences between the methods might be explained by the lengths of Type I interval (the spaced practice view): Such an explanation was supported in some conditions by Izawa (1971a), and was also suggested by Battig (1973). However, in order for the spaced practice view to account fully for both superior and equal performances when the study-test method is compared with the anticipation method, it is necessary to show that: (a) Massed practice produces no significant performance differences, and (b) spaced practice does give rise to significant performance differences when the methods are compared. If these two propositions are not satisfied, i.e., if both significant and nonsignificant differences occur with either massed, or spaced practice, or with both; then, spaced practice effects can not serve as the primary explanation for the performance differences between the methods.

Inferences from Izawa's 1971a data

FIG. 3. Predictive inferences possible from Izawa's results (1971a) concerning performance differences (indicated by vertical distance between the two lines) between anticipation and study-test (RT) methods.

need not be limited to Battig's view (cf. Predictive Inference I, Fig. 3). There exist at least two other possibilities. One is Predictive Inference II, i.e., within the anticipation method, spaced practice effects are too small to be significant, but are large enough to reduce differences between the two methods to a nonsignificant level.

In fact, large performance differences between the two methods with spaced practice using 30 sec Type I intervals are by no means universal (e.g., Kausler, 1963; Izawa, 1974b), contrary to Battig's suggestion. These two studies support the possibility that there are neither large performance differences with massed nor with spaced practice (Fig. 3, center: Predictive Inference II). Therefore, the two necessary propositions above have not been satisfied.

Furthermore, there is a third possibility: As seen in Predictive Inference III (Fig. 3), spaced practice effects are significant within the study-test method but not significant within the anticipation method; but performance differences *between* the two methods are significant not only with spaced practice but also with massed practice. In further support of this third possibility, significant differences do occur with massed practice (Cavanagh & Parkman, 1971; Izawa, 1971a, 1972, 1974a; Wright, 1967). In Predictive Inference III situations, performance differences between the methods are significant at all degrees of spacing providing supplementary evidence again negating the two propositions necessary for a rigorous spaced practice theory.

The fact that both significant and nonsignificant differences coexist with massed as well as with spaced practice indicates that the spaced practice view cannot fully explain the inconsistencies in extant data. Is a more adequate explanation possible?

Under the traditional list design, it is impossible to isolate the effects of the Type I interval from those of the Type II inter-

val, since they are always confounded with each other. In contrast, it is possible to isolate the effects of Type II intervals from those of Type I intervals, since the latter can be completely eliminated when 0 sec intercycle intervals are used. This fact is important. According to the spaced practice view, no performance differences should occur with massed practice, but this is confirmed neither by the studies cited above, nor by Izawa's experiments (1972): Highly significant differences did occur with massed practice. Since the retention interval with massed practice included no (or only brief) Type I intervals, one may conclude that the Type II interval controls most performance differences between anticipation and study-test methods; and this is entirely consonant with the retention interval hypothesis.

EXPERIMENT

Empirical evidence must take precedence over any speculative or a priori arguments in theory construction. Thus, the sole effects of Type II intervals, in the absence of Type I intervals, are examined in order to assess the relative strengths of the two types of intervals in controlling performance differences between anticipation and study-test methods.

Method. Six comparisons (i.e., Sets A, B, C, D, E, and F) were made between the two methods by completely eliminating Type I (intercycle rest) intervals (i.e., 0 sec Type I intervals). 12 CVC-CVC pair lists were each learned in Sets A, B, and C, 15 CVC-two-digit pairs in Set D, and 20 CVC-two-digit pair lists in Sets E and F, respectively. CVCs and two-digits in the first 5 sets were of high to intermediate difficulty (Noble's m' range: 0-1.69), those in Set F were easier (m' : 2.43-2.81); each list was low in intralist similarity. Exposure rates, on a Stowe memory drum B-549, for both reinforcement (R, study) and test (T) events were: 2 sec for Set D and 3 sec for Sets A, B, E, and F, and 6 sec for Set C. These lists and rates were used to take into account varying situations employed in other studies.

To be comparable to the first anticipation cycle, the first test cycle was given prior to the first R cycle under the study-test method. Acquisition was continued until the 10th R and the subsequent T were reached. All subjects had a practice task with two lists: One with three pairs under the study-test method, and the other with three pairs under the anticipation method. A semi-Latin Square design was employed to assign subjects to conditions in the order of appearance. Introductory Psychology students, 140 in all, participated for course credit: 20 each in Sets A to D (10 per method), and 30 each in Sets E and F (15 per method). No ability differences between the two groups within each set were noted on a practice task.

Results and Discussion

Figure 4 gives results in terms of incorrect responses. Tested by the total errors per subject, the study-test method produced significantly better performances than the anticipation method in five of the

FIG. 4. Six comparisons of anticipation and study-test (RT) methods with no Type I (intercycle) intervals (maximally massed practice).

six comparisons: F 's(1, 18) were 4.41, 5.28, 5.90, and 6.47 for Sets A, B, C, and D, respectively, and $F(1, 28)=11.62$ for Set F, each $p \leq .05$. However, the two methods did not differ significantly in Set E: $F(1, 28)=2.19$, $p > .10$.

Both presence and absence of differences between anticipation and study-test methods may be logically explained by the retention interval hypothesis. In contrast, the spaced practice view predicts that with maximally massed practice there should be no difference between the two methods. Set E results as well as those of Bruder (1969) and Izawa (1974a, extremely long and short lists) supported the latter view. However, the results of the other five sets together with Cavanagh and Parkman (1971), Izawa (1972) and Wright (1967, unmixed 1st list) all contradicted spaced practice expectations.

Since the performance differences between anticipation and study-test methods are sometimes reliable and sometimes not in the absence of the Type I interval, attempts to explain the contradictory results by the Type I interval alone would seem to be unsupportable.

The present findings and conclusion, while contrary to Battig's, need not be viewed as surprising: Investigations of performance differences between the two methods involving long and short Type II intervals, respectively, but using the constant Type I intervals address an issue differing qualitatively and fundamentally from investigations of massed and spaced practice within each method employing different Type I intervals but the same Type II intervals (cf. Fig. 2).

The above conclusion, however, does not mean that the Type II interval alone can account for all aspects of performance differences between the two methods, nor that the Type I interval produces no effects at all. We address here the reasons for the relative dominance of the Type II interval over the Type I interval. The controlling effects of the Type II interval

may be more or less safely assumed provided that intercycle intervals are either zero or brief, i.e., in the order of a few seconds, as is often the case in common verbal learning experiments (see also Izawa, 1972, 1974a).

The assumption, however, becomes unsatisfactory when the Type I interval is considerably longer, i.e., with spaced practice. It is indeed true that relative performance differences between the two methods are controlled to a major extent by the Type II interval, supporting the retention interval hypothesis (spaced practice also produces both significant and nonsignificant differences, Fig. 3). The hypothesis, therefore, seems to explain contradictory results both for massed and spaced practice. However, the absolute performance level for each method is controlled by the Type I interval, in agreement with Izawa (1971a). Relative performance differences and absolute performance levels are related, but not identical issues. A weakness of the retention interval hypothesis with spaced practice lies in its inability to account for the absolute performance levels of the two methods.

Fortunately, results that cannot be explained by the retention interval hypothesis can be accounted for by the test trial potentiating model (Izawa, 1971b) which can explicate spaced practice (Type I interval) effects as seen in Izawa (1971a).

Then, the foregoing suggests that our theoretical endeavor may yield fruitful results by combining the retention interval hypothesis, that accounts for Type II interval effects, and the test trial potentiating model, that accounts for Type I interval effects. A new theory which synthesizes both theories within one theoretical framework may be able to explain not only relative (significant or insignificant) performance differences between anticipation and study-test methods, but may also explain the levels of absolute performance under each method. Thus a new, as yet

unnamed theory seems almost within our grasp.

REFERENCES

- BATTIG, W. F. 1973 Evidence that recall-anticipation differences increase with increasing time intervals. *Journal of Experimental Psychology*, **100**, 399-402.
- BOWER, G. H. 1967 A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*, Vol. I. New York: Academic Press, Pp. 229-325.
- BRUDER, G. A. 1969 Effects of method of presentation on paired-associate learning. *Journal of Experimental Psychology*, **79**, 383-384.
- CAVANAGH, J. P., & PARKMAN, J. 1971 A comparison of the study-recall and anticipation methods in steady state paired-associate learning. *Psychonomic Science*, **22**, 361-363.
- IZAWA, C. 1968 Effects of reinforcement, neutral and test trials upon paired-associate acquisition and retention. *Psychological Reports*, **23**, 947-959.
- IZAWA, C. 1971 a Massed and spaced practice in paired-associate learning: List versus item distributions. *Journal of Experimental Psychology*, **89**, 10-21.
- IZAWA, C. 1971 b The test trial potentiating model. *Journal of Mathematical Psychology*, **8**, 200-224.
- IZAWA, C. 1972 Retention interval hypothesis and evidence for its basic assumptions. *Journal of Experimental Psychology*, **96**, 17-24.
- IZAWA, C. 1974 a Retention interval hypothesis and list lengths: Comparison of anticipation and study-test procedures. *Canadian Journal of Psychology*, **28**, 214-224.
- IZAWA, C. 1974 b Performance differences and distributed practice between anticipation and reinforcement(study)-test methods. Paper presented at the meeting of the Psychonomic Society, Boston, November. (A list of 15 pairs was learned by 20 subjects, with 30 sec intercycle intervals. Differences between the two methods were nonsignificant: $F < 1$.)
- KAUSLER, D. H. 1963 Comparison of anticipation and recall methods for geriatric subjects. *Psychological Reports*, **13**, 702.
- TULVING, E., & ARBUCKLE, T. Y. 1963 Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and Verbal Behavior*, **1**, 321-334.
- WRIGHT, J. H. 1967 Effects of stimulus meaningfulness, method of presentation, and list design on the learning of paired associates. *Journal of Experimental Psychology*, **73**, 72-77.

(Received Dec. 13, 1976)