The retention interval model: Qualitative and quantitative examinations¹

CHIZUKO IZAWA²

Department of Psychology, Tulane University, New Orleans, LA 70118 U.S.A.

The retention interval model was formulated to account for variable performance differences between anticipation and study-test item presentation methods on the basis of differential and equal individual retention intervals, consisting of the intercycle (rest) interval and intervening study and test events of other items (work intervals). An experiment was conducted under both methods with massed and spaced practice. The study-test method outperformed the anticipation method significantly with spaced practice, but nonsignificantly with massed practice, in line with the model. Furthermore, numerical predictions from the model satisfactorily accounted for data in spite of three highly restrictive modes of parameter estimations utilized. The present findings suggest that the two item presentation methods may produce nearly the same acquisitions, but differ mainly in their retention processes attributable to the differential and equal distributions of the retention intervals.

Key words: theories, mathematical models, learning methods, item presentation procedures, acquisition vs. retention, parameter estimation modes, massed vs. spaced practice.

Retention processes occurring during the acquisition phase were saliently put into the foreground by the retention interval model (Izawa, 1972, 1981) in investigating varied performance differences between anticipation and study-test (reinforcement-test, RT) methods commonly used in paired-associate learning (PAL) and verbal discrimination learning (VDL). Two cycles of the PAL anticipation method, with random presentation order for an n -item list, may be illustrated as:

$$\left. \begin{array}{l} A_1- \quad , A_1-B_1, A_2- \quad , A_2-B_2, \dots, \\ A_j- \quad , \boxed{A_j-B_j}, \dots, A_n- \quad , A_n-B_n; \\ \text{(Intercycle Interval);} \\ A_3- \quad , A_3-B_3, A_7- \quad , A_7-B_7, \dots, \\ \boxed{A_j- \quad}, \boxed{A_j-B_j}, \dots, A_4- \quad , A_4-B_4; \end{array} \right\} (1)$$

in which j indicates an individual item ($1 \leq j \leq n$). A test event of Item j is enclosed by broken lines, while its study event is shown by solid lines. Here, a test ($A_j- \quad$) is immediately followed by its study event (feedback, A_j-B_j), and therefore this particular method was considered to be advantageous. The principle was indeed promoted by the teaching machine and operant conditioning.

In contrast, under the comparable study-test method, study and test events are separated and administered on alternate cycles as in Equation 2:

$$\left. \begin{array}{l} A_1-B_1, A_2-B_2, \dots, \boxed{A_j-B_j}, \dots, \\ A_n-B_n; \text{(Intercycle Interval);} \\ A_3- \quad , A_7- \quad , \dots, \boxed{A_j- \quad}, \dots, \\ A_4- \quad ; \text{(Intercycle Interval);} \end{array} \right\} (2)$$

¹ The present research was supported, in part, by BRSR Grant SO7 PR07040 awarded by the Biomedical Research Support Grant Program, Division of Research Resources, National Institutes of Health of the United States.

² The author is indebted to M. A. Kelley who wrote computer programs used in the present research, and also gratefully acknowledges services rendered by Jesse Gonzales who collected data, and by Elizabeth Radaj and Nelda Clements who provided general research and clerical assistance, respectively. Without their assistance, this study would have been delayed to a considerable extent.

$$\left. \begin{array}{l} A_6-B_6, \boxed{A_7-B_7}, \dots, A_n-B_n, \dots, \\ A_2-B_2; \text{ (Intercycle Interval); } \\ A_5- \quad, A_1- \quad, \dots, A_7- \quad, \dots, \\ A_j- \quad. \end{array} \right\}$$

Quite mysteriously, the study-test method with no immediate feedback often produced significantly better performances. To make matters worse, findings were not consistent, and the two methods frequently did not differ much. These contradictory findings challenged investigators for decades. For details see Izawa (1972, 1974, 1977), Izawa, Hayden, and Isham (1980), and Kanak, Cole, and Eckert (1972).

The solution to this puzzle began to unfold, however, with the discovery that the critical factor controlling performance differences between the two methods is the retention interval between a study (S, enclosed with solid lines in Eqs. 1 & 2) event of an item and its subsequent test (T, enclosed with broken lines) event. Furthermore, the work (Type II) interval component of the retention interval distributes triangularly with the same lower limits (0 intervening event) but different upper limits ($2n-2$ vs. $4n-4$) for the two methods (Izawa, 1972, 1977).

Thus, the two distribution curves necessarily overlap. If a sufficient number of critical items that reside in short-term memory (STM) falls into the nonoverlapping area, these critical items are likely to survive during the short retention interval as under the study-test method, but are unlikely to do so during the long one as under the anticipation method, thus resulting in a large superiority for the former. However, if a sufficient number of critical items falls into the overlap area, their retention intervals are likely to be about the same for both methods, and therefore, likely to result in about the same levels of performance for both. Similar nonsignificance between the procedures is expected when only a very small number of critical items is generated.

The retention interval hypothesis based

on the above rationale does indeed account for both significant and nonsignificant supremacy of the study-test method vis-à-vis the anticipation method within a single theoretical framework. The hypothesis has, nonetheless, a drawback; it postulates nothing for the intercycle (rest) interval. This cripples our theoretical efforts when we elevate the theoretical level sufficiently to be able to make quantitative predictions.

In accordance with the suggestion by Izawa (1977), a combination of the retention interval hypothesis that accounts for the retention phenomena and the test trial potentiating model (Izawa, 1971b) that accommodates acquisition phenomena via intercycle interval effects was achieved, resulting in the retention interval model (Izawa, 1981). The latter model absorbed all postulates of the test trial potentiating model, with an extensive modification of the inter-presentation interval. In the new model, that interval is subdivided into three operationally distinguishable and theoretically identifiable components: intercycle (Type I), intervening-study-events (a kind of Type II), and intervening-test-events (the other kind of Type II) intervals.

The present study evaluates the retention interval model from various perspectives. Given the importance of the intercycle interval, it seems necessary to have comparisons of the two methods with both massed and spaced practice within a single experiment. Unfortunately, such studies are so scarce that we conducted an experiment suitable to the present purposes.

Experiment

Method

A 2×2 factorial design was used with two methods (study-test vs. anticipation) and two intercycle intervals (0 s for maximal massed practice vs. 30 s for nearly optimal spaced practice for certain situations; cf. Battig, 1973). Conditions 1, 2, 3, and 4 identify: the study-test method

with massed practice, the anticipation method with massed practice, the study-test method with spaced practice, and the anticipation method with spaced practice, respectively.

The list was composed of 20 Consonant-Vowel-Consonant (CVC)-two-digit pairs, with CVCs' meaningfulness (m' , Noble, 1961) ranging from 1.50 to 1.69, and digits' association values (a' , Battig & Spera, 1962) from 0.72 to 2.25. On a Stowe memory drum B-549, each item was presented at the 3 s rate for both study (S) and test (T) events. The item presentation order was randomized from cycle to cycle. To equate test experiences for both methods, a test cycle preceded the first study cycle for the study-test method.

Sixty Tulane University undergraduates enrolled in Introductory Psychology participated, 15 per condition. All participants had a practice task with two short lists with 3 pairs each, one with the study-test method, and the other with the anticipation method. Using a semi-Latin-Square design, the subjects were assigned to their conditions in order of their appearance. Tested by their performance on the

practice task, the four groups did not differ significantly in the subject variables.

Results and Qualitative Discussion

Figure 1 presents data as a function of preceding study trials in terms of proportion of incorrect responses. As expected from the retention interval model (cf. Iizawa, 1977), the study-test method was superior to the anticipation method in both cases, but the extent of that superiority was smaller with massed practice (0 s intercycle intervals) than spaced practice (30 s intercycle intervals). An overall 2×2 analysis of variance in terms of total errors per subject showed an overall significant superiority of the study-test method, $F(1, 56) = 8.92$, $MS_e = 1308.78$, $p < .01$. It is also interesting to observe significant differences between spacing arrangements, $F(1, 56) = 4.60$, $MS_e = 1308.78$, $p < .05$. Because of the unusually large advantage of spaced practice effects with the anticipation method, interactions (method $\times$ spacing arrangement) were nonsignificant.

To procure more fine-grained information, performance differences between anticipation and study-test techniques are analyzed separately for massed and spaced practice: differences were nonsignificant with massed practice, $F(1, 28) = 2.19$, $MS_e = 1370.46$, $p > .10$; but significant with spaced practice, $F(1, 28) = 7.71$, $MS_e = 1247.11$, $p < .01$. (Apparently, a major contribution to the significance of the method factor in the overall analyses came from spaced practice data.)

To check acquisition gains per study trial, the proportion correct for the first time given incorrect on all previous trials was obtained and entered in Table 1. This statistic may be interpreted as a conditioning probability newly occurring on every study trial. Corresponding to the already known phenomenon, the conditioning probability in all conditions increased over the initial one to two study trials, and then more or less stabilized. Seen in the last column, the weighted

Fig. 1. Performance differences between study-test (round dots) and anticipation (square dots) methods with massed practice (0 s intercycle intervals) and with spaced practice (30 s intercycle intervals).

Table 1

Proportion of items correct for the first time given incorrect on all previous tests
(conditioning probability) for each study trial in Conditions 1-4

	Preceding study (S) trials											Weighted Mean
	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈	S ₉	S ₁₀	S ₁₁	
Massed practice												
Study-test	.050	.123	.132	.129	.164	.165	.174	.138	.117	.145	.085	.124
N*	300	285	250	217	189	158	132	109	94	83	71	
Anticipation	.047	.063	.119	.114	.120	.120	.160	.154	.148	.245	.095	.113
N	300	286	268	236	209	184	162	136	115	98	74	
Spaced practice												
Study-test	.087	.124	.200	.182	.268	.243	.241	.303	.261	.206	.333	.183
N	300	274	240	192	157	115	87	66	46	34	27	
Anticipation	.077	.094	.124	.155	.161	.179	.133	.189	.211	.197	.158	.137
N	300	277	251	220	186	156	128	111	90	71	57	

* N: Number of relevant cases.

Table 2
Overt errors per condition

	Massed practice		Spaced practice	
	Study-test	Anticipation	Study-test	Anticipation
Total overt errors	682	783	760	837
Mean / Item	2.27	2.61	2.53	2.79
S.D.	2.16	2.17	2.51	2.63
Total all errors	2 262	2 562	1 843	2 380
Mean / Item	7.54	8.54	6.14	7.93
S.D.	3.57	3.06	3.28	3.19
Overt errors / All errors	0.302	0.306	0.412	0.352

means for the study-test method were greater than those of the anticipation method within each practice arrangement. Similarly, the same statistic within the same method was larger with spaced practice than with massed practice.

It is particularly interesting to note that the conditioning probability increased by 47.6% under the study-test method from massed practice to spaced practice (.124 vs. .183), while the increment of the same statistics under the anticipation method was limited to only 21.2% (.113 vs. .137). In retrospect, it was indeed small wonder why so many studies failed to show positive distributed practice for decades when the

anticipation method was used (cf. Underwood & Ekstrand, 1967), while positive effects were easily demonstrable when the study-test method was used (Izawa, 1971a).

Overt errors were substantial in the present study. Apparently guesses were relatively easy, since response terms were all numbers. Total overt errors, along with the mean overt errors per item and standard deviations are entered in Table 2. The high guessing rate, however, was universal in all present experimental variables and did not differ significantly by spacing or method factor; interactions (spacing $\times$ method) were also nonsignificant.

Quantitative Tests of the Retention Interval Model

The retention interval model (Izawa, 1981) has 11 postulates, and for the purpose of testing it quantitatively, it suffices here to summarize the model's mathematical formulation. The model assumes that any stimulus element in the universe belongs, at any given time, to one of the following four mutually exclusive and exhaustive states:

CA (State 1):	Conditioned (C) and active (A , in the available set),
$C\bar{A}$ (State 2):	Conditioned and inactive ($\bar{A}$, in the unavailable set),
UA (State 3):	Unconditioned (U) and active, and
$U\bar{A}$ (State 4):	Unconditioned and inactive.

In the present learning situation, there are five operationally distinguishable and theoretically identifiable events in reference to the target item: study (S) and test (T) events; intervening-study-events (t_s) and intervening-test events (t_t) intervals; and the intercycle rest (t_i) interval.

According to the model, the learning (conditioning) of new items takes place only on the study event as seen in Equation 3:

$$S = \begin{matrix} CA_{s+1} & C\bar{A}_{s+1} & UA_{s+1} & U\bar{A}_{s+1} \\ CA_s \begin{bmatrix} 1 & 0 & 0 & 0 \\ C\bar{A}_s \begin{bmatrix} 0 & 1 & 0 & 0 \\ UA_s \begin{bmatrix} 1 & 0 & 0 & 0 \\ U\bar{A}_s \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{matrix}, \quad (3)$$

whereas the potentiating effects and the subjective maximization of retrieving the correct response (cf. Izawa, 1971b) occurs uniquely on the test event as:

$$T = \begin{matrix} CA_{s+1} & C\bar{A}_{s+1} & UA_{s+1} & U\bar{A}_{s+1} \\ CA_s \begin{bmatrix} 1 & 0 & 0 & 0 \\ C\bar{A}_s \begin{bmatrix} k' & 1-k' & 0 & 0 \\ UA_s \begin{bmatrix} 0 & 0 & 1-k & k \\ U\bar{A}_s \begin{bmatrix} 0 & 0 & k' & 1-k' \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{matrix}. \quad (4)$$

The effects of $(n-1)/2$ (mean for an n -

pair list under the study-test method) intervening study events, and those of $(n-1)/2$ (mean) intervening test events can be expressed as respective transition matrices in Equations 5 and 6:

$$t_s = \begin{matrix} CA_{s+1} & C\bar{A}_{s+1} & UA_{s+1} & U\bar{A}_{s+1} \\ CA_s \begin{bmatrix} 1-e & e & 0 & 0 \\ C\bar{A}_s \begin{bmatrix} e' & 1-e' & 0 & 0 \\ UA_s \begin{bmatrix} 0 & 0 & 1-e & e \\ U\bar{A}_s \begin{bmatrix} 0 & 0 & e' & 1-e' \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{matrix}, \quad (5)$$

and

$$t_t = \begin{matrix} CA_{s+1} & C\bar{A}_{s+1} & UA_{s+1} & U\bar{A}_{s+1} \\ CA_s \begin{bmatrix} 1-f & f & 0 & 0 \\ C\bar{A}_s \begin{bmatrix} f' & 1-f' & 0 & 0 \\ UA_s \begin{bmatrix} 0 & 0 & 1-f & f \\ U\bar{A}_s \begin{bmatrix} 0 & 0 & f' & 1-f' \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{matrix}. \quad (6)$$

The mean retention interval undergoes the effects of one t_s and one t_t with the study-test method, but two each with the anticipation method (cf. Izawa, 1972, 1981).

Similarly, the transition taking place during an intercycle interval is entered in the following matrix:

$$t_i = \begin{matrix} CA_{s+1} & C\bar{A}_{s+1} & UA_{s+1} & U\bar{A}_{s+1} \\ CA_s \begin{bmatrix} 1-g & g & 0 & 0 \\ C\bar{A}_s \begin{bmatrix} g' & 1-g' & 0 & 0 \\ UA_s \begin{bmatrix} 0 & 0 & 1-g & g \\ U\bar{A}_s \begin{bmatrix} 0 & 0 & g' & 1-g' \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{matrix}. \quad (7)$$

In the above, all components are assumed to be constant throughout the experiment.

With these equations in hand, via Izawa's (1981) assumptions and approximations with regard to expressing the sequence of events for both item presentation methods, theoretical expressions for data (learning functions) can be derived. Let $V_{1,s}$, $V_{2,s}$, $V_{3,s}$, and $V_{4,s}$ be, respectively, the probabilities of being in States 1 (CA), 2 ($C\bar{A}$), 3 (UA), and 4 ($U\bar{A}$) on the s -th cycle on which the T event was preceded by N study events. The start vector is:

$$(V_{1,0}, V_{2,0}, V_{3,0}, V_{4,0}) = (0, 0, J, 1-J), \quad (8)$$

where

$$J = g'/(g + g') \quad (9) \quad \text{and}$$

which is the asymptotic proportion of elements in Set A after a very long intercycle interval, assumed to exist prior to the first study event of the experiment.

Then, the proportion of incorrect responses, after the N th study event for the study-test method, $q_N(\text{st})$, can be derived as:

$$q_N(\text{st}) = {}_{\text{st}}V_{3,N}/({}_{\text{st}}V_{1,N} + {}_{\text{st}}V_{3,N}) \quad (10)$$

where

$$\begin{aligned} {}_{\text{st}}V_{3,N} &= (1-J)U Z_1^{N-1} \\ {}_{\text{st}}V_{1,N} &= 1-D-Z_1^{N-1}(1-D)(1-J) \\ &\quad - (1-k')EFW \left[L_1(1-J) \right. \\ &\quad \left. + J X_1 \left(\frac{1-P_1^{N-1}}{1-P_1} \right) \right], \end{aligned}$$

in which,

$$\left. \begin{aligned} D &= (1-k')\{f + F(w + eW)\} \\ U &= k' + K\{f' + F(w' + e'W)\} \\ P_1 &= 1 - X_1 - Y_1 \\ L_1 &= X_1(1 - Z_1^{N-2}) \\ &\quad + X_1 P_1 \left[\frac{(1 - P_1^{N-3})}{(1 - P_1)} \right. \\ &\quad \left. - \frac{Z_1(Z_1^{N-3} - P_1^{N-3})}{(Z_1 - P_1)} \right] \\ X_1 &= e + (w + fW)E + (1-k') \\ &\quad EFW\{F(w + eW) + f\} \\ Y_1 &= e' + (w' + f'W)E + k' \\ &\quad EFW + (1-k') \\ &\quad EFW\{f' + F(w' + e'W)\} \\ Z_1 &= 1 - [e' + (w' + f'W) \\ &\quad E + k'EFW + KEFW \\ &\quad \{f' + F(w' + e'W)\}] \end{aligned} \right\} \quad (11)$$

In Equations 10 and 11,

$$W = 1 - w - w'$$

where

$$\begin{aligned} w &= g(1 - G^m)/(1 - G), \\ w' &= g'(1 - G^m)/(1 - G) \end{aligned}$$

in which

$$\left. \begin{aligned} m &= 0 \text{ for } 0 \text{ s intercycle interval} \\ &\quad (\text{massed practice}), \text{ and} \\ m &= 3 \text{ for } 30 \text{ s intercycle interval} \\ &\quad (\text{spaced practice}) \end{aligned} \right\} \quad (12)$$

$$\left. \begin{aligned} E &= 1 - e - e' \\ F &= 1 - f - f' \\ G &= 1 - g - g' \\ K &= 1 - k - k' \end{aligned} \right\} \quad (13)$$

Stipulations on m in Equation 12 stemmed from the fact that we let t_1 present the effects of a 10 s intercycle interval, as is to be discussed subsequently.

Similarly, the proportions of incorrect responses after the N th study event under the anticipation method, $q_N(\text{ant})$, can be derived from the model as:

$$q_N(\text{ant}) = {}_{\text{ant}}V_{3,N}/({}_{\text{ant}}V_{1,N} + {}_{\text{ant}}V_{3,N}) \quad (14)$$

where,

$$\begin{aligned} {}_{\text{ant}}V_{3,N} &= (1-J)(1-Z_2)Z_2^{N-1} \\ {}_{\text{ant}}V_{1,N} &= 1 - (Z_2^{N-1} + L_2)(1-J) \\ &\quad - JX_2(1 - P_2^N)/(1 - P_2), \end{aligned}$$

in which,

$$\left. \begin{aligned} P_2 &= 1 - X_2 - Y_2 \\ L_2 &= X_2(1 - Z_2^{N-1}) \\ &\quad + X_2 P_2 \left[\frac{(1 - P_2^{N-2})}{(1 - P_2)} \right. \\ &\quad \left. - \frac{Z_2(Z_2^{N-2} - P_2^{N-2})}{(Z_2 - P_2)} \right] \\ X_2 &= e(1 - k') + (1 - k')E\{f + F \\ &\quad \{w + (e + fE)W\}\} \\ Y_2 &= k' + e'(1 - k') + (1 - k') \\ &\quad E\{f' + F\{w' + (e' + f'E)W\}\} \\ Z_2 &= 1 - k' - e'K - EK\{f' + F \\ &\quad \{w' + (e' + f'E)W\}\} \end{aligned} \right\} \quad (15)$$

Thus, the subject's performances under both item presentation methods are each functions of the fluctuations of elements between Sets A and $\bar{A}$ on the four types of events, expressed in the four matrices, t_s , t_r , t_l , and T , namely $e, e'; f, f'; g, g';$ and k, k' . Once these eight parameters are estimated, one can now make numerical predictions from the model. Parameters can be estimated in many ways. Even if only a single set of parameters is to be estimated, if all data points are used to estimate parameters, resultant deviations be-

tween the model and data are expected to be the least. The present study attempted theoretical work of a much higher order and examined whether the model can predict data that did not participate in parameter estimations in three different and highly restricted ways.

(a) *Mode I: Parameters estimated from the study-test method with spaced practice (Condition 3) only and predicting the data for all other conditions.* No theoretical bases seem to exist for the parameters controlling retention phenomena to differ either between the two methods, or between massed and spaced practice. Involved are element fluctuation probabilities e and e' (t_s), f and f' (t_T), and g and g' (t_I), respectively for intervening-study-events, intervening-test-events, and intercycle intervals. The six parameters, then, can be assumed to be constant throughout all conditions.

In this and all other parameter estimation modes in the present study, the six retention interval parameters were estimated from relevant spaced practice data. This followed from the fact that a unique solution of g and g' is not possible with massed practice with no intercycle intervals (no t_I), with the two parameters appearing only in the start vector (Eqs. 8 & 9).

In contrast, k and k' which reflect acquisition processes in the present model are expected to differ between massed and spaced practice. This follows from the fact that for each learning situation, the conditioning probability was set to be 1 (Eq. 3), and therefore differentials among conditions are reflected by k and k' values which are controlled by ease of retrievability of the correct response. Thus, these two parameters alone must be estimated separately for two variations of intercycle intervals.

In estimating individual parameters, it is necessary to specify the duration of the intercycle interval for which Matrix t_I (Eq. 7) applies. For the present data alone, a t_I that controls a 30 s intercycle interval

suffices. Nonetheless, in order to supply a comparative basis for this and forthcoming studies which involve 100 s intercycle intervals, let t_I present the effects of an intercycle interval of 10 s. Hence, the effects of a 30 s intercycle interval are expressed as t_I^3 . Stipulations regarding m in Equation 12 are based on this consideration.

Let us here use the study-test method for the parameter estimations. The eight values thus estimated from Condition 3 (study-test, spaced practice) alone via Kelley's computer program (least squares, cf. Izawa, 1981) were: $e = .998$, $e' = .111$, $f = .245$, $f' = .069$, $g = .450$, $g' = .010$, $k = .580$, and $k' = .025$. (Parameters with " $\wedge$ " indicate their estimations, whereas those without it, their theoretical values.)

The result from each of the three types of events in the retention interval is very intriguing: the probability of the element escaping from the available set, A , to the unavailable set, $\bar{A}$, is greater than the probability of the element moving in the reverse direction, i.e., $e > e'$, $f > f'$, and $g > g'$. The results of these phenomena correspond with the forgetting or retention loss attributable to activities occurring in these events. The findings are in the direction expected from the model. Intriguingly enough, the loss of the conditioned element from Set A was greatest during the intervening-study-events interval. Apparently, the learning of other items was more interfering with the learning of a target than either other items having been tested, or having no overt activities during the intercycle interval, in the current situation.

Similarly, the model expects $k > k'$. The larger k , the more unconditioned elements are eliminated from sampling in Set A (Eq. 4), and therefore, the more efficient the retrieval of the correct response on a test will be. In contrast, Probability k' controls both conditioned and unconditioned elements returning from Set $\bar{A}$ to A , and therefore, a large k' may not necessarily be advantageous for maximizing correct re-

trieval. Consequently a fairly large $\hat{k}$, and a relatively small $\hat{k}'$ are likely. All of these expectations were well borne out.

Using the above estimated parameter values from Condition 3, the theoretical values were computed via Equations 10–13 with $m=3$, and entered in the left upper panel of Fig. 2 for spaced practice under the study-test method. The predicted values (open round dots connected by solid lines) are compared with actual data (solid round dots). A very close agreement between the model and data was achieved. Tested by a chi-square goodness-of-fit test, commonly used in this field, the model did not deviate from data significantly, $\chi^2=3.129$, $df=4$, $p>.50$. On one hand, such a close agreement is no surprise, since eight parameters were estimated from 12 data points in this condition; on the other hand, such good results may not be forthcoming unless the model is viable.

More crucial and difficult to achieve is: can the parameters estimated from the study-test method data predict the antici-

pation method data that did not participate in parameter estimations in any way? The theoretical predictions for Condition 4 were made via Equations 14–15, and are entered by open square dots (connected by broken lines) in the upper right panel of Fig. 2 to compare with actual data (solid square dots). The model predicted data remarkably well, $\chi^2=12.099$, $df=12$, $p>.30$. The overall chi-squares for both spaced practice conditions were 15.228, $df=16$, $p>.30$.

How about massed practice data? Since k and k' cannot hold constant for both massed and spaced practice in terms of both the model and evidence, these two alone were estimated from massed practice study-test method data (Condition 1), while using the same six retention parameters as spaced practice. Massed practice $\hat{k}$ and $\hat{k}'$, thus estimated, were .238 and .010, respectively. The lower $\hat{k}$ and $\hat{k}'$ values here as compared with those with spaced practice are precisely in the direction predicted from the retention interval model: less acquisition with massed practice and thus harder retrieval of the correct response, and, therefore, smaller $\hat{k}$ and $\hat{k}'$ values.

Using the same procedure as above, by letting $m=0$ in Equation 12, the theoretical values from the model for massed practice study-test method data are entered in the left lower panel of Fig. 2 (connected with broken lines) along with observed data. The model did not deviate from data significantly, $\chi^2=4.209$, $df=10$, $p>.50$. Similar comparisons are made for massed practice data under the anticipation method in the right lower panel. Since this latter condition did not participate in parameter estimation at all, and is the most remote condition from the one by which all six retention parameters were estimated (in terms of both method and intercycle intervals), deviations between the model and data are not extremely small. Yet these deviations were small enough to support the model. It is indeed

Fig. 2. Comparisons of the retention interval model and data when the parameters were estimated solely from the study-test method. (All six parameters that control retention processes were estimated from spaced practice study-test method data (Condition 3, top left) only.)

remarkable that only 10 parameter values were needed to predict all 48 data points (based on a total of 14 400 observations) satisfactorily as seen in the four panels of Fig. 2. The total χ^2 was as small as 51.240 and nonsignificant, $p > .05$ with $df = 38$.

(b) *Mode II: Parameters estimated from the anticipation method with spaced practice (Condition 4) only and predicting the data for all other conditions.* To examine whether similarly satisfactory accounts of data can be made when parameters are estimated from the anticipation method, we repeated the same processes here, excepting that the anticipation method data provided the parameter values. $\hat{e}$, $\hat{e}'$, $\hat{f}$, $\hat{f}'$, $\hat{g}$, $\hat{g}'$, $\hat{k}$, and $\hat{k}'$ estimated from Condition 4 were, respectively, 1.000, .086, .238, .106, .487, .010, .570, and .030. Although these values were mostly different from (though close to) those under the study-test method, they exhibited similar trends among the parameters. $\hat{k}$ and $\hat{k}'$ for massed practice estimated from the anticipation method

(Condition 2) were .200 and .006, respectively. The relationships between massed and spaced practice were also consistent with those with the study-test method discussed under (a).

Figure 3 compares theoretical predictions and observed data. The closest agreement between data and the model was achieved for anticipation spaced practice (Condition 4, top left panel) from which all six retention parameters as well as the two retrieval parameters were estimated, $\chi^2 = 2.973$, $df = 4$, followed by massed practice anticipation data (Condition 2, left lower panel), $\chi^2 = 5.403$, $df = 10$, with $p > .50$ each and nonsignificant. The two right panels present study-test method data, with spaced and massed practice, respectively, despite their non-participation in any parameter estimations at all; the model did not deviate from data significantly either. The overall χ^2 was very small, a mere 32.703, $df = 38$, $p > .50$, and reconfirmed the viability of the retention interval model, supported independent of the item presentation method used in parameter estimations.

Fig. 3. Comparisons of the retention interval model and data when the parameters were estimated solely from the anticipation method. (All six parameters that control retention processes were estimated from spaced practice anticipation method data (Condition 4, top left) only.)

Fig. 4. Comparisons of the retention interval model and data when parameters were estimated from both study-test and anticipation methods. (All six parameters that control retention processes were estimated from spaced practice data, Conditions 3 and 4 (top panels) only.)

(c) *Mode III: Parameters estimated from both methods with spaced practice only, and predicting massed practice data.* An implication of the above results in (a) and (b) is that the two item presentation methods may not differ fundamentally in terms of their acquisition power. If so, parameters can be estimated from both methods simultaneously. When this was done with spaced practice data (Conditions 3 & 4), the values turned out to be: $\hat{e} = .983$, $\hat{e}' = .101$, $\hat{f} = .260$, $\hat{f}' = .084$, $\hat{g} = .465$, $\hat{g}' = .009$, $\hat{k} = .595$, and $\hat{k}' = .015$.

Relationships among the parameters are very similar to those estimated from either method alone. Numerical predications from the model based on these parameters are entered in the top panels of Fig. 4. The results were excellent, $\chi^2 = 7.509$, $df = 16$, $p > .50$. We used 8 parameters for predicting a total of 24 data points. As a matter of course, agreement between data and model was closer here than in situations where only one method was used for parameter estimations, vis-à-vis the two top panels of Figs. 2 and 3.

$\hat{k}$ and $\hat{k}'$ values estimated from massed practice data under both methods were .293 and .001, respectively. Via these plus the six retention parameters, numerical predictions for massed practice data were computed and entered in the lower panel of Fig. 4. The deviations of these theoretical values from data were also minimal and nonsignificant, $\chi^2 = 12.900$, $df = 22$, $p > .50$. When all four conditions were considered together, the resultant χ^2 was as small as 20.409, $df = 38$, and was clearly nonsignificant ($p > .80$).

General Discussion and Conclusion

Supportive demonstrations of a mathematical model or the failure thereof may, in part, be dependent on the parameter estimation modes utilized. The more parameters are estimated from the more data points, the easier it is for the model to achieve satisfactory results. Therefore, in

the present study, the aim was set for the highest possible level, and only a single set of parameters was estimated for the experiment. Furthermore, attempts were made that a single set of necessary parameters was estimated from the least possible data points: in (a) from the study-test method alone, and in (b) from the anticipation method alone, and in (c) from both methods; but in every case, the six retention parameters were obtained from spaced practice data only.

In spite of these most highly restrictive parameter estimation modes, the retention interval model predicted data closely throughout all estimation modes utilized herein including data which did not participate in parameter estimations.

The fact that successful predictions were possible from the model independent of whether parameters were estimated from the study-test method alone, the anticipation method only, or both methods may imply that the two methods may not differ fundamentally in terms of the acquisition processes, although they may differ in their retention processes as spelled out in the model (Izawa, 1981), in the present learning situation.

The above conclusion may appear to be inconsistent with the conditioning probabilities (proportion correct for the first time, given incorrect on all previous trials) entered in Table 1. Conditioning probabilities are greater with the study-test method than with the anticipation method, not only with massed practice, but also with spaced practice. This particular fact may challenge our assuming Parameters k and k' being the same for both study-test and anticipation methods within massed practice and within spaced practice, respectively.

This apparent inconsistency, however, can be resolved when we note the following important fact: in collecting data in Table 1, the responses were all measured at different mean retention intervals between the two item presentation methods.

The difference is inherent in the respective methods. The proportions of these relevant items correct for the first time, which are likely to be still in STM, were measured on the average after $n-1$ intervening events after the critical conditioning occurred on the study event under the study-test, but on the average after $2n-2$ intervening events under the anticipation method, as spelled out in the retention interval hypothesis.

Therefore, even if the two methods had about the same levels of acquisition on every study event, the statistics (conditioning probabilities) for the study-test method should be greater than for the anticipation method. Consequently, the differentials in these statistics may merely reflect differential retention processes based on differential retention intervals between the two item presentation procedures.

In conclusion, the present investigation seems to demonstrate major differences between anticipation and study-test methods, which were considered to be totally different in acquisition processes when our attention was focussed on operant conditioning, teaching machine, or immediate feedback principles; they may be very similar in terms of acquisition processes, and differ mainly in terms of their retention processes.

This position seems well supported by retrieval parameters that reflect acquisition processes more directly than statistics in Table 1 which are contaminated by differential retention intervals inherently involved between the two methods. k and k' estimated from spaced practice the study-test method alone were .580 and .025, respectively, whereas those from the corresponding anticipation method were .570 and .030; the resultant $Ks (= 1 - k - k')$ were .395 vs. .400, and, therefore, acquisition processes reflected in these theoretical constructs between the two methods seem nearly identical. A similar picture emerged from k and k' values with massed practice: they were .238 and .010 with the

study-test method respectively, and .200 and .006 for the anticipation method.

The retention interval model that postulates differential processes for three different events occurring during the retention interval is herein supported in spite of the ambitious usage of our three highly restrictive procedures of parameter estimations. The model seems to be capable of predicting data, not only those which participated in parameter estimations, but also those that did not, independent of either or both of the item presentation methods used for parameter estimations. Furthermore, the estimated values for each parameter across the three different modes utilized in the present study were quite similar, suggesting high internal consistency of the present theoretical work.

References

- Battig, W. F. 1973 Evidence that recall-anticipation differences increase with increasing time intervals. *Journal of Experimental Psychology*, **100**, 399-402.
- Battig, W. F., & Spera, A. J. 1962 Rated association values of numbers from 0-100. *Journal of Verbal Learning and Verbal Behavior*, **1**, 200-202.
- Izawa, C. 1971a Massed and spaced practice in paired-associate learning: List versus item distributions. *Journal of Experimental Psychology*, **89**, 10-21.
- Izawa, C. 1971b 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 Retention interval hypothesis and list lengths: Comparison of anticipation and study-test procedures. *Canadian Journal of Psychology*, **28**, 214-224.
- Izawa, C. 1977 Theoretical and empirical performance differences between anticipation and study-test methods: Retention interval hypothesis vs. spaced practice effects. *Japanese Psychological Research*, **19**, 31-38.
- Izawa, C. 1981 Toward a quantitative theory for performance differences between anticipation and study-test procedures: The retention inter-

- val model. *Scandinavian Journal of Psychology*, **22**. (In press)
- Izawa, C., Hayden, R. G., & Isham, K. L. 1980 Sensory modality and method of item information presentation in memory. *Acta Psychologica*, **44**, 131-145.
- Kanak, N. J., Cole, L. E., & Eckert, E. 1972 Implicit associative responses in verbal discrimination acquisition. *Journal of Experimental Psychology*, **93**, 309-319.
- Noble, C. E. 1961 Measurements of association value (a), rated associations (a'), and scaled meaningfulness (m') for the 2 100 CVC combinations of the English alphabet. *Psychological Reports*, **8**, 487-521.
- Underwood, B. J., & Ekstrand, B. R. 1967 Effect of distributed practice on paired-associate learning. *Journal of Experimental Psychology*, **73**, (4, Pt. 2), 1-21.

(Received Feb. 9, 1981; accepted May 16, 1981)