1976

Judgments of Knowing: A Learning-to Learn Analysis

Joseph King
Loyola University Chicago

JUDGMENTS OF KNOWING: A LEARNING-TO LEARN ANALYSIS

By

Joseph F. King

A Thesis Submitted to the Faculty of the Graduate School
of Loyola University of Chicago in Partial Fulfillment
of the Requirements for the Degree of
Master of Arts

May

1976

ACKNOWLEDGMENTS

The author wishes to express his sincere gratitude to the director of this thesis, Dr. Eugene B. Zechmeister, for his persistent and constructive criticisms during the preparation of this manuscript. The author is also indebted to the other members of his committee, Dr. Frank L. Slaymaker and Dr. Emil J. Posavac. Comments by Chris Gude and Stan Pasko are also acknowledged.

VITA

The author, Joseph F. King is the son of LaVerne Edward King and Mary (LeSage) King. He was born November 17, 1951, in Kankakee, Illinois.

The author graduated from Bishop McNamara High School in Kankakee, Illinois. In September, 1969, the author entered the University of Illinois-Chicago Circle. One year later, he enrolled at Loyola University of Chicago; and in 1973, he graduated Cum Laude with a bachelor of science degree in psychology.

In the fall of 1973, Joseph F. King was granted a research assistantship in experimental psychology at Loyola University of Chicago. In 1974, the author became a student member of the Midwestern Psychological Association and the American Psychological Association.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

CONTENTS OF APPENDICES

INTRODUCTION

The present study seeks to understand how the human subject may attempt to influence the learning process on the basis of what he knows, or assumes that he knows, about what is to be learned. While traditional verbal learning research has generally emphasized the effects that stimulus characteristics and presentation conditions have on learning, the present concern is with how the subject might use knowledge about learning in general, or knowledge about the specific to-be-learned material, to affect subsequent encoding and retrieval.

The adult human subject brings to the learning situation numerous verbal habits and substantial knowledge of associative relationships that have been acquired through a history of processing verbal material. These habits and information undoubtedly influence what is actually learned. For example, research on stimulus selection has revealed that the nominal (experimenter-defined) stimulus is not always identical to that stimulus the subject employs (Underwood, Ham, & Ekstrand, 1962). The distinction between nominal and functional stimuli is necessary since it is apparent that the learner may "modify" the learning situation so that

more meaningful elements of the stimulus array or more easily associated stimuli are acquired. Stimulus selection, therefore, is one obvious example of subjects bringing to the learning task habits which may influence how the to-be-learned material is encoded.

The adult learner may also change or adjust his manner of encoding on the basis of his judged progress toward an immediate learning goal. For example, the student who begins reading a difficult chapter in a textbook, gets halfway through the first page and says, "I am not understanding this.", and then begins again reading more slowly, is exhibiting this kind of behavior.

An experiment by Woodsen (1974) provides an illustration of how subjects may let future encoding activity be determined by successful behavior in the task. The task was to learn the English equivalents of 50 Japanese words under an anticipation method of paired-associate learning. Following one presentation of each of the 50 Japanese-English word pairs, the Japanese words were divided into five 10-item lists. The lists remained in the subject's view throughout the learning session which lasted approximately 2 hours. The subject was then asked the English equivalent of one of the Japanese words from each list, beginning with the leftmost list and proceeding across the five lists, with this procedure repeated constantly throughout the session. The subject responded to one

word at a time and was instructed to type his response on a teletype joined to a computer which was programmed to record the responses and display the correct English equivalent if the subject erred. The manner in which the next item was selected was the major independent variable (only two of the six different selection procedures are particularly relevant to the present discussion). One group was allowed to select the next Japanese word for testing and was given specific instructions regarding its selection. Subjects were specifically encouraged to select from those items which they felt they did not yet know. Another group was not given the opportunity to select items. For subjects in this group, the computer was programmed to display one item which had not been correctly anticipated on two successive trials. Subjects in these two groups performed substantially better on a test of retention given one week later than did all other subjects in the experiment (including subjects allowed to select items under no specific instructions). Therefore, the strategy of allocating study time to those items which were not yet learned, or which the subject himself judged to be not learned, proved to be quite successful. Since the subject himself, in at least one condition, guided the selection of "not yet known" items, this study illustrates the subject's ability to let success under the current learning conditions determine future encoding.

The ability to "know what you know" may best be viewed as a self-regulatory process. At some time in the learning process, the subject will likely ask himself, "Do I know this material well enough to remember it later?" It is on the basis of a negative answer that to this question that subsequent mnemonic activity follows. On the other hand, if the learner answers "yes" to this question, processing is likely to cease. Since current theoretical emphasis is placed on the activities of the learner (Jenkins, 1974), such self-regulatory abilities are of increasing importance. Tulving and Madigan (1970), for example have suggested that any theory of memory must include an analysis of the relation between the information stored in memory and the learner's awareness of what is stored.

Judgments of Knowing. Requiring the subject to judge what he knows during learning, and then later testing the veracity of these judgments, is a method of demonstrating his capacity for self-regulation. A judgment of knowing (JK) can be defined as a subjectivity rated likelihood of later retention of presently studied information. The time of the rating is crucial to the definition. Since retention tasks generally entail a study phase and a test phase, the JK response is, by definition, made during the study phase and before the test phase. For the purposes of the present discussion, both explicit and implicit JKs will be considered.

One example of the use of explicit JKs in the study of human memory is an experiment by Arbuckle and Cuddy (1969). Subjects learned a long series of short PA lists. As each pair was presented for study, the subjects were required to respond "YES" indicating that they would get the item correct or "NO" indicating that they would not get the item correct. Test trials immediately followed the single study trials. It was found that subjects could accurately predict recall or non-recall at greater chance levels. Furthermore, an additional group of subjects was asked to rate each pair on its ease or difficulty of learning. It was found that apparent difficulty of the pair was related to the probability of predicting correct recall. This was interpreted to mean that subjects were assessing the strength of the associative connection between stimulus and response terms at the time of presentation and were using this information as a basis for the prediction. More will be said about the use of item characteristics as a basis for JKs later in this introduction.

A second example of the use of explicit JKs is an experiment by Zechmeister and Shaughnessy (Note 1) in which subjects were asked to make JK ratings of individual items during the learning of a lengthy free-recall list. Rated items in the study list were either once- or twice-presented items, the latter having been presented under massed or

distributed conditions. Of experimental concern was the degree to which the JK ratings matched actual recall for both massed and distributed items.

Under such so-called MP-DP presentations, recall for massed items is generally lower than recall for distributed items (Shaughnessy, Zimmerman, & Underwood, 1972), a so-called "spacing effect." An attentional explanation of this phenomenon says that subjects are less likely to continue processing of the second presentation of an item under massed presentation than under distributed presentation (see Hintzman, 1974, for a review of this and other explanations of the spacing effect). It was suggested that subjects shift attention because they overestimate their memory for massed items relative to distributed items, and differences in JK ratings would reflect this perceived difference in the judged likelihood of recall.

The results of the Zechmeister and Shaughnessy study indicated that subjects could make JK ratings with some degree of accuracy. Once-presented items were rated lower (less likelihood of recall) than twice-presented items, and recall was, in fact, lower for once- than twice-presented items. Furthermore, the hypothesis regarding the ratings of massed versus distributed items was upheld in that, although massed items were recalled significantly less than distributed items, massed items received ratings similar to distributed items. This suggests that the

subjects might have been "misled" by the presentation conditions, and the attentional explanation of the spacing effect was supported.

A JK is a request for an assessment of how successfully an item has been stored in memory. It is assumed that one purpose such a judgment might serve during learning is that processing time may be allocated to items on the basis of this judgment. Experimental paradigms which allow the assessment of the subject's distribution of attention can be viewed, therefore, as examples of implicit judgments regarding the memory for an item.

A common finding is that the number of items learned during a certain interval is invariant regardless of the distribution of presentation times within the interval (Cooper & Pantle, 1967). Zacks (1969) was interested in testing the generality of this law of "total time invariance" when subjects were given increased control over presentation and testing conditions during the learning of two lists of paired associates. That is, study time was either free (subject-paced) or constrained (experimenter-paced), and the order of study and test trials was either free (subject-controlled) or constrained (experimenter controlled). Conditions for the learning of the first list represented a factorial combination of these two factors. The second list was learned under one of two conditions; either constrained study time and

ordering of study and test trials, or free study time and ordering of study of test trials. Thus there were eight groups in the experiment. Learning was to a criterion of one perfect trial.

Some of the results reported by Zacks are especially important to the present discussion. In a paired-associate list, one would not expect all items to be of equal difficulty, and thus subjects might spend more time on items perceived as difficult. To test this notion, for each subject Zacks ranked the items according to performance across all learning trials (ranking was from fewest to most errors). Similarly, the study times for each of the items were ranked for each subject. A positive rank-order correlation would indicate that more time was spent on difficult than easy items. For List 1 learning, the average median correlation across all groups was .66, for List 2, .72. Thus the subjects were appropriately allowing more study time to difficult items; and furthermore, with experience at the task, there was a slightly stronger relationship between time and item difficulty.

One would also expect that for a given pair, study time would be longer if an error was committed on the previous attempt than if the previous test was successful. Indeed, Zacks reported that the mean study time following an error was twice as long as that following a correct response.

One cannot claim that these findings unambiguously demonstrate that subjects were making implicit JKs. However, a relationship is suggested. That is, it appeared that subjects were allocating study time "as if" they were asking themselves whether or not they knew a pair well enough. This is an assumption which is important to later discussion.

Zacks employed a transfer paradigm to examine whether subjects under self-paced conditions learned specific "skills" which would aid future learning. Since self-paced and experimenter-paced presentations led to equal rates of learning on List 1, the groups were equivalent and the appropriate comparisons could be made. She reported that groups who had learned List 1 under self-paced conditions learned List 2 faster than subjects who had learned List 1 under experimenter-paced conditions. This suggests that subjects who were allowed to control their study time acquired skills which other subjects did not. Furthermore, the beneficial effects of learning List 1 under self-paced conditions were apparent even though second list learning was experimenter paced.

Since the assumption was made above that the process of self-pacing of study items is related to implicit JKs, these latter findings concerning positive transfer suggest an interesting question: can the performance of implicit JKs be similarly regarded as a "skill" which could facilitate

learning in subsequent tasks? That is, could successful learning based on correct implicit JKs be one of the many skills associated with learning-to-learn (cf. Postman, 1969)? The relationship between JKs and learning-to-learn is important for the present discussion.

Information which contributes to a JK. Several sources of information which aid the subject in making a JK can be suggested. One obvious source is item characteristics. For example, if a subject were shown a list of high-frequency words and a list of low-frequency words and was asked which would be easier to learn, he would probably select the high-frequency list. Through experience, certain item characteristics have become associated with "ease" or "difficulty" of learning. In a JK task, stimulus knowledge, or a knowledge of item characteristics, will likely influence the subject's judgment. Although, it must be pointed out that JKs need not depend on stimulus characteristics. If an item is learned sufficiently, that item's background frequency should not *per se* affect a JK.

The understanding of task demands may also influence a JK. If the above example were changed such that the subject were asked which list was easier to recognize, he *should* give the opposite response since it has been shown that low-frequency words are more easily recognized than high-frequency words (Underwood & Freund, 1970). An

understanding of task demands in general, or how item characteristics can interact with task demands, will be called process knowledge. Thus in a JK task, an awareness of how the material will be tested may affect the JK response. Other examples of process knowledge would include an understanding of retro- and pro-active interference. That is, if a to-be-judged item was very similar to one studied in a previous list, the JK may be influenced if the subject has some understanding of proactive interference.

A JK may also be influenced by implicit retrieval attempts. For example, during paired-associate learning, the subject may try to retrieve the mediator he had employed for that item before making his JK rating. If the task involved JKs after reading textual material, the subject may attempt to retrieve the "gist" of the sentences or paragraphs, and the success of these implicit retrieval attempts may affect the JK response.

Therefore, three major sources of information for a JK are: (a) stimulus knowledge, (b) process knowledge, and (c) implicit retrieval. The importance of each of these sources is likely to vary from task to task. For example, if homogeneous lists are used, differences in item characteristics would be minimal, and perhaps, process knowledge would be a major determinant of the JK response. However, it is suggested that across a variety

of laboratory tasks, the validity of these sources of information may be demonstrated.

Stimulus Knowledge

Subjects apparently know what characteristics of a stimulus are important in determining learning ease. Many types of items (words, nonsense syllables, consonant syllables) have been scaled for meaningfulness or association value. Richardson and Erlebacher (1958) examined the relationship between rated ease of learning and degree of associative connection between pairs of verbal items. Subjects were given a sample paired-associate list followed by a list of pairs of adjectives, nonsense syllables, or consonant syllables. Some subjects were asked to imagine that each pair were inserted in the sample list and then to rate the ease-of-learning each for pair as if it were a member of the sample list. Other subjects were asked to rate the associative connection between members of the pairs as compared to pairs in the sample list. Very high correlations between rated ease-of-learning (EL) and meaningfulness of the pairs were obtained.

In a more extensive analysis, Underwood (1966) demonstrated that subjects are capable of accurately judging the ease or difficulty of an item prior to actual learning of the item. He first instructed subjects to imagine that they were in a free-recall experiment and then presented a list of trigrams. Subjects were to make judgments of the

speed with which they thought they would learn each item in the imagined task. Following this, an incidental recall trail was given and then six actual study-test learning trials. Other groups of subjects simply learned the items or made pronunciability or meaningfulness ratings for the items.

Underwood correlated the ease-of-learning judgments with a measure of item recall. The results of the Underwood study and some of the problems identified with this type of research are relevant to the present discussion.

Mean EL ratings obtained from the group of subjects correlated highly (.96) with the mean number of times each item was correctly recalled across the six learning trials. However, one source of criticism of this correlational technique is that the requirement of EL judgments could produce a "bias." That is, subjects might have produced recall protocols which "matched" the ratings. An item judged as difficult before learning might have been treated differently during learning simply because it had received a "difficult" rating. Nevertheless, Underwood showed that the correlation between EL and learning measures taken from the same group was not higher than that obtained when EL ratings from one group were correlated with the learning scores of a group that did not make EL ratings. Thus, one can conclude that the relationship between EL and recall was reliable.

An additional concern of the Underwood study was whether an individual subject's EL ratings would predict his own learning. The correlation between an item's EL rating and the number of correct recalls across six learning trials was computed for each subject. The mean of the individual correlations was .48. Thus, subjects were able to predict their own learning of the trigrams with a substantial degree of accuracy. However, these correlations were substantially lower than the correlations between group EL and group learning. Also, Underwood reported that group EL scale values predicted group learning better than individual subject's EL ratings predicted group learning.

The difference in the relationships between individual ratings and learning and group EL ratings and learning suggests that across a group of subjects, a greater number of item characteristics which determine learning ease of difficulty are likely to be taken into account. This suggests that an EL value is more valid if it represents a summation of perceived characteristics which determine learning ease across a group of individuals (for example, EL values could be seen as analogous to an item's association value which is usually determined by a group of subjects). Underwood suggested that the reliability of an individual's ratings should be examined. Since, in Underwood's experiment, individual EL-learning correlations were determined on the basis of one observation which was

subject to error of measurement; the individual correlations may approach the magnitude of the group correlation if several EL ratings were made for each item.

It is also possible that an EL rating may contain idiosyncratic factors. That is, a subject may judge an item as easy to learn because of some idiosyncratic association or because of some personal familiarity with an item. In order to test this, Underwood, randomly repaired each subject's EL ratings with another subject's recall scores. With this repairing, the EL-learning correlations tended to decrease, and this is exactly what one would expect if the EL scale was, in fact, reflecting a specific individual's perception of an item.

In Underwood's analysis, individual differences in learning may also have affected the correlations. That is, for rapid learners a greater number of items would be recalled on all six trials, and a low correlation would result because of the attenuation in the range of possible learning scores. Underwood, therefore, ordered the subjects into six groups according to learning ability. As learning ability increased the mean correlations tended to decrease. This finding could suggest that fast learners were simply poor raters of EL. However, this is unlikely since for the fast learners, EL ratings correlated highly with the group recall scores. Therefore, it must be concluded that the decline in the correlations for the fast

learners resulted from a statistical rather than a conceptual problem.

One additional concern with the Underwood experiment is that artificially high correlations could have resulted because of the wide range of item types employed. The study list contained both high-meaningful three-letter words and low-meaningful consonant-consonant-consonant trigrams. Such a "mixed" list may have artifically drawn attention to item characteristics that are related to learning ease. Nevertheless, Lippman and Kintz (1968) replicated Underwood's findings even though a more homogeneous list of items was used (all CVC trigrams).

The important conclusion from the research cited above is that even before specific instructions to learn the items, subjects can accurately "predict" their learning of these items. It is assumed that this is done by attending to those item characteristics which determine learning ease. Stimulus knowledge is likely to be acquired through learning experience with a variety of items. The assumption that stimulus knowledge is acquired through learning tasks suggests a close relationship between stimulus knowledge and process knowledge. This relationship will be considered in the next section.

Process Knowledge

A second general source of information that is available for the performance of a JK can be called process

knowledge. This term is meant to refer to a general understanding of the "laws" of learning and a knowledge of specific task demands. For example, a student, upon being told of a forthcoming examination, will invariably ask whether the test will be an essay test or a multiple-choice test. This implies that the type of test will determine how the student studies for the exam. The student appears to have an implicit understanding of the difference between a recall test and a recognition test. A JK may depend on the expected manner of testing. Similarly, students may decide whether or not they have adequately prepared for a test on the basis of how successful certain study habits have been in the past. A student who has always read the textbook three times and outlined the material before a test and who has been consistently reinforced with acceptable grades, is likely to say, "I read the material three times _therefore_ I know it well enough to do well on the exam." Through experience, individuals develop general learning skills or study habits which can be applied to a variety of learning situations.

Development of Process Knowledge--Metamemory. The development of process knowledge actually begins with an understanding of what "remember" means. Flavell (1971) and Hagen (1971) have noted that very young children may equate remembering with repetition and that processes

such as meditation or rehearsal are not performed effectively by young children.

In addition to the development of memory performance, the development of knowledge about one's memory has recently been investigated by Flavell and his associates (Flavell & Wellman, in press). This knowledge about memory has been termed "metamemory." Kruetzer, Leonard, and Flavell (1975) interviewed young children to access the degree to which they understood their own memory capabilities. Children were asked, for example, "Suppose you looked up your friend's telephone number. If your brother asked you a question just before you begin to dial, do you think you would be able to remember the number?" Answers to this question suggested that children possess an implicit understanding of the concept of interference in immediate memory. From this and similar questions, the authors concluded that children show clear evidence of comprehending basic "laws" of human memory.

The understanding of immediate memory span was specifically tested by Flavell, Fredricks, and Hoyt (1970). Nursery school, kindergarten, second- and fourth-grade children were asked to predict their own memory span. They were shown a series of pictures of increasing lengths, and the maximum number of items the children thought they could remember was recorded. When these predictions were compared with actual recall, it was found that, while all children

tended to overestimate their memory capacity, the degree of overprediction dropped with age.

Yussen and Levy (1975) pointed to a possible methodological flaw in the Flavell et al. study and extended the investigation to include adult subjects. Yussen and Levy questioned the use of two demonstration trials (lengths of one and two items) on which nearly all Flavell's subjects were successful. These trials could have led to an unwarranted expectancy for success. To control for this bias, they presented stimulus arrays in decreasing as well as increasing order of difficulty. Furthermore, normative information was given to the subjects. That is, they were told "your friends can remember this many," when the series of appropriate length was presented. For all subjects (ages 4, 8, and 20 years) the tendency to over-predict was observed, but accuracy of prediction for all but the youngest subjects increased when normative information was given. Apparently, 4 year olds were uninfluenced by knowledge of their peer group's performance. It is interesting to note that adults were quite accurate in predicting their immediate memory span. The mean predicted recall was 5.83 items, and the mean actual recall was 5.52 items.

In addition to an understanding of one's general memory capacity, the awareness of certain "rules" of memory performance is important in the development of process knowledge. Moynahan (1973) examined first-, second- and

third-graders' awareness of the ease of recalling conceptually related items as opposed to unrelated items. Subjects were shown two cards, one containing eight unrelated items and one containing eight categorized items (four instances of two categories). Subjects were then asked to select that card which would be easier to remember. First-graders were less likely to select the categorized card than were third- or fifth-graders, yet the first-graders selected the categorized card more often than would be expected by chance. Thus, even first-graders showed some understanding of the facilitory effects of categorization on recall.

In an earlier part of this paper, the differential apportionment of study time to items within a list was assumed to be related to the JK process (cf. Zacks, 1969). Masur, McIntyre, and Flavell (1973) extended Zack's findings to include first- and third-graders as well as college-age subjects. The task was to learn a list of items which was 50% longer than each subject's memory span. After presentation of the list for 45 seconds, recall was tested. On subsequent trials, each subject was told that he could study only one half of the items and he was to indicate which items he wanted to study. However, as the subject's learning approached criterion (one perfect recitation) new items were added to the list prior to the subject's selection of study items.

In order to determine if the selection of study items was affected by the success of previous recall attempts, the proportion of items not recalled on the previous trial that were selected for study was computed for each subject. The proportions for first-graders were significantly less than for the other groups of subjects. In post-experimental interviews, none of the first-graders reported conscious selection of previously missed items. It was not likely that first-graders failed to understand the task since they, like the older subjects, tended to select "new" items which were added to the study list on the final trials. Furthermore, it was unlikely that first-graders failed to remember what they had recalled on previous attempts. Masur _et al_. retested eight subjects, and before study trials began, they asked subjects to indicate which items they had just recalled. Subjects were extremely accurate at this recognition task. The authors suggested that while the adults might have tried to remember all the items, younger subjects might have spent most of their effort on remembering those items which had been selected for study. Thus for the younger students, the strategy of allocating additional study time to unlearned items would not have been appropriate for the task as they perceived it. Masur _et al_. concluded that this strategy was probably absent from the young children's repertoire of study behaviors.

From the above research, two general conclusions regarding the development of process knowledge can be stated. First, even young children exhibit an awareness of the general "laws" of memory and are aware of the capacities of their memories. Secondly, metamemory and memory performance seem to develop in parallel. This suggests a rather strong relationship between process knowledge and learning ability.

Process knowledge: knowledge of task demands. A subject who differentially encodes verbal material on the basis of the type of test he expects is exhibiting process knowledge. Gude and Zechmeister (Note 2) reported that subjects anticipating a recognition test showed poorer recall than subjects expecting a recall test. They suggested that this effect was due to the reduced level of attention subjects found necessary to produce efficient performance on a previous recognition task.

Kellas, McCaully and McFarland (1975) also demonstrated that subjects employ task-specific skills during encoding. Subjects were given successive 15-item lists for either serial learning or free recall learning. Presentation was self-paced and thus the distribution of study times for serial and free recall learning could be examined. Under serial learning instructions, a direct linear relationship between study-time and serial position was observed. This suggests that rehearsal of items

was cumulative. For free-recall, this relationship was less systematic. This difference suggests that the perception of task demands led to different encoding behavior.

It can be suggested that the subject's perception of the appropriateness of the encoding of an item might influence his JK response. This would require process knowledge.

Process knowledge--"Mis-judgments" of knowing. Erroneously perceived task demands can lead to what may be called implicit "misjudgments" of knowing. As was discussed above, Zechmeister and Shaughnessy (Note 1) suggested that the spacing effect could be related to the subject's mis-perception of task demands. The attentional explanation of the spacing effect is somewhat dependent on the subject's misjudgment of what is necessary for successful retention of massed items.

It is likely that other memory phenomena can be explained on the basis of erroneous implicit JKs. Craik (1970) and McCabe and Madigan (1970) have reported that if a subject is presented with a series of short free recall lists, terminal items will be recalled almost perfectly on an immediate test, but on a delayed final test given after the series of short lists, these terminal items will be recalled very poorly. This recall decrement has been called the "negative recency effect."

Light (1974) encouraged subjects to increase rehearsal of terminal items and informed subjects that they would be given a final test. Even with this clarification of task demands, the negative recency effect was still obtained. She suggested that terminal items were rehearsed in a different manner than other items. That is, terminal items were rehearsed in a manner that merely maintained the items, whereas other list items were given more "elaborate" rehearsals (cf. Craik, 1973). Thus subjects were misjudging what was necessary to assure retention of terminal items. Watkins and Watkins (1974) found that when subjects were given a series of lists of varying lengths, and thus could not know when the end of the list was approaching, the negative recency effect vanished. Therefore, under these conditions, it can be assumed that subjects did not treat terminal items differentially. The subjects were "prevented" from making a misjudgment of knowing at the time of encoding, and this apparently increased retention.

These examples suggest that with learning experience, it is likely that "inappropriate" as well as appropriate learning strategies can be acquired. Therefore, part of what is called process knowledge is the knowledge of the relationship between task demands and appropriate encoding processes.

Implicit Retrieval

In addition to stimulus knowledge and process knowledge, a JK may also depend on a direct examination of the contents of memory, that is, implicit retrieval. According to the definition of JKs stated earlier in this paper, a JK response is made during the study phase of a learning task. Therefore, a pre-test phase retrieval attempt is assumed to be made without the benefit of explicit retrieval attempts (or test trials).

An obvious way to decide if you are going to be able to remember certain information at a later time is to try to retrieve that information. If this attempt is successful, attention may be directed to less well-established information. For example, if a JK is required during paired-associate learning, the subject may attempt to retrieve the mediator he had employed for that pair on earlier trials. In free recall, the subject may review associative connections between list items or rehearse retrieval strategies when making a JK. If these retrieval attempts are successful, the subject may rate an item as likely to be remembered.

However, it must be noted that a delayed test may render information gained from an implicit retrieval attempt invalid. That is, in paired-associate learning, a subject may successfully retrieve the mediator for a list pair and make a JK response. However, during the retention

interval, forgetting of the mediator may be just as likely as forgetting of the response term. Thus the JK would become less reliable if no other relevant sources of information were available to the subject.

As was mentioned earlier, a JK may be seen as a self-regulatory process. The assumed purpose of a JK is to allow processing time to be allocated in the most efficient manner. That is, distinguishing between "known" and "unknown" material will allow extra attention to be given to unknown material. Furthermore, it must be pointed out that the JK response for an item may be made <u>relative</u> to the degree to which other to-be-learned material is successful encoded. Thus it is possible that during an implicit retrieval attempt, complete retrieval of information may indicate to the subject that the material is well-established and that further processing would be unnecessary. However, a partially successful retrieval attempt may indicate the degree to which further processing is necessary. A partially successful retrieval attempt tells the subject that information is not completely absent from memory and that with additional study time, that information may easily become well-established. Therefore, information gained from partial retrieval may be helpful to the subject who is making a JK response. Two rather common memory phenomena which have been related to partial retrieval will now be considered.

Relationship between JK and Tip-of-the-Tongue Phenomena. The tip-of-the-tongue (TOT) phenomenon was empirically demonstrated by Brown and McNeill (1966). Subjects were presented with the definition of an uncommon English word and were then asked to supply the word. If the subject could not recall the word but indicated a TOT state, he was asked what he knew about the target word. If an items was not retrieved, information such as the first letter, number of syllables, letters in various positions, syllabic stress, and suffixes could often be accurately reported. It appeared as though the greater the number of attributes recalled, the greater the likelihood of eventual successful retrieval of the target word.

Koriat and Lieblich (1974) improved the analysis of TOT states by accounting for guessing rates. For example, if you are searching for the name of a chemical element, you are more likely to report that the target word has four syllables than one syllable based on your general knowledge of chemical terminology. This more analytic approach led them to suggest that both knowledge about the class of items to which the target word belongs (class detection) and the retrieval of information specific to the target word (differential detection) can result in a TOT experience.

Recently Yarmey (1974) studied the TOT phenomenon in relation to non-verbal information. Subjects were shown

pictures of famous people and were required to supply their names. In a manner similar to the Brown and McNeill study, subjects reporting a TOT state were asked to indicate the occupation of the target person, the places where the target person would most often be seen, and how recently the target person had been encountered. Orthographic and phonetic information about the target person's name was also requested. Yarmey found that initial letters, number of syllables, and sound of the target name were not as frequently supplied as were reports of the target person's profession, or the place or time of recent encounter with the person. Thus while Brown and McNeil demonstrated that subjects in a TOT state could accurately report verbal information, Yarmey showed that spatial-temporal or nonverbal information may also be retrieved in a TOT situation.

A very similar memory phenomenon, called the "Feeling-of-Knowing" experience (FK), occurs when one reports that unrecallable information is recognizable. Hart (1965, 1966, 1967) has extensively analyzed this phenomenon, and since the methods and analyzes he employed are similar to those which will be employed in the present experiment, these studies will be considered in detail.

Hart exploited the well-established fact that recognition is generally better than recall. The general procedure was as follows. Subjects were given general

information questions that had only a single answer. For each question that could not be answered, subjects were told to make a yes-no judgment as to whether the correct answer could be selected from several alternatives. Following completion of this task, multiple-alternative recognition tests were given. The technique has been called the recall-judgment-recognition (RJR) paradigm.

An accurate FK judgment could represent one of two cases. When a subject said he knew the answer (YES judgment) and, in fact, he could select it from the alternatives on the recognition test, the response was called a FK hit. Likewise, when the subject reported that he could not recognize the answer (NO judgment) and, in fact, could not select the correct alternative on the recognition test, the response was called a Feeling-of-NOT-knowing hit ($\overline{\text{FK}}$ hit). A FK miss and a $\overline{\text{FK}}$ miss represent instances in which the judgments did not match the subsequent recognition performance. To test the accuracy of FK judgments, recognition of items the subject "felt he knew" should be compared with recognition of items the subject "felt he did not know"; that is, FK hits versus $\overline{\text{FK}}$ misses.

Hart (1965) found that more non-recalled items given a YES FK rating were recognized than items given a NO FK rating. While this indicated that accurate judgments can be made about what is in storage, judgments of what is

not in memory were found to be less accurate than would be expected. It was found that 43% of the items subjects judged as not in memory were nevertheless subsequently correctly recognized. Hart (1966) reduced this $\overline{FK}$ miss rate by providing more equally-likely alternatives on the recognition test and by including a six-point rating scale rather than the yes-no dichotomy for FK judgments. The $\overline{FK}$ miss rates was reduced but did not vanish under these more sensitive conditions.

Hart (1966) also improved the RJR technique by requiring subjects to guess on every recall attempt, since subjects who withheld retrieved information could have effectively inflated the FK accuracy scores. That is, if a subject remembered that he had earlier judged an item to be unknown, he could increase the accuracy of his FK responses by simply withholding the correct response (in this case, the $\overline{FK}$ miss rate would remain low, thereby increasing the difference between FK hits and $\overline{FK}$ misses). The guessing requirement tended to reduce the relative magnitude of FK hits to $\overline{FK}$ misses; however, Hart concluded that the FK experience was still demonstrated.

In a later experiment, Hart (1967) employed the RJR paradigm with word-trigram paired-associates. After one, two, or three study trials, a recall test was presented. A response was demanded for each stimulus on the test list, and if the subject believed he was responding

correctly, he was to circle the response. For all uncircled items, the subject was to make a FK judgment. Three learning trials produced a greater difference between FK hits and $\overline{\text{FK}}$ misses than did one or two trials. (Both FK hits and $\overline{\text{FK}}$ misses increased as the number of study trials increased, but the former increased at a faster rate than the latter.)

Perhaps a weakness in Hart's research concerns the problem of criterion differences across subjects. That is, some subjects may be more cautious in predicting later recognition than others, and given some method of measuring criterion, the accuracy of FK judgments might be shown to increase. (See Murdock, 1966, for a discussion of criterion problems in memory tasks.)

In general, it is important to consider information available to the subject which tells him that to-be-learned material is *not* yet learned. This information can be seen as just as important to the subject as is information which leads him to believe that the material is already learned. Therefore, under certain conditions, the TOT experience and the FK experience could be related to the performance of a JK.

JKs and Learning to Learn

As was mentioned earlier in this paper, the capacity to make JKs is related to the notion of learning-to-learn in that JKs may represent a skill which can be applied

across a variety of learning situations. The present experiment is directly aimed at the question of JKs as a learnable skill.

In research in human learning, transfer effects have been traditionally divided into two classes: specific and non-specific transfer. The present experimenter deals with one aspect of so-called non-specific transfer, learning to learn. Non-specific transfer effects are assumed to be independent of any purposeful similarity of the to-be-learned material across a series of tasks. Specific transfer effects, on the other hand, are traditionally seen as dependent on planned relationships between items on successive lists to be learned.

The above classification of transfer effects has recently been challenged. In an extensive analysis of learning-to-learn, Postman (1969) has presented the view that no clear discontinuity between specific and non-specific transfer can be postulated. That is, types of transfer are seen as lying along a continuum of applicability. Transfer represents a "carrying over" of an acquired habit or skill from the learning of one list to the learning of another. Some skills are more generally applicable than others.

The effects of learning-to-learn can be demonstrated by the learning of two lists which reflect no systematic relationship or similarity between the items comprising

the two lists (e.g., A-B, C-D). During learning of the first list, response integration skills (among others) may be acquired which facilitate learning of the second list. For example, if List 1 contains low-meaningful trigrams as response terms, then learning experience with these items will facilitate second list learning to a greater extent than if List 1 was comprised of readily available high meaningful words (Postman, Keppel & Zacks, 1968). A skill was acquired. This skill was the efficient processing of low-meaningful items.

Other conditions of paired-associate learning may lead to the acquisition of higher order "rules" for learning lists of a specific structure. Postman (1969) suggests that rules of inclusion and exclusion are employed to discriminate between correct and incorrect responses. For example, in the specific transfer paradigm, A-B, C-B, response terms are identical on the two successive lists, but the stimuli differ. This paradigm requires the use of the rule of inclusion, or the use of "old" response terms. Furthermore, the use of these rules can be learned across a series of pairs of lists. That is, employment of the rules can result from learning-to-learn (Postman, 1964).

Postman (1969) concludes that such skills as response integration, method of practice (serial vs. paired associate learning), and higher order rules of response

selection may be included under the classification of learning-to-learn.

LaPorte and Voss (1974a) furthered the analysis of learning-to-learn effects by independently manipulating (1) the number of lists learned before the critical transfer list, and (2) the criterion of learning each list. Thus, the number of associations to be learned could be held constant while the criterion differed. Learning four eight-item lists to a 50% criterion requires the same number of associations as does the learning of two eight-item lists to a 100% criterion. Some of the results reported by LaPorte and Voss are especially relevant to the present discussion. They found that if a constant number of associations were required for pre-critical list learning, positive non-specific transfer was more likely to result if the initial lists were learned to a 100% criterion than if the initial lists were learned to a 50% criterion. Thus it appeared that, in certain situations, one aspect of positive transfer is the ability to process unlearned pairs when part of the list has already been learned. A "mastery" of the list requires the subject to discriminate the pairs that have been learned from the pairs that have not been learned and then to allocate processing efforts accordingly.

In a related study of paired-associate learning, LaPorte and Voss (1974b) substituted a computational

filler task in place of test trials. Groups were given 5, 10, 20, or 40 such non-test trials and then were required to learn the list under normal study-test conditions to a criterion of one perfect trial. A control group was given only the normal study-test learning trials. Performance of the first test trial by the groups who received no initial test trials was compared to the number of study-test trials which was necessary for the control group to reach an equal degree of learning. The control group required about half as many trials to reach the level of performance of the groups who were not previously tested.

The authors concluded that in the testing conditions, valuable "feedback" information was available to the subject. This feedback allows the subjects to determine which pairs are learned and which pairs require further processing. If testing were absent, it would become more difficult for subjects to discriminate between known and unknown pairs. Also, when test trials are removed, the subjects' rehearsal may be unrelated to the degree to which an item is successfully stored in memory. LaPorte, Voss, and Bisanz (1974) attempted to extend these findings to a transfer paradigm. However, no evidence was obtained which suggested that elimination of test trials during the learning of List 1 produced a detriment in the second list learning. In the present experiment, the role of test trials in the JK process will be examined.

THE PRESENT EXPERIMENT

The present experiment was designed to answer the following questions. First, under paired-associate (PA) learning conditions, to what extent do JKs depend on knowledge of previous test trial performance? Second, does JK accuracy improve as familiarity with specific task demands increases? Third, does JK experience lead to the development of learning "skills" which can be employed when conditions of learning change?

A learning-to-learn paradigm was employed. Four groups of subjects learned three different PA lists, and two critical aspects of the learning procedure comprised the independent variables.

Differences in the learning procedure during the _first two_ lists defined the first independent variable. One half the subjects learned the pairs under an alternating study-test trial procedure (STUDY-TEST groups); while, for the other two groups, intervening test trials were omitted and a single test trial was given for each list (STUDY-ONLY groups). On the _third_ list, all subjects learned the pairs under the STUDY-ONLY conditions.

A second important aspect of the procedure involved the presence or absence of a JK rating trial after each

PA list. Two of the groups were shown the pairs after learning and were asked to make JKs (JK groups). For the remaining two groups, no JKs were required for the first two lists, and subjects were given an additional study trial (CONTROL groups). For the third list (STUDY-ONLY), all subjects were asked to make JK ratings after learning and before the test trial.

Therefore, four groups, representing a factorial combination of the two independent variables (STUDY-TEST vs. STUDY-ONLY, and JK vs. CONTROL) were employed. These groups will be referred to as STUDY-TEST-JK, STUDY-TEST-CONTROL, STUDY-ONLY-JK, and STUDY-ONLY-CONTROL. A diagram of the procedure is contained in Table 1. Both the improvement in JK accuracy across the first two lists and the level of accuracy on the third list were of interest.

Regarding the major purposes of the present experiment, the following hypotheses can be stated.

First, since subjects in the STUDY-TEST-JK group will have knowledge of test trial performance available to them while the subjects in the STUDY-ONLY-JK group will not, JK accuracy will be greater for the former group than for the latter group on the first list.

Second, since subjects in the STUDY-TEST-JK and STUDY-ONLY-JK groups will gain experience with task demands across the first two lists, both these groups will show improvement in JK accuracy. Since the

Table 1

Schematic Diagram of the Procedure

Group	Lists 1 & 2			List 3		
A. STUDY-TEST-JK	S-T, S-T, S-T	JK	TEST	S, S, S	JK	TEST
B. STUDY-ONLY-JK	S, S, S, S, S	JK	TEST	S, S, S	JK	TEST
C. STUDY-TEST-CONTROL	S-T, S-T, S-T	S	TEST	S, S, S	JK	TEST
D. STUDY-ONLY-CONTROL	S, S, S, S, S	S	TEST	S, S, S	JK	TEST

S-T - PA study and test trial

S - study trial only

JK - Judgment of Knowing trial

STUDY-TEST-JK group should have more information relevant to a JK available, this group should improve at a faster rate than the STUDY-ONLY-JK group across the first two lists.

Third, the STUDY-TEST-JK group should show a decrease in JK accuracy on List 3 when intervening test trials are removed. That is, for this group, conditions of learning were changed such that knowledge of test trial performance would not be available. If this information was crucial for the STUDY-TEST-JK group's JK performance on the first two lists, then JKs should be less accurate on List 3.

The fourth hypothesis states that subjects in the STUDY-ONLY-JK group will exhibit continued improvement in JK accuracy across all three lists. This would indicate that subjects learned to make JKs on knowledge other than that provided by test trials.

The fifth hypothesis concerns JK accuracy on the third list. The STUDY-TEST-JK group will show greater JK accuracy than the STUDY-TEST-CONTROL group since the former group will have had experience with the JK task while the latter group will not have had this experience. Similarly, the STUDY-ONLY-JK group will show greater JK accuracy than the STUDY-ONLY-CONTROL group on the third list.

A secondary issue which was investigated concerned the effect on recall of making JKs. That is, does the requirement of making JKs lead the subject to more efficient

processing than when no JK requirement is present and the subject is left to his own devices? Within the STUDY-TEST groups, it is expected that those subjects making JKs will recall more items than those subjects not making JKs. Similarly for the STUDY-ONLY groups, the STUDY-ONLY-JK group will show greater recall than the STUDY-ONLY-CONTROL group.

METHOD

Design. Four groups of subjects learned three lists of 24 PAs. Two groups learned the first two lists under an alternating study-test trial procedure (STUDY-TEST groups). The other two groups learned the first two lists with no test trials intervening between study trials and only a single test trial following several study trials (STUDY-ONLY groups). Within the STUDY-TEST groups, one group was asked to make JK ratings before the last test trial (STUDY-TEST-JK). The other STUDY-TEST group received an additional study trial in place of the JK rating trial (STUDY-TEST-CONTROL). Similarly for the STUDY-ONLY groups, one group received a JK trial before the single test trial (STUDY-ONLY-JK), while the other group received an additional study trial (STUDY-ONLY-CONTROL). For the third list, all groups learned under the STUDY-ONLY conditions, and all groups made JK ratings before the single test trial. Therefore, the four groups were distinguished by the learning procedure and by the presence or absence of the JK rating trial on each of the first two lists.

Materials. Three 24-item lists of PAs were constructed. From the Paivio, Yuille, and Madigan (1968)

norms, all two-syllable words were selected with frequencies between 1 and 50 per million (Thorndike & Lorge, 1944). The words in this pool were ranked on imagery value, and the 72 items with the lowest imagery values were selected for use as response terms. These items were randomly divided into three sets of 24. Stimulus terms were 72 CVC trigrams randomly selected from Noble's (1961) norms. Stimuli ranged in association value from 70% to 78%. From the pool of 72 CVCs, items were randomly divided into three sets of 24 items with the qualification that, within an item set, no two trigrams could have the first two letters in common. Stimulus and response item sets were paired, and the 24 CVCs and two-syllable words within a set were randomly paired to form three lists of 24 PAs. Stimulus-response pairs were typed on index cards for study trial presentation.

In order to prevent serial position from serving as a cue in PA learning, on each study trial, subjects received a different order of study items. To construct different orders, each of the three study lists was considered as four blocks of six items. Items within a block were randomly rearranged on each trial. Also, the order of the blocks was random on each trial. Furthermore, the order of item presentation for rating and test trials was different than that for study trials.

Study, rating, and test lists were constructed such that the interval between studying and rating an item and the interval between rating and testing an item would be controlled. For this reason, all lists were divided into blocks of items, and membership in a block was systematically controlled. Study lists were arranged into four blocks of six items. Rating lists included six blocks of four items, and each block was composed of one item randomly selected from each of the four study list blocks. Using this procedure, two different orders of list items were constructed to be used for JK ratings for each item set. Test lists were also divided into four blocks of six items. Each test block contained one item randomly selected from each rating list block. Subjects in the STUDY-TEST groups received several test trials over each item set, and in order to prevent repeated exposure to the same order of test items, three different orders of test items were constructed using the same procedure.

The rating and test lists were printed on half sheets of paper and inserted in envelopes such that only one item could be exposed at a time.

Pilot Study. A pilot study was conducted in order to determine the necessary number of learning trials to reach a criterion of approximately 50% for each of the conditions of the present experiment. Using the same materials as described above, subjects in one pilot group were required

to learn the PAs under STUDY-TEST conditions, and subjects in a second pilot group learned the PAs under STUDY-ONLY instructions. A 50% criterion was attained after three trials for the STUDY-TEST pilot group and after five trials for the STUDY-ONLY pilot group. Items were presented at a 3 second rate. Hence, these parameters were used in the present experiment.

JK Task. A JK is defined as the subjectively rated likelihood of later retention of presently studied information. Accuracy of the prediction is determined by eventual recall or non-recall of the rated item. To insure the validity of the JK, it is necessary to prevent subjects from withholding responses. That is, if a subject remembered at the time of test that he had not predicted recall for that particular item, he/she might withhold the response in order to achieve correct prediction. To prevent this, a game devised by Pasko (Note 3) was employed, and game points were administered on the basis of correct recall and use of the JK scale.

A six point scale was provided for the JK rating. The scale was intended to be dichotomous and to reflect the degree of confidence in the prediction. Table 2 contains the JK scale and the conditions under which points were assigned. Subjects were instructed that they would receive +5 points for every word they recalled and -5 points for every word not recalled. Additional bonus or

Table 2

Summary of the Rules For Receiving Points in the JK Task

1 2 3	4 5 6
No I will not recall the item.	Yes I will recall the item.

1. For each response term recalled on the final trial you will get +5.
2. For each response not recalled on the final trial you will get -5.
3. If you recall an item, and you indicated a YES, (e.g. 4, 5, or 6) then you get bonus points.

1	2	3	4	5	6
-3	-2	-1	+1	+2	+3

   If you indicated on your JK that you would <u>not</u> recall the item (e.g. 1, 2, or 3) then you will lose points.

4. If you do not recall an item you lose 5 points but you may gain some of the points back if your JK matched your recall. That is, if you said you would <u>not</u> recall an item, then you get bonus points.

1	2	3	4	5	6
+3	+2	+1	-1	-2	-3

   Note, however, that if you reported that you <u>would</u> recall the item, but you didn't, then in addition to losing 5 points for not recalling the item, you would be penalized an additional 1, 2, or 3 points.

penalty points were to be assigned on the basis of the specific JK response. Briefly, subjects were told that if their prediction matched their recall performance, they would receive bonus points corresponding to the degree to which they were sure that recall would or would not follow. Likewise, penalty points were determined by the extremeness of the JK response when it did not match recall performance. It should be noted that for an accurate prediction of not knowing (or non-recall), an item was rewarded by as many as +3 points but that -5 points were assigned for a missed item. Consequently, maximum game points could only be gained by recalling as many items as possible.

Procedure. Subjects were seen in pairs and were assigned to groups by a blocked randomization procedure upon appearance at the laboratory. All subjects were told that they were to participate in a memory study and that their ability to predict what was known would be of importance. They were instructed that pairs of items (CVC-two-syllable word pairs) would be presented and production of the two-syllable word would be required. They were told that they would be allowed to see the list several times; but they were not told that more than one PA list would be presented for learning.

For the STUDY-TEST-JK and STUDY-ONLY-JK groups, the specific details of the JK task were explained before presentation of the first list. Cards displaying the JK

scale and the game's payoff matrix (see Table 2) were placed in front of the subjects, and the specific rules were briefly explained. The concept of prediction was frequently mentioned throughout the JK task instructions. Since subjects in the two CONTROL groups did not make JKs until after presentation of the third list, these subjects were told that, later in the experiment, they would participate in a game regarding their ability to predict what was known.

For the learning of the first two lists, study decks were placed in front of the subject, and he/she was instructed to turn to the next card at the sound of a tone. Tape recorded tones occurred at a 3 second rate. Following a study trial, subjects in the STUDY-TEST groups were given an envelope containing a test sheet and were told to pull the sheet out of the envelope to expose the next stimulus item whenever they heard a tone. Next to each stimulus term, they were to write down the response term. Subjects were encouraged to guess and were instructed to place an "X" next to those CVCs for which they could not produce a response. Tones during test trials occurred at a 5 second rate. Following three such study-test-trial cycles, subjects in the STUDY-TEST-JK group were given an envelope containing the list of study pairs and a JK rating sheet. Tones occurring every 5 seconds paced the subjects through the JK list. Subjects wrote a scale

value next to each pair as it was revealed. In place of the JK rating list, subjects in the STUDY-TEST-CONTROL group were given a study deck for an additional study trial. These subjects were told that the list would be presented again at a slower rate, and for this trial, items were also presented at a 5 second rate. Then, for both STUDY-TEST groups, a last test trial was given.

Subjects in the STUDY-ONLY groups did not receive test trials following each study trial, and five study trials were presented. Items were shown at a 3 second rate. After the study trials, subjects in the STUDY-ONLY-JK group were given the JK rating list followed by a recall test as described above for the STUDY-TEST-JK group. The STUDY-ONLY-CONTROL group received a sixth study trial before the test trial in place of the JK task. This procedure was followed for the first two lists.

For the learning of the third list, all groups were told that the learning procedure would be slightly different for the third list but that the same type of PAs would be presented. Subjects in the STUDY-TEST groups were told that now they would not be tested after each study trial but would receive only a single test. Subjects in the STUDY-ONLY groups were told that the procedure would be the same except that fewer learning trials would be allowed. For subjects in the CONTROL groups, the JK task instructions were presented at this time, and for all other subjects,

the instructions were reviewed. The interval between the end of List 2 and the beginning of List 3 presentation was approximately equal for all groups. Following these instructions, all subjects studied the third list for <u>three</u> study trials with no intervening study trials followed by a JK trial and a final test trial.

The six possible orders of the three item sets were employed, and within a group, five subjects learned the item sets in each order.

<u>Subjects</u>. Subjects were Loyola University undergraduates participating to fulfill a course requirement. Thirty subjects served in each group. Four subjects had to be replaced due to a failure to follow instructions.

RESULTS

Initial analyses were performed to determine whether differences in recall could be attributed to item sets or rating-test forms. None of these *F* ratios reached significance, and, subsequently, these were not included as factors in any of the analyses of variance to be reported.

The following terms were used to identify factors in the analyses of variance: *study condition* (i.e. STUDY-TEST vs. STUDY-ONLY); *JK experience* (i.e. JK vs. CONTROL); and *Lists* (1, 2, & 3).

Recall. Analyses were performed to determine if study condition and/or JK experience led to different levels of recall. Figure 1 displays the mean number of correctly recalled response terms for the three lists for each of the four groups. For the STUDY-TEST groups, these means are based on the last test trial; and for the STUDY-ONLY groups, the means are based on the single test trial recall scores. The ANOVA source table is contained in Table 3. The main effect for study condition reached significance, $F\ (1,116) = 8.94$, $p < .005$. Although pilot work had indicated that recall following three study-test trials was equivalent to that after five study-only trials, the STUDY-TEST groups recalled more items on each list

than did the STUDY-ONLY groups. Furthermore, the main effect for lists was significant, $F\ (2,232) = 49.35$, $p < .001$. As can be seen in Figure 1, more items were recalled on the second list than on the other two lists. The study condition by list interaction also reached significance, $F\ (2,232) = 7.71$, $p < .05$. Examination of Figure 1 reveals that across the first two lists, the increase in recall for the STUDY-TEST and STUDY-ONLY groups was similar. The decrease in recall between Lists 2 and 3 was slightly greater for the STUDY-TEST groups than for the STUDY-ONLY groups. As can also be seen in Figure 1, there was no main effect for JK experience, $F < 1.0$. Subjects making JKs did not recall more items than subjects not making JKs on each list. Furthermore, there was no JK experience by list interaction, $F < 1.0$.

Simple effects analyses were performed to determine if List 3 recall was affected by study condition and JK experience. Contrary to expectations, the STUDY-ONLY groups did not recall more items on List 3 than the STUDY-TEST groups, $F < 1.0$. Furthermore, the JK groups did not recall more items on List 3 than the CONTROL groups, $F < 1.0$.

The unexpected difference in degree-of-learning between the STUDY-TEST and STUDY-ONLY groups might have been due to differences in recall criterion or willingness to respond with a "guess." To test for recall

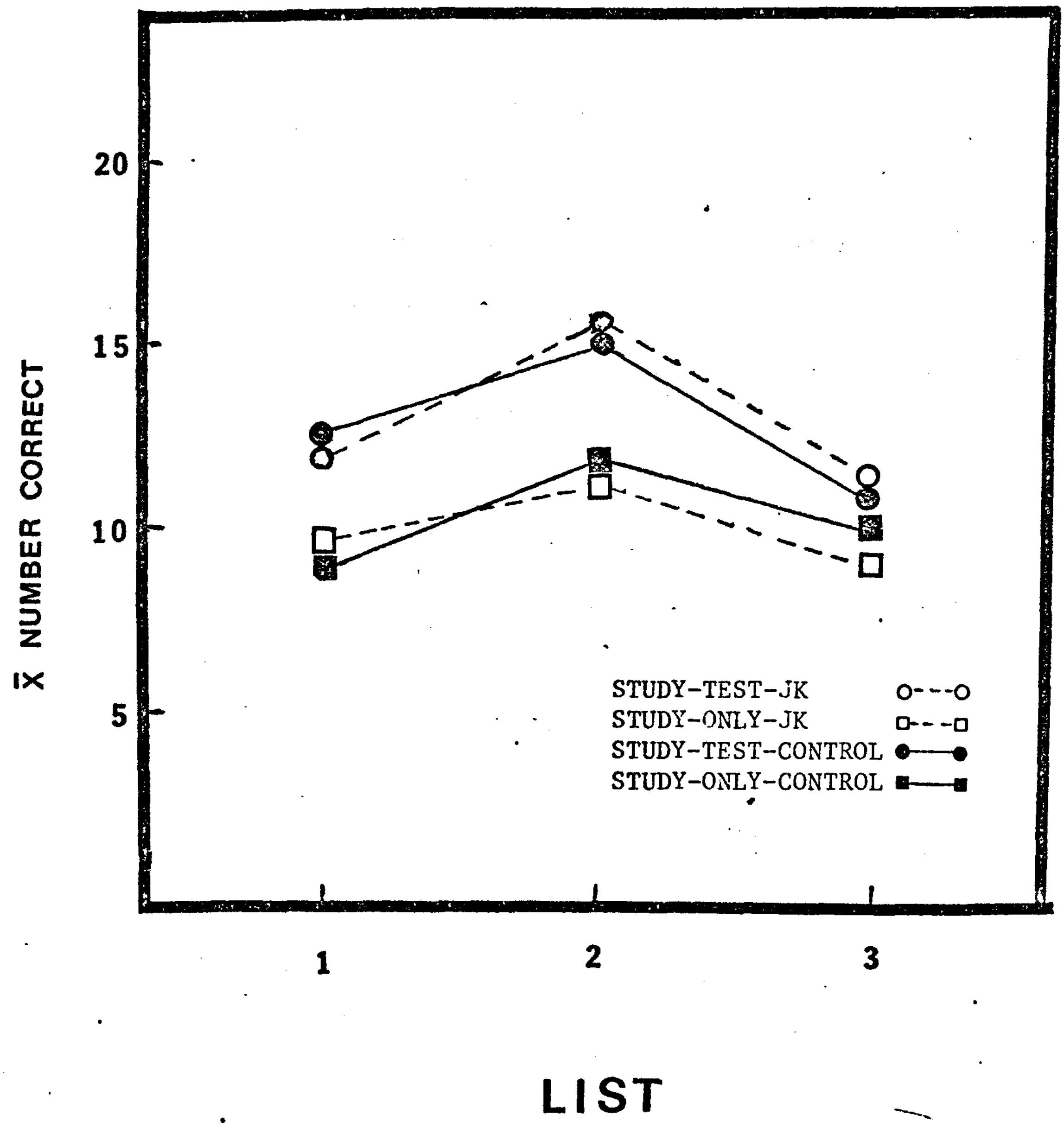

Figure 1. Mean number of correctly recalled items as a function of list.

Table 3

Analysis of Variance Summary Table

For Correct Recall

Source	SS	df	MS	F	
JK experience	2.34	1	2.34	<1.0	
study condition (ST)	621.47	1	621.47	8.94	**
JK by ST	8.40	1	8.40	<1.0	
error (between)	8063.25	116	69.55		
lists (L)	718.11	2	359.05	49.35	***
L by JK	8.90	2	8.90	<1.0	
L by ST	112.21	2	56.10	7.71	*
L by JK by ST	26.94	2	13.47	1.85	
error (within)	1688.12	232	7.28		

* $p < .01$

** $p < .005$

*** $p < .001$

criterion differences, intra-list intrusions and extra-list intrusions were tabulated for each subject. Also, the number of incorrect responses which were "close" to being correct (i.e. sounded like the correct response or highly associated with the correct response) were counted. When the frequency of intra-list intrusions was used as the dependent measure, an analysis of variance revealed neither a significant main effect for JK experience nor a significant main effect for study condition, Fs < 1.0. Intra-list intrusions did increase across lists ($\bar{X}$s = 1.38, 1.23, 1.02 respectively) but the F ratio was only marginally significant, $F\ (2,232) = 2.53$, $p < .07$. None of the interactions reached significance. When extra-list intrusions and "close" responses were tabulated, so few were produced by each subject that the analyses of variance were not performed. Thus it is not likely that group differences in recall criterion existed in the present experiment.

The relationship between recall and JK ratings was examined by comparing the proportion correct recall for items given each of the six JK ratings. For example, for items given a rating of "5", the proportion correctly recalled was calculated. These proportions for each group are plotted in Figure 2. Since not all subjects used each of the six JK categories, no analyses of variance were performed on these proportions. Panels A and

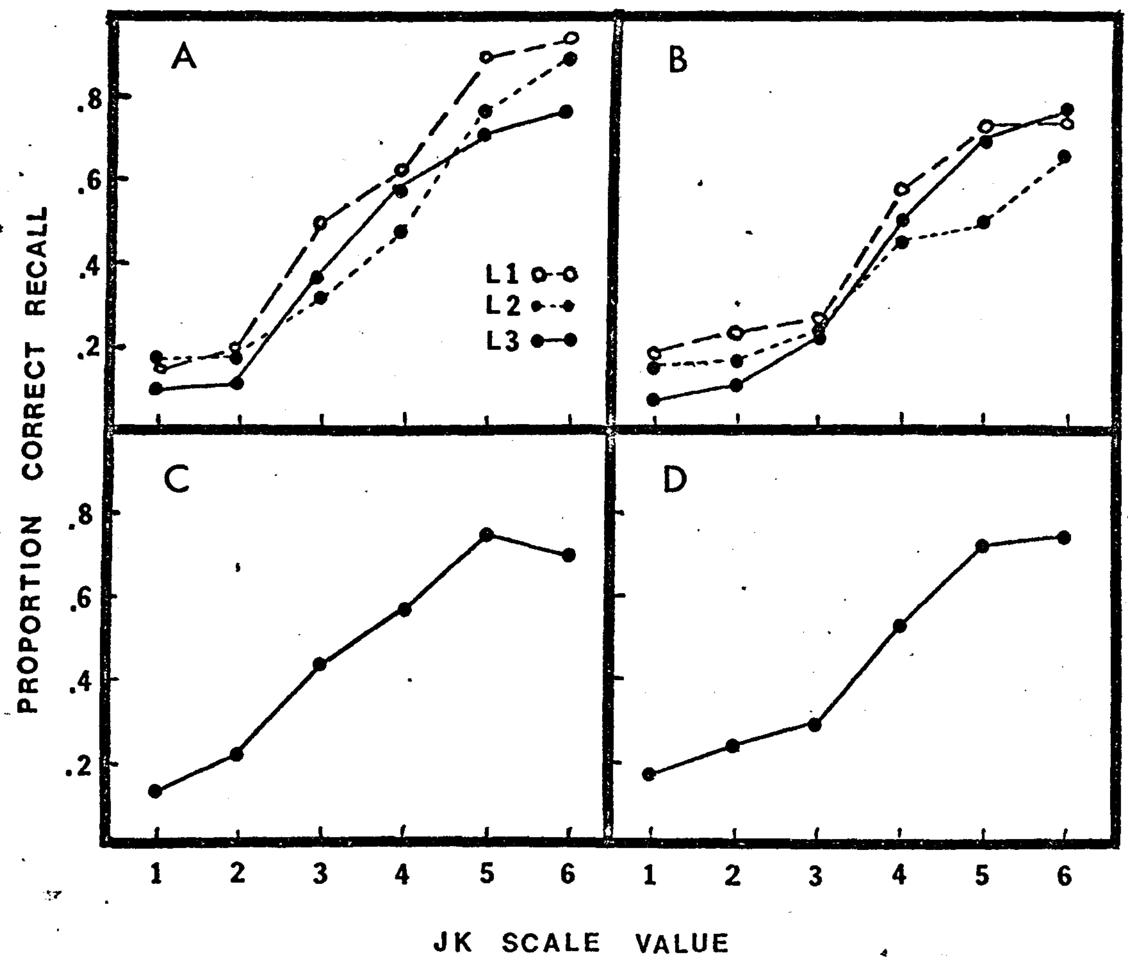

Figure 2. Proportion correct recall as a function of JK rating.

(Panels: A--STUDY-TEST-JK; B--STUDY-ONLY-JK; C--STUDY-TEST-CONTROL; D--STUDY-ONLY-CONTROL.)

B contain data from three JK trials for the STUDY-TEST-JK and STUDY-ONLY-JK groups respectively. The two CONTROL groups (STUDY-TEST-CONTROL & STUDY-ONLY-CONTROL) made JKs only on the third list, and recall proportions for these groups are plotted in Panels C and D respectively. Examination of the curves reveals an increase in proportion recall as a function of JK rating. The higher the rating the greater the probability of recall.

In describing the results for the STUDY-TEST-JK group (Panel A), it appears that ratings 1 and 2 and ratings 5 and 6 can be grouped together. That is, a very low JK rating was likely to be followed by unsuccessful recall, and a very high rating was likely to be followed by successful recall. For this group, ratings of 3 or 4 resulted in approximately .50 probability of recall.

The pattern of results is slightly different for the STUDY-ONLY-JK group. In Panel B it can be seen that ratings 1, 2, and 3 tend to be grouped together in terms of recall probabilities, as are ratings of 4, 5, and 6. Furthermore, the probability of recall for items given high JK ratings (5,6) by the STUDY-ONLY-JK group was not as high as the corresponding probability for the STUDY-TEST-JK group. For each of the control groups, (Panels C & D) the List 3 curves are very similar to the List 3 curves for the corresponding JK groups.

JK Analyses. Measures of JK accuracy should reflect the ability to predict what will be recalled and also the ability to predict what will not be recalled. Furthermore, the measure should not be affected by degree-of-learning. That is, it should be possible for subjects recalling 30% and subjects recalling 70% of the items to produce equal accuracy scores.

To derive an accuracy measure, the JK task was seen as analogous to an absolute-judgment recognition test. That is, the subject could respond "YES" (i.e. 4, 5, or 6 on the JK scale) or "NO" (i.e. 1, 2, or 3 on the JK scale) and, of course, recall or non-recall will follow for an item. The combination of these events results in the contingency table contained in Table 4. A hit (H) represents the case of recall being predicted and successful recall occurring. A correct rejection (CR) indicates that recall was not predicted and, in fact, recall did not occur. These two outcomes represent correct predictions. Errors in prediction were termed false alarms (FA, i.e. recall was predicted but did not follow), and misses (M, i.e. recall was not predicted but did occur). The task was seen as the "detection" of a recallable memory trace. Thus, in a manner similar to signal detection paradigms, the trace could either be detected or fail to be detected when the trace was either actually there or actually not there.

Table 4

Contingency Table for JK Performance and Hypothetical Data

	Recall	Non-recall
"YES" JK	HIT	FALSE ALARM
"NO" JK	MISS	CORRECT REJECTION

	Subject #1		Subject #2		Subject #3	
	R	$\overline{R}$	R	$\overline{R}$	R	$\overline{R}$
YES	11	1	11	11	11	1
NO	1	11	1	1	11	1

Measures of JK accuracy	Subject #1	Subject #2	Subject #3
HITS	11	11	11
P (Hit "YES")	.916	.50	.916
P (Hit RECALL)	.916	.916	.50
HIT-FALSE ALARMS	10	0	10
JK ERRORS	2	12	12
BIAS	0	-.833	+.833

Accuracy can be defined in terms of JK errors, or M + FA. Underwood (1974) has shown that errors in a recognition task can be used as a measure of recognition sensitivity. He reported that errors were highly correlated with d', the sensitivity measure derived from signal detection theory. Furthermore, Underwood has demonstrated that the subject's tendency to respond "YES" or "NO," or his/her criterion can be quantified by the following formula:

$$BIAS = (M - FA)/(M + FA) \tag{1}$$

Underwood reported high correlations between values resulting from Formula 1 and <u>beta</u> as determined by signal detection theory. The bias measure and the sensitivity measure (JK errors) are independent of one another. The value produced by the bias formula would be positive if a subject exhibited a tendency for Ms (i.e. prediction of non-recall when recall followed), and a negative value would indicate a tendency for FAs (i.e. prediction of recall when non-recall followed). Therefore, in the present experiment, JK errors was used as the primary measure of accuracy, and the bias score (Formula 1) was used to examine differences in criterion.

In addition to JK errors, four other dependent measures were considered for use: (1) total Hs; (2) H minus FAs; (3) p (H/"YES"), and (4) p (H/Recall). By examining data from three hypothetical subjects (see lower half of

Table 4), these measures can be contrasted, and the reasons for using JK errors (M + FA) can be clarified.

If Hs were the sole determinant of JK accuracy, it can be seen in the examples shown that all three hypothetical subjects would perform equally well even though the subjects differ in the frequency of FAs and Ms. Further, this measure would be correlated with overall level of recall, since greater recall would provide more possibilities for Hs. Also, this measure would be clearly inadequate as a measure of prediction accuracy since the ability to make CRs is not reflected by the score.

Hart (1965) used H minus FAs as a measure of accuracy in an analogous paradigm. This is a common "correction for guessing" strategy used in recognition memory tasks. The FA rate is considered a guessing rate, and the hit rate is reduced to account for guessing. As seen in Table 4, Subjects #1 and #3 would result in equal accuracy scores if this measure were used. Subject #3 performed less accurately when NO judgments are considered. Thus, this measure is inadequate since the accuracy of "NO" judgments does not influence the score.

A third possible measure is the number of Hs relative to the number of "YES" JK ratings, or:

$$p\ (H/\text{"YES"}) = H/(H + FA) \qquad (2)$$

If this measure were used, again Subjects #1 and #3 in Table 4 would produce equal accuracy scores. As was seen

above, these subjects differ in their ability to accurately make NO judgments.

A fourth measure to be considered is the number of Hs relative to the number of correctly recalled items, or:

$$p\ (H/Recall) = H/(H + M) \tag{3}$$

Using this measure of accuracy, subjects #1 and #2 from Table 4 would be equally accurate. This measure does not detect differences in prediction for words which are not recalled. Misses rather than FAs are crucial to this measure, and it can be seen that Subject #2 produced many FAs which did not affect the accuracy score.

Table 4 also contains the accuracy score as measured by M + FA and the criterion scores produced by Formula 1 for the three subjects. By examining both measures, differences in JK performance can be determined. Subject #1 is most accurate, and Subjects #2 and #3 are equally accurate. However, when criterion or bias scores are examined for the latter two subjects, different response tendencies can be observed. Therefore, the present experiment used JK errors as the primary accuracy measure and Formula 1 as the measure of bias.

In the present experiment, JK accuracy should be independent of overall degree of learning. As was stated above, the STUDY-TEST groups recalled more items than the STUDY-ONLY groups, and thus if the JK errors measure was sensitive to degree-of-learning, any differences in JK

accuracy between groups may be due to the fact that unequal degree of learning was obtained. To be sure that this was not the case, within each group, fast and slow learners were determined. The analyses of variance included the fast-slow distinction as a factor. If no main effect for fast-slow learners is obtained and if no interactions between fast-slow learners and any other factor is obtained, then it can be concluded that the JK errors measure was independent of degree-of-learning. Fast and slow learners were determined by the total number recalled on the first two lists. Within a group, the 15 subjects above the median constituted the fast learners, and the 15 subjects below the median constituted the slow learners. When more than one subject scored the median value, the assignment was made randomly in order to assure an equal number of subjects in each group.

The mean number of JK errors for each of the four groups (collapsed across fast and slow learners) is displayed in Figure 3. Two separate analyses were done on these data. First, only those groups which made JKs on all three lists were considered (i.e. the JK groups). Table 5 contains the summary table for the 2 (study condition) by 2 (fast-slow learners) by 3 (lists) repeated measures analysis of variance using JK errors as the dependent measure. A significant main effect for study condition was obtained, $\underline{F}$ (1,56) = 10.07, $\underline{p} < .001$.

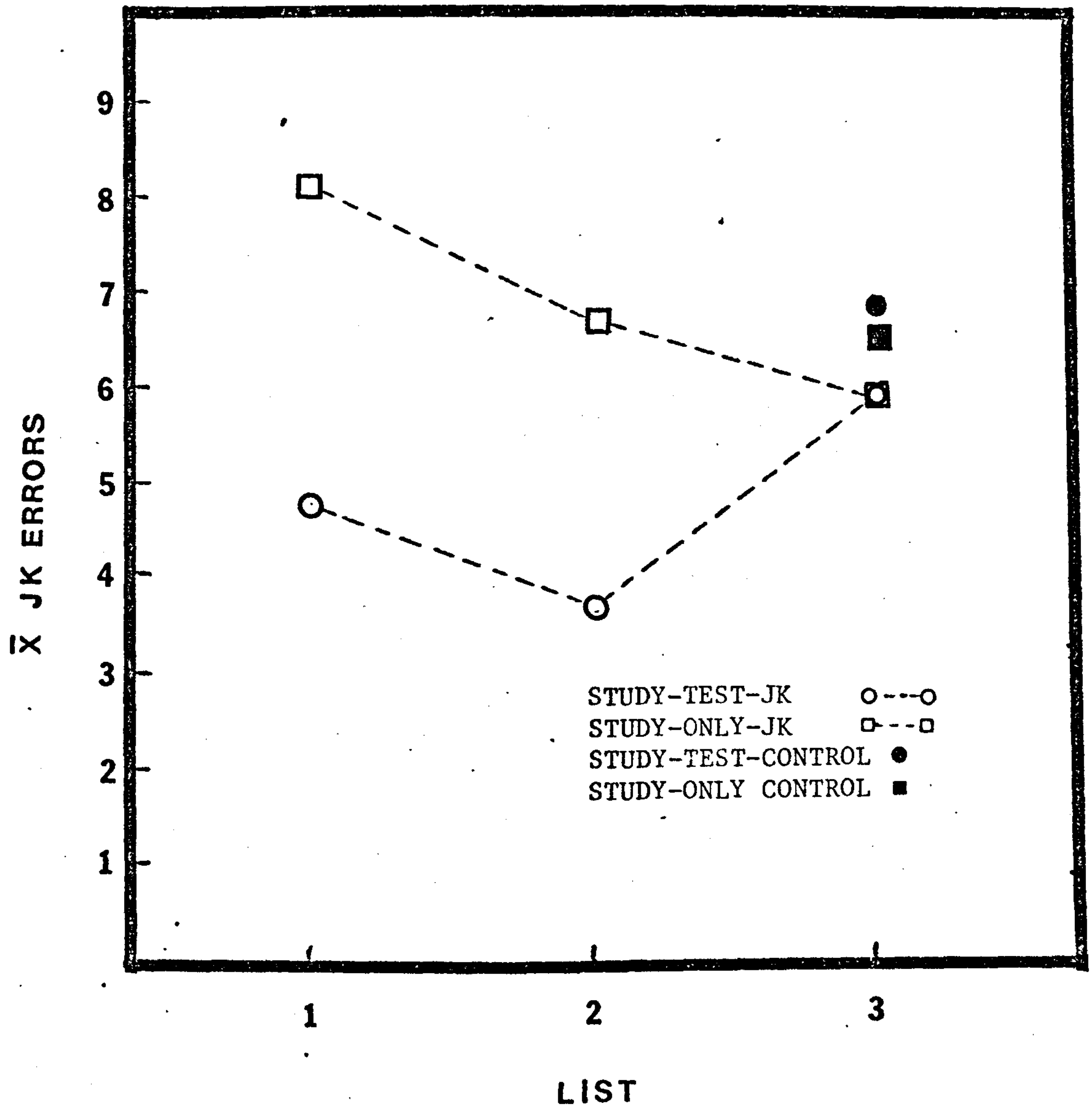

Figure 3. Mean number of JK errors as a function of list. (Control group JK data is for List 3 only.)

Table 5

Analysis of Variance Summary Table for JK Errors (MISSES + FALSE ALARMS) for the STAND-JK And STUDY-JK Groups for All Three Lists

Source	SS	df	MS	*f*
Study cond. (ST)	186.05	1	186.05	10.07 **
Ability (A)	18.05	1	18.05	<1.0
ST by A	.93	1	.93	<1.0
error (between)	1034.35	56	18.47	
Lists (L)	48.14	2	24.07	5.44 *
L by ST	90.83	2	45.42	10.26 **
L by A	16.03	2	8.02	1.81
L by ST by A	7.08	2	3.54	<1.0
Error (within)	495.88	112	4.43	

* $p < .01$

** $p < .001$

Fewer JK errors were made by the STUDY-TEST-JK group ($\bar{X}$ = 4.88) than by the STUDY-ONLY-JK group ($\bar{X}$ = 6.92). A significant main effect for lists was also obtained, F (2,112) = 5.44, p < .01. Errors generally declined across lists. Furthermore, the study condition by list interaction was highly significant, F (2,112) = 10.26, p < .001. Inspection of Figure 3 reveals that the mean JK errors for the STUDY-TEST-JK and the STUDY-ONLY-JK groups decreased at equal rates across the first two lists. Furthermore, it can be seen that the STUDY-ONLY-JK continued to show a decrease in errors between Lists 2 and 3. For the STUDY-TEST-JK group, on the other hand, a sharp increase in JK errors between List 2 and 3 resulted. Finally, the main effect for fast-slow learners failed to reach significance, F < 1.0. The mean JK errors for the fast learners were not different than for the slow learners. Also, the fast-slow factor did not interact with study condition (F < 1.0) or lists, F (2,112) = 1.81, p > .10.

A second JK analysis used data from all four groups. A 2 (JK experience) by 2 (study condition) by 2 (fast-slow learners) analysis of variance was performed to examine JK accuracy (M + FA) for List 3 only. The source table is contained in Table 6. As can be seen, none of the F ratios reached significance. Study condition for List 1 and 2 apparently had no effect on List 3 JK accuracy. Furthermore, experience with the JK task, regardless of study

Table 6

Analysis of Variance Summary Table

For JK Errors on List 3 Only

Source	SS	df	MS	F
JK Experience (JK)	19.20	1	19.20	1.86
Study cond. (ST)	.83	1	.83	< 1.0
Ability (A)	1.20	1	1.20	< 1.0
JK by ST	1.20	1	1.20	< 1.0
JK by A	.30	1	.30	< 1.0
ST by A	3.33	1	3.33	< 1.0
JK by ST by A	4.03	1	4.03	< 1.0
error	1157.84	112	10.34	

condition, had no effect on JK accuracy. The study condition by JK experience interaction also failed to reach significance. Also, none of the F ratios involving fast-slow learners reached significance.

The fast-slow learner factor used in the previous two analyses of variance produced no significant effects. This suggests that the JK errors measure of accuracy is not dependent on degree-of-learning, and gives support to the validity of this measure.

Analysis of JK response bias. A second important aspect of the difference in JK performance between the STUDY-TEST and STUDY-ONLY groups concerns response biases. That is, the STUDY-TEST-JK group and the STUDY-ONLY-JK group may have had different tendencies to respond YES or NO on the JK scale. In order to make strong conclusions about JK accuracy (sensitivity) differences in bias must be examined.

The bias score derived from Formula 1 can be used to determine response tendencies. This measure is theoretically independent of the JK errors (sensitivity) measure; and in general, this measure reveals differences in criterion or "cutoff" levels. That is, it measures what level of confidence or what degree of perceived (memory) strength is necessary in order for a subject to respond YES on the JK scale. A positive bias score indicates a realtively "strict" or "conservative" criterion (i.e. more Ms than

FAs were observed). In this case, the subject must be relatively sure of the presence of an item in memory in order for a YES JK to be made. A negative value, on the other hand, indicates a less "strict" or more liberal criterion, and less confidence is necessary in this case in order to respond YES on the JK scale.

Therefore, differences in JK performance could either be due to differences in the true ability to detect or assess an item's strength in memory or be due to differences in the criterion above which an item's presence in memory is acknowledged. By examining bias as well as JK errors greater understanding of the JK performance can be attained.

Table 7 contains the mean bias score (from Formula 1) for the STUDY-TEST-JK and STUDY-ONLY-JK groups for all three lists. An analysis of variance on the data from Table 9 revealed only a marginally significant main effect for study condition, $\underline{F}\ (1,56) = 3.94$, $\underline{p} < .1$. It can be seen that the STUDY-ONLY-JK group had a slightly greater tendency to commit FAs than did the STUDY-TEST-JK group. That is, the criterion for the latter group was more strict than that for the former group. The study condition by list interaction did not reach significance, $\underline{F}\ (2,112) = 1.63$, $\underline{p} > .1$. When List 3 bias scores were examined for all four groups, no differences between groups was observed.

Table 7

Mean Bias Score as a Function of Group and List

	List			
	1	2	3	4
STUDY-TEST-JK	-.03	.09	-.16	-.03
STUDY-ONLY-JK	-.47	-.20	-.29	-.32
STUDY-TEST-CONTROL			.00	
STUDY-ONLY-CONTROL			-.14	

Additional JK Analyses. In the JK analyses reported thus far, JK errors have been of concern. In order to more fully understand differences in JK performance between groups, correct JK responses (i.e. Hs and CRs) must also be examined. Thus separate analyses of Hs, Ms, FAs and CRs were performed in order to describe the pattern of JK outcomes across lists. Furthermore, the mean number of YES JKs (i.e. ratings of 4, 5, or 6) for each group was tabulated, since this measure provides a context within which changes in Hs or FAs can be explained. A YES judgment can only result in an H or FA, and conversely, a NO judgment can only result in an M or CR. Thus the total number of YES JKs (or NO JKs) must provide a "baseline" around which changes in the specific JK outcomes can be explained. Later in this paper, this pattern of Hs, Ms, FAs, and CRs, will be used to infer differences in the ability "to know what is known" and "to know what is not known."

Table 8 contains the mean number of Hs, Ms, FAs, and CRs across lists for the STUDY-TEST-JK and STUDY-ONLY-JK groups as well as the means for List 3 for the STUDY-TEST-CONTROL and STUDY-ONLY-CONTROL groups. Also, the mean number of "YES" JKs for each list for each group is contained in Table 9.

Five separate 2 (study condition) by 2 (fast-slow learners) by 3 (lists) analyses of variance were performed

Table 8

Mean Number of Hits, Misses, False Alarms, and Correct Rejections as a Function of Lists

	List 1	List 2	List 3	$\overline{X}$
Hits				
STUDY-TEST-JK	9.76	13.83	8.57	10.72
STUDY-ONLY-JK	7.76	8.80	7.43	8.00
STUDY-TEST-CONTROL			7.76	
STUDY-ONLY-CONTROL			7.70	
Misses				
STUDY-TEST-JK	2.16	2.07	2.40	2.21
STUDY-ONLY-JK	1.90	2.43	1.66	2.00
STUDY-TEST-CONTROL			3.03	
STUDY-ONLY-CONTROL			2.70	
False Alarms				
STUDY-TEST-JK	2.76	1.76	3.50	2.67
STUDY-ONLY-JK	6.23	4.26	4.26	4.91
STUDY-TEST-CONTROL			3.86	
STUDY-ONLY-CONTROL			3.83	
Correct Rejections				
STUDY-TEST-JK	9.30	6.33	9.53	8.39
STUDY-ONLY-JK	8.10	8.50	10.63	9.08
STUDY-TEST-CONTROL			9.33	
STUDY-ONLY-CONTROL			9.76	

Table 9

Mean Number of YES Judgments as a Function of List

	List			
	1	2	3	$\bar{X}$
STUDY-TEST-JK	12.53	15.60	12.06	13.40
STUDY-ONLY-JK	14.00	13.06	11.70	12.92
STUDY-TEST-CONTROL			11.63	
STUDY-ONLY-CONTROL			11.53	

in order to describe the pattern of Hs, Ms, FAs, CRs, and YES JKs for the two JK groups.

The pattern for Hs will be described first. The analysis of variance yielded a significant study condition by list interaction, F (2,112) = 10.53, $p < .001$. A simple effects analysis revealed that for the STUDY-ONLY-JK group, Hs were constant across lists, F (1,112) = 2.80, $p > .10$. For the STUDY-TEST-JK group, the simple effects analysis revealed significant differences in Hs across lists, F (1,112) = 42.05, $p < .001$. Hits increased across the first two lists and then decreased on List 3 for the STUDY-TEST-JK group.

When Ms were used as the dependent measure (see Table 8) there were no differences between the STUDY-TEST-JK and STUDY-ONLY-JK groups across lists, $F < 1.0$. Therefore, differences in Ms could not have accounted for differences in JK performance.

The third analysis of variance concerned FAs. The study condition by list interaction reached significance, F (2,112) = 6.65, $p < .01$. A simple effects analysis revealed differences across lists for the STUDY-ONLY-JK group, F (1,112) = 6.91, $p < .01$. For this group, FAs were high on List 1, and then decreased and remained constant for the second and third lists. The simple effects analysis also revealed differences across lists for the STUDY-TEST-JK group, F (1,112) = 4.05, $p < .05$.

For this group, FAs were initially lower than for the STUDY-ONLY-JK group but then increased on List 3. Therefore, changes in FAs are likely to account for differences in JK performance.

The fourth analysis of variance used the frequency of CRs as the dependent measure. The study condition by list interaction was significant, $F\ (2,112) = 6.93$, $p < .01$. Again, a simple effects analysis revealed differences in CRs across lists for the STUDY-ONLY-JK group, $F\ (1,112) = 7.38$, $p < .01$. For this group, CRs steadily increased across lists. For the STUDY-TEST-JK group, CRs decreased from List 1 to List 2 and then increased on List 3, $F\ (1,112) = 12.67$, $p < .01$.

The fifth analysis used the frequency of YES judgments as the dependent measure. There was no significant main effect for study condition, $F < 1.0$. Thus, even though the STUDY-TEST-JK group recalled more items than the STUDY-ONLY-JK group, the mean number of YES JKs did not differ for these two groups. The study condition by list interaction reached significance, $F\ (1,112) = 4.40$, $p < .05$. YES JKs consistently decreased across lists for the STUDY-ONLY-JK group. For the STUDY-TEST-JK group, YES judgments increased from List 1 to List 2 but then decreased on List 3.

The frequency of Hs, Ms, FAs, CRs, and YES JKs was also examined for List 3 taking into account the CONTROL

groups as well as the JK groups. No significant main effects or interactions were obtained when Hs, FAs, CRs, or YES JKs were used as dependent measures. However, when Ms were considered, a significant main effect for JK experience was obtained, F (1,112) = 4.39, p < .05. That is, the CONTROL groups committed more Ms ($\bar{X}$ = 2.86) than did the JK groups ($\bar{X}$ = 2.03). Thus, while JK accuracy on List 3 for the JK groups did not differ from that for the CONTROL groups, a difference in the type of errors produced was obtained.

Testing Effects. It has been suggested above that the JK task is analogous to a signal detection paradigm. That is, an item in memory can be reported to be either present or absent when, in fact, that item was either present or absent. The analogy is not perfect, however. That is, in a signal detection task, the experimenter unambiguously controls the presence or absence of the signal, but in the JK task, the presence of an item is inferred from the subject's later recall performance. Furthermore, since a considerable time lapse occurs between the JK rating and testing of an item, the item could be forgotten during this interval. Thus, suppose a subject assigned a high JK rating to an item, and suppose that the item was, in fact, available. Interference from other list items could result in the item being unavailable at time of test, and thus the subject's

prediction, in this case, would be incorrect. The opposite situation could also occur. Facilitation from other list items could cause an item, previously rated as unavailable (i.e. a NO JK), to become available at the time of test. Again, the JK in this case would be considered inaccurate.

To consider these possibilities, the change in the frequency of Ms and FAs within a test list was examined. For this analysis, the number of items intervening between the rating and the testing of each pair (i.e. lag) was tabulated. For each test form, the distribution of lags was divided into quartiles. The mean proportion of Ms and FAs as a function of lag quartile is plotted in Figures 4 and 5 respectively. Two 2 (study condition) by 3 (lists) by 4 (lags) repeated measures analyses of variance were performed on these data. For Ms, there was a significant main effect for lags, $\underline{F}$ (3,174) = 8.66, $\underline{p} < .01$. Overall, there was a decrease in Ms across lags. Neither the main effect for study condition nor any of the interactions reached significance. Furthermore, it can be seen in Panels C and D of Figure 4 that for the two CONTROL groups, List 3 Ms decreased as a function of lag. ($\underline{F}$ (3,348) = 8.95, $\underline{p} < .01$.)

For FAs, the analysis of variance revealed a significant main effect for lags, $\underline{F}$ (3,174) = 12.65, $\underline{p} < .01$. A general increase across lags was observed for FAs. The study condition by list by lag interaction reached

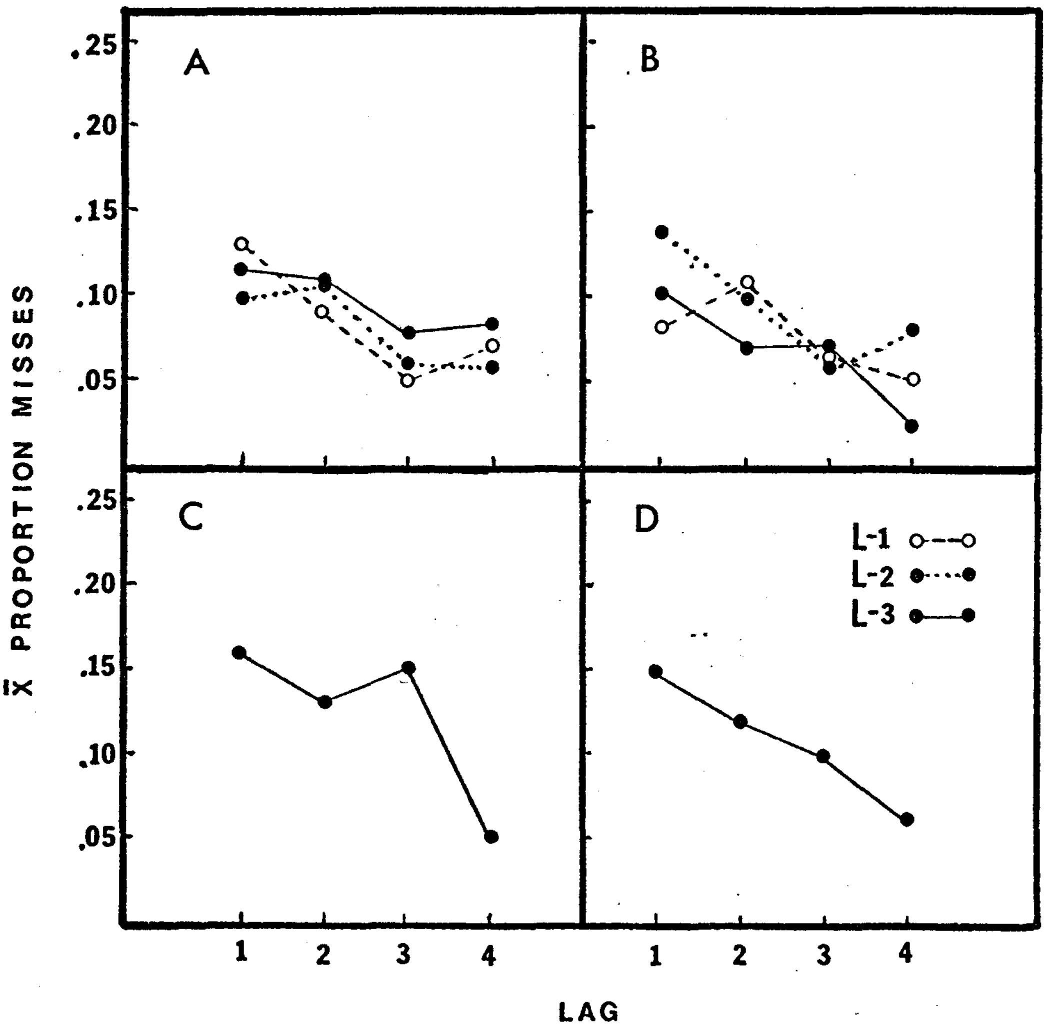

Figure 4. Mean proportion misses as a function of lags and list.

(Panels: A--STUDY-TEST-JK; B--STUDY-ONLY-JK; C--STUDY-TEST-CONTROL; D--STUDY-ONLY-CONTROL)

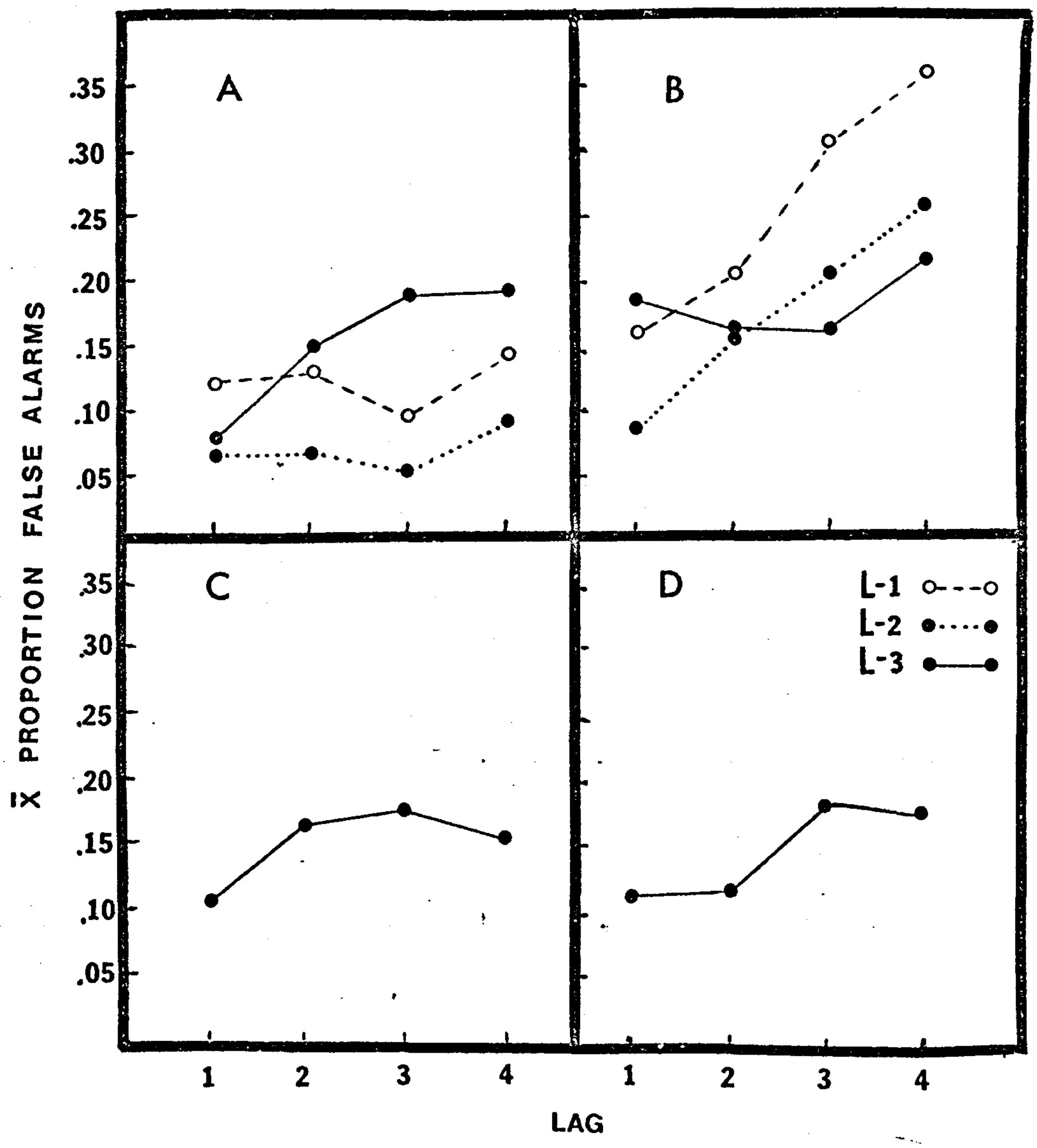

Figure 5. Mean proportion false alarms as a function of lags and list.

(Panels: A--STUDY-TEST-JK; B--STUDY-ONLY-JK; C--STUDY-TEST-CONTROL; D--STUDY-ONLY-CONTROL)

significance indicating that the change in FAs across lags was not identical for the STUDY-TEST-JK and STUDY-ONLY-JK groups across the three lists, (F (6,348) = 4.07, $p < .01$.) A simple effects analysis revealed that for the STUDY-TEST-JK group, FAs did not differ across lags for the first list, F (1,464) = 2.21, $p > .10$, nor for the second list, $F < 1.0$. However, for the third list, a significant linear trend across lags was obtained, F (1,484) = 5.84, $p < .01$. For the STUDY-ONLY-JK group, the simple effects analysis revealed differences in FAs as a function of lags on List 1 and 2, but not on List 3. A significant linear trend was obtained for List 1, F (1,464) = 18.46, $p < .001$, and for List 2, F (1,464) = 8.95, $p < .01$. No difference in linear trend was observed for the first two lists, $F < 1.0$. In Figure 5, it can be seen that when the two CONTROL groups (Panels C & D) were included in the analysis. FAs generally increased as a function of lag on List 3, F (3,348) = 3.39, $p < .05$. It must be noted that these results are partially dependent on the total number of recalled items as a function of lag. Table 10 contains the proportion correct recall as a function of lag. It can be seen that as lag increases, proportion correct recall decreases. Thus, conclusions about FAs and Ms as a function of lag should be interpreted in light of the fact that recall also decreases as a function of lag.

Table 10

Proportion Correvt Recall as a Function

Of List and Lag

List	Lag			
STUDY-TEST-JK	1	2	3	4
1	.533	.544	.438	.472
2	.711	.644	.643	.649
3	.561	.444	.405	.466
STUDY-ONLY-JK				
1	.527	.405	.344	.333
2	.610	.444	.410	.406
3	.478	.355	.372	.311
STUDY-TEST-CONTROL				
3	.605	.361	.433	.399
STUDY-ONLY-CONTROL				
3	.488	.477	.427	.338

DISCUSSION

The results of this experiment show that, in general, subjects can predict which PAs will be recalled on an upcoming test. The distinction between the STUDY-TEST-JK and STUDY-ONLY-JK groups allows several specific conclusions to be made regarding the ability to judge what is known.

The Role of Test Trials. In a PA task, JK accuracy (M + FA) appears to benefit from knowledge of previous test trial performance. The results obtained in this experiment provide two lines of evidence in support of this conclusion. First, the STUDY-TEST-JK group made more accurate predictions on each of the first two lists than did the STUDY-ONLY-JK group. Secondly, the STUDY-TEST-JK group revealed a marked increase in JK errors on the third list when test trials before the JKs were removed. That is, from the increase in errors, it can be inferred that the subjects in the STUDY-TEST-JK group had developed a strategy for making JKs which was at least partially based on the success of previous test trial attempts.

The conclusion that JKs are facilitated by the presence of preceeding test trials is partially dependent on the assumption that subjects can remember what they have

recalled earlier, and some evidence exists which supports this assumption. Masur et al. (1973) reported that young children could accurately recognize items they had recalled on an immediately preceeding test trial. Using adult subjects, Lockhart (1975) presented a series of 10 free recall lists followed by a final recognition test on the words from all 10 lists. Subjects were then asked to circle any recognized items which they remembered having recalled. Accuracy of discriminating recalled from non-recalled items was found to be quite high. Thus, given this evidence, in the present study it can be reasonably assumed that subjects in the STUDY-TEST-JK group could distinguish those items that they had recalled from those that they had failed to recall on the test trials preceeding the JK trial. Therefore it is likely that subjects were implicitly reasoning "since I got it right on the last two test trials, I should get it right again."

The Role of Familiarity. As familiarity with task demands increases, JK accuracy increases. The STUDY-TEST-JK group revealed an increase in JK accuracy across the first two lists, and for the STUDY-ONLY-JK group JK accuracy increased across all three lists. From the performance of the two CONTROL groups, it appears that the familiarity with PA learning, rather than the familiarity with the JK task itself, is crucial. That is, on the third list, the two CONTROL groups that had no previous JK task experience,

performed as accurately on the third list as did the two JK groups. Thus while JK accuracy can be said to increase with practice, actual JK task experience is not as important as is experience with learning under the constraints of the task at hand.

Specific JK Performance Differences. In order to more fully understand differences in JK performance between the STUDY-TEST-JK and the STUDY-ONLY-JK groups, two specific questions must be answered. First, both these groups revealed a decrease in JK errors across the first two lists, and it should be determined whether, for both these groups, the decrease occurred for the same reason. That is, did the same pattern of Hs, Ms, FAs, and CRs result for both these groups on the first two lists? The second question concerns the reasons for the increase in JK errors on the third list for the STUDY-TEST-JK group.

The results relevant to the improvement in JK accuracy by the STUDY-ONLY-JK group will be summarized first. From the analyses of JK outcomes, the following four findings can be noted: (1) YES JKs declined across lists; (2) Hs and Ms remained constant; (3) FAs generally decreased across lists, and (4) CRs increased across lists. Given this pattern of results, any changes in JK performance must have been due to FAs and CRs, since the other outcomes did not change. Furthermore, since YES judgments decreased across lists, differences in the accuracy of

NO judgments were likely to account for the JK improvement. That is, FAs indicate the lack of the ability to detect what is not in memory; and CRs reflect the ability to successfully indicate that an item is not in memory. Since FAs were decreasing while Hs remained constant, and since CRs increased while Ms remained constant, the ability to know what is not known was apparently accounting for the improvement in JK accuracy for the STUDY-ONLY-JK group.

With respect to the increase in JK accuracy across the first two lists shown by the STUDY-TEST-JK group, the following findings should be considered: (1) YES JKs and Hs increased; (2) Ms did not change; (3) FAs decreased slightly across the first two lists, and (4) CRs decreased sharply across the first two lists. From this pattern of results, it can be shown that Hs are responsible for the increase in accuracy across the first two lists. The reasoning is as follows. Since FAs decreased while Hs increased, the ability to know what is known (as opposed to what is not known) must have improved. The ability to know what is not known was less likely to account for the improvement because CRs decreased and Ms remained constant. In fact, the ability to predict what was not known was probably constant across the first two lists for this group. That is, since the overall frequency of NO judgments declined, and since Ms did not change, a decline in

CRs would have to occur. This would not necessarily imply a decrease in the accuracy of NO judgments.

Correlations support the conclusion that Hs are responsible for the increase in accuracy. That is, if the correlation between Hs and JK errors increases from List 1 to List 2 for the STUDY-TEST-JK group, Hs are likely to be implicated in the improvement. The correlation increased from -.41 to -.49 on List 2.

Thus for the STUDY-TEST-JK group, the increase in JK accuracy across the first two lists was probably due to an increase in the ability to know what was known; while for the STUDY-ONLY-JK group, the improvement was due to an increase in the ability to know what was not known.

The STUDY-TEST-JK group showed a marked increase in JK errors between Lists 2 and 3. This increase corresponded with the removal of test trials before the JK task on the third list. Given the analyses of the JK outcomes, the following results relevant to this increase in errors can be considered: (1) YES JKs and Hs decreased; (2) FAs slightly increased; (3) CRs increased across lists, and (4) Ms did not change. These observations are made relative to List 2 JK performance. The decline in YES judgments (or increase in NO judgments) was accompanied by an increase in CRs (with Ms remaining unchanged). Thus, on the one hand, accuracy of NO judgments cannot account for the decrease in accuracy because, in fact, the accuracy

of NO judgments improved on List 3. On the other hand, FAs are likely to account for the increase in errors. That is, YES JKs and Hs both declined but not at the same rate. Given that a YES JK can only result in an H or FA, it must be concluded that FAs would have to increase since the decline in YES JKs was <u>less</u> than the decline in Hs. This suggests that a greater likelihood of FAs was causing the decrease in accuracy. Correlations support this conclusion. The correlation between JK errors and FAs rose from .48 on List 2 to .55 on List 3. An increase in FAs indicates a decrease in the ability to know what is not known.

<u>Bias Scores and Testing Effects</u>. Additional differences in the JK performance of the STUDY-TEST-JK and STUDY-ONLY-JK groups can be pointed out by examining the criterion or bias scores and the testing (or lag) effect data. Furthermore, it should be determined whether or not the reasons given for JK improvement for these two groups are consistent with bias scores and lag effects.

Conclusions about the relationship between specific patterns of JK outcomes (i.e. Hs, Ms, FAs, and CRs) and bias scores should be prefaced by a brief discussion of the possible redundancy (or interdependence) of these measures. Bias scores reflect a relative tendency for Ms or FAs, and from this measure, either a "strict" or "lax" criterion is inferred. Theoretically, this measure is

distinct from the <u>absolute</u> level of FAs or Ms considered separately. Thus a subject exhibiting many FAs would not show a lax criterion if he also exhibits an equal number of Ms. However, this assumption cannot be unequivocally accepted given the fact that, in the present experiment, FAs correlated -.71 with bias scores, and Ms correlated .78 with bias scores. Thus, the following statements must be interpreted cautiously, knowing that the bias scores and the Ms and FAs are not totally independent.

For the STUDY-ONLY-JK group, it can be concluded that subjects made JKs using a relatively "lax" criterion. That is, since there was a greater tendency for FAs for the STUDY-ONLY-JK group than for the STUDY-TEST-JK group, it can be inferred that less "strength" of the memory trace was required for a YES judgment for the former group than for the latter group. As accuracy of NO judgments increased across lists, one would expect to observe a more and more "strict" criterion being employed. An increase in the accuracy of NO judgments implies an increase in the ability to asses an item's memory strength as inadequate for later retrieval. This can be seen as a "raising" of the level of memory strength necessary for a YES JK response. In fact bias scores did become less negative (or more strict) on the second and third lists relative to the first list. Thus, for the STUDY-ONLY-JK group, the criterion data are consistent with the conclusion

that accuracy of NO judgments accounted for the improvement in JK accuracy across lists.

As was stated earlier, an increase in FAs within a list as lag (number of intervening events between rating and testing) increases indicates that other list items interfered with items given YES JKs and thus caused these items to be unavailable at time of test. If a lax criterion were being used, then one would expect greater lag effects than if a strict criterion were being used. That is, items with low levels of memory strength are more likely to be interferred with than items with higher levels of memory strength. For the STUDY-ONLY-JK group, lag effects for FAs were greater on List 1 and List 2, and then became less extreme on the third list. This is as expected since the criterion became more strict across lists. A strict criterion indicates that only well-learned items are given YES JKs, and well-learned items are less likely to suffer from interference across time.

The improvement in JK accuracy across the first two lists shown by the STUDY-TEST-JK group was due to an increase in Hs. An H implies that the perceived level of memory strength was adequate to assure later recall. This should correspond with a rather "strict" JK criterion; and relative to the STUDY-ONLY-JK group, a more strict criterion was employed on the first two lists by the STUDY-TEST-JK group. Furthermore, lag effects for FAs were minimal

across the first two lists for the STUDY-TEST-JK group. This indicates that, in general, only well-learned items were given YES JKs.

On List 3, the STUDY-TEST-JK group revealed a decrease in JK accuracy and it was suggested that FAs or inaccurate YES JKs accounted for this poorer performance. That is, subjects in this group more frequently gave YES JKs to items which were eventually not recalled when test trials were removed than when test trials were present before the JK trial. As would be expected, this corresponded with a more lax criterion on the third list than on the first two lists. Again, the lag effect data for FAs for this group is consistent with the lowering of the criterion. For the STUDY-TEST-JK group, there was a greater lag effect for FAs on the third list than on either of the first two lists.

For all groups, Ms generally decreased as a function of lags. Two possible interpretations are available for this effect. First, subjects could have developed a strategy of withholding responses or of selectivity ignoring items for which they remembered giving NO JKs. Thus, if an item was given a NO judgment, non-recall would result in an accurate prediction. Perhaps, the subjects realized this only as they progressed through the test list, and thus, the greater the lag, the greater the likelihood of their having developed a strategy of selective rehearsal or withholding of known responses. This interpretation

can be rejected for several reasons. First, subjects were instructed that under the "game" rules, they would always be penalized for non-recall regardless of the prediction. In fact, if they were trying to "beat" the game, Ms should have increased since an M always resulted in gaining points and a CR (i.e. deliberate withholding or selective rehearsal of an item given a NO JK would result in a CR) always resulted in losing points. A more convincing argument against the notion that subjects were withholding or selectively rehearsing items and thereby causing the lag effects is that the lag effects for Ms were constant across all three lists. If subjects were becoming aware of these strategies with increasing JK task experience, then the lag effect should have been "learned" by the time the subjects reached the third list. Thus, one would not expect lag effects for Ms on the third list. Furthermore, if a strategy of withholding or selectively rehearsing were being employed, memory for an items rating at the time of test would be necessary. The shorter the interval between rating and testing, the more likely subjects would be to remember which rating was assigned. Thus, at short lags, one would expect fewer Ms than at longer lags. The opposite results were observed.

Lag effects for Ms were also obtained in a similar experiment by Arbuckle and Cuddy (1969) and the procedures they employed were also designed to prevent selective

rehearsal or withholding. In the first experiment reported by these authors, lag effects were obtained with highly practiced subjects who were specifically warned against the use of withholding or selectively rehearsing items. In their second experiment, they used a probed recall technique, such that the subject did not know which of five presented items would be tested. Many Ms were observed when the probed item was the last item in the list. Thus these authors also found lag effects for Ms when the possibility of withholding and selective rehearsal were removed.

A more acceptable interpretation of the lag effects for Ms concerns facilitation and interference from other list items and the subject's perception of the task. It should be noted that, in general, items with short lags occurred relatively late in the rating list and relatively early in the test list. Furthermore, an M means that an item judged to be of insufficient strength was in fact of sufficient strength to assure recall. The misjudgment could be due to either a "misperception" of an item's strength or an unexpected increase of an item's strength. First consider "misperceptions" of memory strength. A subject had no way of telling which items would be tested and thus he might have been judging an item's memory strength as sufficient or insufficient to assure recall at the "expected" lag. The expected or average lag would

have been about 24 items. Thus, an item judged to be of insufficient strength at the expected lag may have been of sufficient strength at lags less than the expected 24. Thus, if Ms were to occur, they would most likely occur at the shortest lags. The second reason for misjudgments could be that some unexpected increase in an item's memory strength occurred. This increase could be said to be due to facilitation from other list items. However, since over time, interference would be causing a decrease in an items availability, an item would become available because of facilitation only at the shorter lags. That is, if an item's strength is low to begin with and is decreasing across time, any events which could increase that item's strength and would be most effective when strength is highest. In other words, facilitation would be most likely at the shortest lags. This was likely to account for the lag effects for Ms.

<u>JK Performance by the CONTROL Groups</u>. On List 3, the two CONTROL groups predicted their recall just as accurately as did the two JK groups. Furthermore, the specific patterns of JK outcomes, the bias scores, and the testing effects found for the CONTROL groups were not markedly different from the JK groups. The only difference was that the CONTROL groups exhibited more Ms than the JK groups on the third list indicating that the CONTROL groups had a slightly greater tendency to be

inaccurate with NO judgments. However, overall accuracy as measured by JK errors (M + FA) did not differ between groups on the third list. Two interpretations of the failure to find a difference between the CONTROL and JK groups can be offered.

It can be suggested that the kinds of information that contribute to JKs are inherently involved in the learning of a PA list. That is, perhaps subjects in the STUDY-TEST-CONTROL group were covertly deciding, for example, that more study time should be spent on items that they had missed on preceeding tests. Subjects in the STUDY-ONLY-CONTROL group may have been allocating attention or processing effort on the basis of what can be called implicit JKs. Arbuckle and Cuddy (1969) have suggested this commonality of demands between a JK task and a standard PA learning task. Zacks (1969) also suggested that covert self-regulation of processing efforts may be occurring in PA learning. If this interpretation is to be accepted, fast learners should have been more accurate than slow learners, and this was found not to be the case. Therefore, it cannot be concluded that PA learning inherently involves those processes on which JKs are based.

A second interpretation of the equal JK accuracy shown by the CONTROL and JK groups on List 3 is based on the role of tests. In any learning task (i.e. free recall, recognition as well as PA learning) tests provide "feedback"

to learners (LaPort & Voss, 1974a, 1974b). Since all groups were tested on each list, all subjects must have gained an understanding of what was required for "mastery" of a list given the constraints of the learning procedures. As was concluded above, familiarity with the learning procedures rather than with the JK task was crucial to eventual JK performance. Furthermore, with learning experience with a particular type of learning material (i.e. PAs) subjects could also attain a greater understanding of what is necessary for learning. Since subjects in the CONTROL groups had an equal amount of learning and testing experience as the JK groups, their understanding of task demands or their acquisition of "process knowledge" would have been equivalent to that for subjects in the JK groups. Thus, JK performance for the JK groups and the CONTROL groups might have been equal because both groups acquired the appropriate process knowledge on the first two lists. This explanation is admittedly speculative, and the present experiment did not include control groups relevant to this question. This issue could be directly tested if the experiment were replicated with the addition of two groups. Suppose one group was instructed to learn the first two PA lists but was not tested on the items, and then learned the third list, made JK ratings, and received a test trial. If this group was found to be less accurate than a group given the STUDY-ONLY-JK instructions, one could conclude

that testing was crucial to the development of process knowledge useful for later JK performance. Furthermore, suppose one group of subjects received the STUDY-ONLY-JK instructions but that for the first two lists as different type of PA was to be learned (i.e. letter-digit pairs). If the JK accuracy shown by this group was less than that shown by a group that learned CVC-word PAs on all three lists, then it could be concluded that process knowledge is partially dependent on the type of material that is to be learned. If these results were obtained, it could be concluded that equivalent process knowledge had been acquired by the JK and CONTROL groups in the present experiment.

Stimulus Knowledge. Thus far in this discussion, only two proposed sources of information which contribute to JKs have been discussed: process knowledge and implicit retrieval. In the introduction to this paper stimulus knowledge, or an understanding of those item characteristics which determine learning ease, was suggested as a source of information relevant to a JK. Arbuckle and Cuddy (1969) concluded that the perception of the ease or difficulty of items was important for the prediction of recall or non-recall. Perhaps certain pairs used in the present experiment were more easily associated than others, and thus if a subject were aware of these a priori differences in ease of learning (EL), JKs might have

been influenced by this stimulus knowledge. To determine whether JKs were related to the perceived EL for the given pairs, and to see if JKs contained any information other than stimulus knowledge the following analysis was performed.

The proportion correct recall for each of the 72 pairs was determined by collapsing recall data for the first two lists from the CONTROL groups. Data from these groups were used in order to obtain measures of recall probability that are uninfluenced by the performance of JKs. Further, the mean JK rating assigned to each pair was computed by collapsing JK ratings from the STUDY-TEST-JK and STUDY-ONLY-JK groups for all three lists. Thus, each mean was based on 60 ratings.

To obtain EL ratings, the 72 PAs were shown to 36 additional subjects who were naive to the purposes of the present experiment. These subjects were instructed to pretend that they were shown these pairs several times and that for testing, only the CVC would be shown and the production of the response term would be required. The subjects were asked to rate each pair on how easily it could be learned. A six-point scale ranging from very difficult (1) to very easy (6) was provided. Five seconds were allowed for the rating of each pair.

It was found that the items' mean EL ratings were highly correlated with the items' probability of recall

($r(70) = .63$, $p < .001$) and with the mean JK rating assigned to an item ($r(70) = .73$, $p < .001$). Moreover from the original subjects it was learned that the probability of recall of an item correlated .83 ($df = 70$, $p < .001$) with the mean JK rating. If stimulus knowledge (i.e. mean EL rating) completely accounted for the relationship between recall and mean JKs, then a part correlation between probability of recall and mean JK predicted by EL, should be near zero. The part correlation was significantly greater than zero, ($r(70) = .54$, $p < .001$). The decrease in the correlation indicates that to a certain extent, JKs were related to perceived EL. However, stimulus knowledge does not completely account for an item's JK rating. Under different learning conditions perhaps stimulus knowledge would become a much more important determinant of JKs.

General Conclusions and Implications. The present experiment demonstrates that learners can predict what will be recalled with a reasonable degree of accuracy. This "ability to know what you know" benefits from the knowledge of the success of previous retrieval attempts. Therefore, this implies that students' study habits should include self-testing. Self-testing or practice testing would lead to an accurate assessment of what is known and thus study time could efficiently be allocated to that which is not known. Furthermore, since JK accuracy generally improved with increasing familiarity with specific

learning and testing procedures, the ability to know what is known can be said to depend on an understanding of the manner in which the information will be tested. With greater practice with a certain type of test or with a certain kind of to-be-learned material, accuracy of knowing what is known will increase.

A limitation of the present experiment was that only one type of stimulus material and one type of testing procedure were employed. Subsequent research should determine the degree to which prediction ability or experience acquired during PA learning transfers to other learning situations. For example, does the experience of predicting PA recall facilitate the prediction of sentence retention? Furthermore, in addition to the prediction of recall, can subjects accurately predict which items will be recognized? Future research should answer these questions.

REFERENCE NOTES

1. Zechmeister, E. B., & Shaughnessy, J. J. When you know and when you think that you know but you don't. Paper presented at the 15th Annual Meeting of The Psychonomic Society, Boston, November 1974.
2. Gude, C., & Zechmeister, E. B. Role of test experience and expectation on free recall. Paper presented at the American Psychological Association 83rd Annual Convention, Chicago, August, 1975.
3. Pasko, S. J. Judgments of knowing relative to stimulus knowledge, practice and individual differences. Unpublished dissertation, Loyola University of Chicago.

REFERENCES

Arbuckle, T. Y., & Cuddy, L. L. Discrimination of item strength at time of presentation. Journal of Experimental Psychology, 1969, 81, 126-131.

Blake, M. Prediction of recognition when recall fails: exploring the feeling of knowing phenomenon. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 311-319.

Brown, R., & McNeill, D. The "tip-of-the-tongue" phenomenon. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 325-337.

Cooper, E. H., & Pantle, A. J. The total-time hypothesis in verbal learning. Psychological Bulletin, 1967, 68, 221-234.

Craik, F. I. M. The fate of primary memory items in free recall. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 143-148.

Craik, F. I. M. A "levels of analysis" view of memory. In Pliner, P., Krames, L., & Alloway, T. (eds.) Communication and affect: Language and thought. New York: Academic Press, 1973.

Flavell, J. H. What is memory the development of? Human Development, 1971, 14, 272-278.

Flavell, J. H., & Wellman, H. M. Metamemory. In Kail, R. V., & Hagen, J. W. (eds.) _Perspectives on the development of memory and cognition_. Hillsdale, N.J.: Lawrence Erlbaum Associates, in press.

Flavell, J. H., Fredricks, A. G., & Hoyt, J. D. Developmental changes in the memorization process. _Cognitive Psychology_, 1970, _1_, 324-340.

Hagen, J. W. Some thoughts on how children learn to remember. _Human Development_, 1971, _14_, 262-271.

Hart, J. T. Memory and the feeling-of-knowing experience. _Journal of Educational Psychology_, 1965, _56_, 208-216.

Hart, J. T. A methodological note on the feeling-of-knowing experiments. _Journal of Educational Psychology_, 1966, _57_, 347-349.

Hart, J. T. Memory and the memory-monitoring process. _Journal of Verbal Learning and Verbal Behavior_, 1967, _6_, 685-691.

Hintzman, D. L. Theoretical implications of the spacing effect. In R. L. Solso (ed.) _Theories in cognitive psychology: The Loyola symposium_. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Hyde, T. S., & Jenkins, J. J. The differential effects of incidental tasks on the organization of recall of a list of highly associated words. _Journal of Experimental Psychology_, 1969, _82_, 472-481.

Jenkins, J. J. Can we have a meaningful theory of memory? In R. L. Solso (ed.) Theories in cognitive psychology: The Loyola symposium. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Kellas, G., McCauley, C., & McFarland, C. E. Reexamination of externalized rehearsal. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1, 84-90.

Koriat, A., & Lieblich, I. What does a person in a "TOT" state know that a person in a "Don't know" state doesn't know? Memory and Cognition, 1974, 2, 647-655.

Kreutzer, M. A., Leonard, C., & Flavell, J. H. An interview study of children's knowledge about memory. Monographs of the Society for Research in Child Development, 1975, 40, (1, Serial No. 159).

LaPorte, R., & Voss, J. F. Paired-associate acquisition as a function of a number of initial non-test trials. Journal of Experimental Psychology, 1974a, 103, 117-123.

LaPorte, R., & Voss, J. F. Nonspecific transfer as a function of number of lists acquired and criterion of learning. American Journal of Psychology, 1974b, 87, 463-474.

LaPorte, R., Voss, J. F., & Bisanz, G. Nonspecific transfer with and without test trials on the first list. American Journal of Psychology, 1974, 87, 475-479.

Light, L. L. Incentives, information, rehearsal and the negative recency effect. Memory and Cognition, 1974, 2, 295-300.

Lippman, L. G., & Kintz, B. L. Group predictions of item differences of CVC trigrams. Psychonomic Science, 1968, 12, 265-266.

Lockhart, R. S. The facilitation of recognition by recall. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 253-258.

McCabe, L., & Madigan, S. Negative effects of recency in recall and recognition. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 307-310.

Masur, E. F., McIntyre, C. W., & Flavell, J. H. Developmental changes in apportionment of study time among items in a multitrial free recall task. Journal of Experimental Child Psychology, 1973, 15, 237-246.

Moynahan, E. D. The development of knowledge concerning the effects of categorization upon free recall learning. Child Development, 1973, 44, 238-246.

Murdock, B. B. The criterion problem in short-term memory. Journal of Experimental Psychology, 1966, 72, 317-324.

Noble, C. E. Measurements of association value (a), rated associations (a'), and scaled meaningfulness (m') for the 2100 CVC combinations of the English alphabet. Psychological Reports, 1961, 8, 487-521.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, Imagery, and Meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph Supplement, 1968, (Vol. 76, No. 1, Part 2).

Postman, L. Studies of learning-to-learn II: Changes in transfer as a function of practice. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 137-147.

Postman, L. Experimental analysis of learning-to-learn. In Bower, G. H., & Spence, J. T. (eds.) The psychology of learning and motivation, Vol. 3. New York: Academic Press, 1969.

Postman, L., Keppel, G., & Zacks, R. Studies of learning-to-learn VII: The effects of practice on response integration. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 776-784.

Richardson, J., & Erlebacher, A. Associative connection between paired verbal items. Journal of Experimental Psychology, 1958, 56, 62-69.

Shaughnessy, J. J., Zimmerman, J., & Underwood, B. J. Further evidence on the MP-DP effect in free recall learning. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 2-12.

Tulving, E., & Madigan, S. Memory and verbal learning. Annual Review of Psychology, 21, 437-484.

Underwood B. J. Individual and group predictions of item difficulty for free learning. Journal of Experimental Psychology, 1966, 71, 673-679.

Underwood, B. J. The role of the association in recognition memory. Journal of Experimental Psychology Monograph, 1974, (Vol. 102, No. 5, 917-939).

Underwood, B. J., & Freund, J. S. Word frequency and short-term recognition memory. American Journal of Psychology, 1970, 83, 343-351.

Underwood, B. J., Ham, M., & Ekstrand, B. R. Cue selection in paired-associate learning. Journal of Experimental Psychology, 1962, 64, 405-409.

Watkins, M. J., & Watkins, O. Processing of recency items for free recall. Journal of Experimental Psychology, 1974, 102, 488-493.

Woodsen, C. E. Learner judgment in instructional decisions for learning meaningful paired-associates. Journal of Experimental Psychology, 1974, 102, 167-169.

Yarmey, A. D. I recognize your face but I can't remember your name: further evidence of the tip-of-tongue phenomenon. Memory and Cognition, 1973, 1, 287-290.

Yussen, S. R., & Levy, V. M. Developmental changes in predicting one's own span of short-term memory. Journal of Experimental Child Psychology, 1975, 19, 502-508.

Zacks, R. T. Invariance of total learning time under different conditions of practice. Journal of Experimental Psychology, 1969, 82, 441-447.

APPENDIX A

Paired-Associate Lists

1.	nos--gadfly	het--foible	kol--chaos
2.	lom--offshoot	wof--preview	nep--belief
3.	tol--deceit	buk--adage	gos--welfare
4.	kew--nephew	tem--maker	fen--traction
5.	yep--unit	gel--nonsense	yeg--savant
6.	laq--session	vug--research	fim--namesake
7.	lem--charter	hup--vapor	sek--vision
8.	pum--venom	dof--proxy	gid--context
9.	mac--tidbit	mot--friction	jow--forethought
10.	wer--buffoon	ril--frontage	nas--steerage
11.	bik--madness	xic--hatred	lor--essence
12.	vam--folly	raq--vigor	hus--blessing
13.	dut--upkeep	tif--patent	jun--franchise
14.	nup--abbess	neb--array	cuz--feline
15.	fub--encore	kav--outcome	fok--hardship
16.	lur--item	ren--prestige	tog--satire
17.	dal--garret	fis--rating	dil--assault
18.	tux--concept	wes--workhouse	rew--onslaught
19.	nes--impact	rox--hearing	cid--hindrance
20.	pom--chloride	jor--tribute	gur--gender
21.	siz--pressure	fer--victim	pit--mischief
22.	dow--impulse	bod--kindness	bov--northwest
23.	hib--reflex	fow--crisis	sik--malice
24.	mik--boredom	las--rosin	lox--surtax

APPROVAL SHEET

The thesis submitted by Joseph F. King has been read and approved by the following committee:

Dr. E. B. Zechmeister, Director
Associate Professor, Psychology, Loyola

Dr. E. J. Posavac
Associate Professor, Psychology, Loyola

Dr. F. L. Slaymaker
Assistant Professor, Psychology, Loyola

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the thesis is now given final approval by the Committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment of the requirements for the degree of Master of Arts.

5-12-76
Date

Director's Signature