



1976

The Judgment of Knowing as Related to Practice, Ease of Learning Judgments, and Individual Differences

Stanley J. Pasko
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

 Part of the [Psychology Commons](#)

Recommended Citation

Pasko, Stanley J., "The Judgment of Knowing as Related to Practice, Ease of Learning Judgments, and Individual Differences" (1976). *Dissertations*. 1695.
https://ecommons.luc.edu/luc_diss/1695

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](#).
Copyright © 1976 Stanley J. Pasko

THE JUDGMENT OF KNOWING
AS RELATED TO
PRACTICE, EASE OF LEARNING JUDGMENTS, AND INDIVIDUAL DIFFERENCES

by
Stanley J. Pasko

A Dissertation Submitted to the Faculty of the Graduate School
of Loyola University in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy
November
1976

ACKNOWLEDGEMENTS

General gratitude is present for the gifts of health, good-fortune, and ability which have allowed the completion of this dissertation.

Specific appreciation must be offered to my parents, brother, and grandparents who have consistently encouraged my fascination with verbal materials; to the Felician Sisters of St. Wenceslaus Grammar School who must have felt blest with class size under fifty; to the priests at Quigley and St. Mary of the Lake Seminary who taught a respect for knowledge; to the faculty of Loyola University who have created an atmosphere conducive to learning, and to the Department of Psychology which has supplied financial and emotional support over the years.

It is also a privilege to thank Dr. Eugene B. Zechmeister for his direction of this dissertation. His confident guidance in a generally unexplored area has increased the likelihood that this research may offer insight to other experimenters. Further gratefulness is present to Drs. Emil J. Posavac and John R. Crocker, S.J., for their serving on the committee of a dissertation which mirrors their own concern that students should profit from school and study. Dr. John J. Shaughnessy and Joseph F. King ought also to be thanked for discussing and reading this paper.

Lastly, I am fortunate in being able to rejoice that Julieta Rosales Flores de Pasko, M.A., has helped make her prediction of task success a self-fulfilling prophecy.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
TABLE OF CONTENTS	iii
LIST OF TABLES.	vii
LIST OF FIGURES	viii
CONTENTS OF APPENDICES.	ix
 Chapter	
I. INTRODUCTION.	1
Judgment of Knowing.	4
Comparison with ease of learning and feeling of knowing judgments.	9
The EL judgment and the Judgment of Knowing	9
The FK judgment and the Judgment of Knowing	10
Research relevant to memorial self-knowledge	13
Educational psychology and memorial self-knowledge. .	14
Developmental psychology and memorial self- knowledge	16
Physiology and memorial self-knowledge.	18
Learning-memory and memorial self-knowledge	22
The Present Research	29
Question 1. Accuracy, practice, and the Judgment of Knowing	31
Question 2. The EL judgment and the Judgment of Knowing	32

Question 3. Feeling of knowing judgments, scholastic achievement, and the Judgment of Knowing .	34
Question 4. Learning ability and the Judgment of Knowing	35
II. METHOD	36
Design	36
Lists	36
Procedure	39
Subjects	45
III. RESULTS	47
Question 1. Accuracy, practice, and the Judgment of Knowing	47
The JK as a dichotomous measure	51
The JK as a six-point scale.	57
Question 2. The EL judgment and the Judgment of Knowing	66
Correlations between EL and JK ratings by individuals and by groups	66
Predictions of recall performance by EL and JK ratings	71
Secondary Analysis: EL and JK congruence and JK accuracy	77
Secondary Analysis: Stimulus characteristics as predictive of the JK	80
Imagery as predictive of the JK	82

	Page
Learnability as predictive of the JK	85
Question 3. Feeling of knowing judgments, scholastic achievement, and the Judgment of Knowing	88
The feeling of knowing judgment	88
The feeling of knowing judgment and the Judgment of Knowing	92
Secondary Analysis: Listwise comparisons between JK and FK accuracy	95
Secondary Analysis: FK accuracy in comparison with EL and JK accuracy	97
Overall scholastic achievement and the Judgment of Knowing	98
College course achievement and the Judgment of Knowing	99
Summary	102
Question 4. Learning ability and the Judgment of Knowing	103
Learning ability and practice at the Judgment of Knowing	104
Summary	106
The relationship between EL and JK ratings and learning ability	107
Secondary Analysis: Predictive congruence and learning ability	107
Individual differences and learning ability	108

	Page
Time at the EL task and the Judgment of Knowing.	109
IV. DISCUSSION	111
Three general considerations regarding research on the	
JK	112
Accuracy, practice, and the Judgment of Knowing	115
The ease of learning judgment and the Judgment of	
Knowing	116
Feeling of knowing judgments, scholastic achievement,	
and the Judgment of Knowing	120
Learning ability and the Judgment of Knowing.	122
Summary of the present research	125
REFERENCES	127
APPENDIX A	135
APPENDIX B	140

LIST OF TABLES

Table	Page
1. Outline of sessions, stages, and tasks	30
2. Information given subjects regarding the game-point manipulation	42
3. Terminology from signal detection theory as applied to JK accuracy	48
4. Comparisons of relative (JK Accuracy) and absolute (JK Errors) measures of JK sensitivity	50
5. Mean performance per subject relevant to the evaluation of the sensitivity and bias of the dichotomous JK	54
6. Recall proportions for the JK scale values on four lists . .	59
7. JK performance scored as negative game-points, positive game-points, and their components: Mean game-points per subject	62
8. Measures of central tendency for EL and JK correlations . .	67
9. Measures of central tendency for biserial correlations between EL and JK ratings and recall performance	73
10. Predictive congruence-incongruence and probabilities of correct JK predictions	79
11. Proportions of correct recognitions for the six FK scale positions	89
12. Pearson product-moment correlations for overall JK and FK measures	94

LIST OF FIGURES

Figure		Page
1	Recall proportions for JK scale values on four lists	58
2	JK Errors over lists as a function of learning ability . . .	105

CONTENTS OF APPENDICES

	Page
APPENDIX A Stimulus and response terms for the PA lists	135
APPENDIX B Feeling of knowing questions	140

CHAPTER I

INTRODUCTION

Believing that learning and memory are best characterized as under stimulus control, traditional experimentation on memory processes has assumed that the systematic manipulation of materials and tasks would lead to a theory of memory adequate both within and outside the laboratory (Jenkins, 1974a, 1974b).

However, research has gradually delineated a role in memory for processes which are not directly under the control of the external stimulus. For example, stimulus selection has been found to occur during paired-associate (PA) learning (Underwood, Ham, & Ekstrand, 1962) and subjective organization during free recall (Tulving, 1962). Likewise, the negative results of experimentation on the interference theory of forgetting in the context of extraexperimental sources of interference have generated much research on the hypothesized operation of subjective performance variables such as recoding, differentiation, and the "selector mechanism" (Postman, 1963). Such data and theory thus support the hypothesis that learning and memory are determined by the interaction of stimulus- and subject-controlled processes.

A subject-controlled process which has been given minimal attention in psychological research has been the human memory system's self-knowledge (Tulving & Madigan, 1970). A particular kind of memorial self-knowledge which must be acquired for general competence outside the laboratory is the capacity to determine when study has been sufficient

to allow later retention. The basis of such self-knowledge would seem to be an ability to estimate the memorial status of presently studied material.

An experiment by Bugelski and Rickwood (1963) indicates the presence of such a skill in the laboratory. Subjects in this study were allowed to self-pace their study of nonsense syllable paired-associates to a criterion of two successive correct repetitions of each response term. The mean total response exposure time per item for these subjects did not significantly differ from that required by subjects who learned the items to the same criterion under various experimenter-paced presentation rates (Bugelski, 1962). It would seem that the subjects were rather adept at judging when learning was sufficient to allow the discontinuance of their study of individual word-pairs.

A study of Kellas, McCauley, and McFarland (1975) may be interpreted similarly. In this experiment, subjects self-paced their study of lists of words. Each list was presented only once and a free recall test trial occurred immediately afterwards. One group was allowed to study covertly; another group was required to rehearse overtly. Both groups recalled a similar number of words. However, the overt rehearsal group utilized far more study time than the covert rehearsal group. If one accepts Landauer's (1962) conclusion that subvocal speech is not essentially faster than overt speech, the temporal difference between the two groups would not seem to have resulted from verbalization itself. Instead, it seems plausible that subjects using overt rehearsal were at least implicitly aware of the limitations placed on their future retention by the requirement of overt rehearsal and therefore self-paced their study at a much slower rate than that chosen by the covert rehearsal

subjects.

However, in other laboratory studies, estimation of one's own knowledge state has not always been so accurate. Shaughnessy, Zimmerman, and Underwood (1972) allowed subjects to self-pace their study of a list of words in which some items were repeated contiguously (massed presentation or MP items) while other items were repeated only after different words intervened (distributed presentation or DP items). Subjects chose to study MP items for less time than DP items and subsequently recalled less MP than DP items.

To examine why subjects allot less study time to MP than to DP items, Zechmeister and Shaughnessy (1974) asked subjects to predict their likelihood of recalling individual words on an immediately forthcoming free recall trial. It was found that, while recall was greater for DP than for MP items, the predictions of future recall were higher for the MP items. It would seem that the method of presentation misled subjects into overevaluating the present memorial status of the MP items.

The gist of the above studies would seem to be that subjects are able to some degree to estimate the present memorial status of to be learned material but that such judgments are imperfect. The further evaluation of the nature of this skill is the theoretical concern of the present research.

Research on the subjective awareness of one's degree of knowledge regarding the memorial status of information while it is being learned may also have practical implications. That is, teachers are often at a loss to evaluate a student's statement that he or she had studied substantially for a test but still did poorly. When item difficulty and personal veracity are not at issue, one wonders why a correlation between

study time and test scores did not materialize. Perhaps such low performance is the result of a poorly developed capability at evaluating one's present state of knowledge. Should one falsely evaluate a present state of knowledge regarding, for example, the formula for variance, and on this basis discontinue its further memorization, even a good grasp of the concept will not later be of much assistance in either formula recall or computation. If the factors influencing such personal evaluations can be clarified, some service may be done students of all ages.

Judgment of Knowing

The present research is concerned with what is herein called a judgment of knowing (JK). A JK is defined as the estimation of one's later memory for presently studied information. It is suggested that, to be accurate, this prediction of future performance must involve at least two components. The first necessity is an evaluation of the present memorial status of the information in question. The second requirement is that the evaluation be made in light of the conditions under which retention will be tested, i.e., retention interval, number of other items to be retained, type of retention test.

Several sources of knowledge for a JK can be identified. These sources may involve stimulus knowledge, process knowledge, and/or memory knowledge. Stimulus knowledge refers to subjects' awareness of the relative ease of learning of individual information. Thus, past experience with words might allow one to predict differential recallability on the basis of such stimulus characteristics as pronunciability or association value. Since ratings of both item pronunciability and association value correlate quite highly with recall (Underwood, 1966),

the accuracy of a JK founded on this basis may be rather high. It may further be theorized that, if the basis of a JK is stimulus knowledge, a more direct memorial examination is unnecessary.

A second source of information for a JK can be labeled process knowledge. Process knowledge refers to a familiarity with the overall adequacy of either one's own learning operations or of learning processes in general. In an experimental situation, for example, a person who twice fails to recall any items in an entire list may well be influenced by this process knowledge to make predictions of future nonrecall if he is asked to predict his performance on a third list. As with stimulus knowledge, process knowledge need not require a memorial examination in order to be accurate.

It is plausible that a combination of stimulus and process knowledge is sufficient to predict future retention. That is, if a PA item appears hard to learn and if the pair has been rehearsed only one time, a prediction of future nonrecall would seem to have a high likelihood of later being proven correct.

However, evidence is present that the human memory system has some self-knowledge of its own contents (v.g., Hart, 1965; Brown & McNeill, 1966) and that subjects rather accurately indicate whether an answer in PA recall is correct (v.g., Suboski, Pappas, & Murray, 1966a). Therefore, the possibility exists that such knowledge of one's own memory is also present more immediately after a learning attempt.

Memory knowledge may be taken to refer to the knowledge that a memory has been formed, that an attribute now exists in memory which was not previously present. Such knowledge is seen as different from stimulus knowledge because it results from an attempt to learn information

and because it is not based on already known information. Similarly, memory knowledge is seen as other than process knowledge because process knowledge merely states that certain operations have been performed on information.

The definition of the three types of knowledge and their respective roles in the JK may be better clarified by an examination of three learning situations.

In one learning situation (v.g., the PA learning of tub-venom), a subject may be immediately aware of a strong associate to one term (v.g., snake to venom). Such knowledge is already present in the subject's memory and hence is defined as stimulus knowledge. It may lead to a judgment that the PA is learnable and thus may serve as the basis for a JK. If a mental image uniting the to be learned terms (v.g., a bathtub filled with snakes spewing liquid) also becomes present, such a visual associate is seen as a new attribute present in one's memory and is labeled memory knowledge. Further, awareness of the efficacy of a vivid image as a retrieval cue may be labeled process knowledge. Although all three types of knowledge are present in such an example, the sine qua non character of memory knowledge in such a situation would seem to validate its role as the basis of a positive JK.

It may also be noted that there is at least the logical necessity of a less explicit type of memory knowledge in all similar situations in which past knowledge is utilized in new learning. Thus, in the above example, a subject must have some means of distinguishing between cues (v.g., snakes, spewed liquid, poison) and the to be learned associate (v.g., venom). Neither stimulus knowledge nor process knowledge can be effective in this regard. Some attribute must therefore be present in

memory to allow the appropriate discrimination. This may be a memory for the relatively greater rehearsal frequency of the to be retained term, of its temporal priority, or perhaps even the memory that the response term is an associate of the mental picture. Explicit awareness of such memory knowledge is not as important as its logically necessary presence in many situations in which future recall is predicted on the basis of information already in memory.

A second learning situation is one in which a subject discovers an associative mechanism whose adequacy as a retrieval aid is uncertain. For the PA 'bat-hindrance', for example, a subject may picture a baseball bat held horizontally at arm's length. This learning experience may be classified as similar to that previously mentioned. However, the possibility exists that the association between bat and hindrance was learned in such a way that the imaginal cue was secondary. The basis for the prediction may be the learned material itself. Thus, the example may also be considered analogous to the third learning situation. A less ambiguous designation is not presently possible.

A third learning situation is one in which the utility of associative or mediational cues is limited, v.g., ~~names~~ of colleagues, foreign vocabulary, or PA learning when no mediators arise. In such circumstances, rehearsal is of paramount importance during original learning. However, the decision must eventually be made that an item is "known" so that rehearsal may be discontinued. Although self-testing may be utilized for such decisions, the possibility exists that the determination to end rehearsal and to begin self-testing is itself the result of some knowledge about an item's memorial status.

Since stimulus knowledge and process knowledge offer little

assistance regarding decisions as to the degree of response-learning, the knowledge involved would seem to be memory knowledge. In such instances, however, the new attributes added to memory would have to be the phonetic, visual, or orthographic attributes of the to be learned material rather than associative attributes related to the material. Further, for subjects to have confidence in such judgments, some cues must be available which indicate that learning has occurred.

Potential candidates as cues are the physiological changes which have been found to accompany stages of learning. That is, variations in skin resistance levels have been found for different degrees of learning (Kintsch, 1965). Similarly, differential eye-movements have been discovered during different stages of learning (McCormack & Hal-trecht, 1966). If subjects have a minimal degree of awareness of such correlates of learned states (and hence of memory knowledge), the changes in arousal level may allow somewhat accurate JK ratings.

A more hypothetical basis for a memorially based JK is the sensory feedback from a response which closed-loop theory postulates as being laid down during learning (Adams, 1968). The existence of such a perceptual trace is theorized to serve as a referent for determining the correctness of future occurrences of the response. If proprioceptive feedback has a role during recall, a "sense" of the extent of its formation during learning would seem to be a potential substratum for a JK.

Although it has seemed necessary to establish differences between various potential bases for the JK, it must be emphasized that the present research design allows only very general statements regarding sources for the JK.

Comparison with ease of learning and feeling of knowing judgments

A JK is seen as related to two other types of judgments which have received some attention in research on learning and memory—ease of learning (EL) ratings and the feeling of knowing (FK) which was introduced by Hart (1965).

The EL judgment and the Judgment of Knowing. To introduce the concept of an EL judgment, the procedure for one group of subjects in an experiment by Underwood (1966) may be described. Subjects were first shown a list of 27 three-letter combinations which covered a wide range of learning difficulty, from three-letter words to difficult consonant syllables. They were told to imagine seeing the trigrams presented singly at a 2-sec. rate for a series of study trials and having to recall them on alternating test trials. After being asked to examine all the trigrams "to get a 'feel' for the range of difficulty," the subjects were asked to draw lines next to each trigram to indicate the speed with which they thought they would learn it. The lines were to be drawn proportionate to item difficulty so that, for example, a trigram judged twice as difficult to learn as another would receive a line twice as long. Afterwards, the trigrams were presented singly for six study-test trials. The average length of the EL lines for individual trigrams was then correlated with the average number correct for that item to serve as a measure of EL predictive accuracy. This correlation was very high (.91) and was similar to that between this group's mean EL scaling for the items and mean learning scores for the items by a group of subjects which did not make EL judgments.

In being asked to make an EL judgment, therefore, a subject is asked to draw on past experience with similar stimuli (i.e., on what

has been previously labeled "stimulus knowledge") to make a probabilistic decision as to the amount of time or effort required to establish a memory.

In contrast to the EL rating, the JK is an estimate of the present status of knowledge rather than of the relative potential learnability of particular items. Hence, it can only be made during an attempt to learn information. Further, a JK is only valid until some definite future time.

An example can more clearly distinguish EL and JK ratings. If two subjects with identical past histories were to have a study trial on the same material and one were then asked for an EL judgment and the other for a JK, both on similarly ascending six-point scales, the EL ratings for a certain item may be a 4 to indicate a moderately easy to learn item, while the JK may be a 2 to indicate the improbability of recall on the next trial--and both judgments could be accurate. However, if the same judgments were again asked after five more study trials, the ideal EL rater would be expected to make the same judgment, while the ideal JK subject (under the assumption that learning had occurred for the particular item) would be expected to give a much higher JK than the earlier evaluation.

The FK judgment and the Judgment of Knowing. Though the feeling of knowing paradigm introduced by Hart (1965) has undergone methodological variations (cf. Hart, 1967a; Blake, 1973), an examination of the original procedure permits a clear differentiation between the feeling of knowing and the JK. In the Hart (1965) experiment, subjects first were asked general information questions, such as the name of the painter of "Afternoon at La Grand Jatte." If a subject could not supply the

answer for a question, he was asked to indicate whether or not he felt he knew the answer to the extent that he could recognize it among several alternatives. A YES response was thus considered an indication of a subject's being in a FK state. After all the questions had been presented, the subject was asked to select the correct answer from four similar alternatives for each question. As an example, the alternatives for the above question were Monet, Cezanne, Seurat, and Dufy.

To assess FK accuracy, Hart subtracted the proportion of correct recognitions given a NO FK judgment from the proportion of correct recognitions given a YES FK judgment. It was theorized that, if feelings of knowing were accurate indicators of memory storage, subjects would recognize more test items about which they had a feeling of knowing than items they had predicted would be unrecognizable. The difference supported this prediction. However, this and later studies (Hart, 1967a, 1967b; Blake, 1973) have also shown that subjects were able to select the correct alternative at better than chance levels when they had previously indicated no feeling of knowing. This inability to predict non-recognition was interpreted by Hart (1965) as due to the lack of equally likely alternatives so that "a person ignorant of the correct answer but possessing some related knowledge of the field covered by the question might be able to narrow the alternatives down to three or two, thereby raising his guessing probability to .33 or .50" (p. 211).

To the extent that a FK judgment is a subjective indication that a memory has or has not been established (Underwood, 1969), both a FK judgment and a JK are similar. However, there are important differences.

An appropriate reference point for determining dissimilarities between a JK and a FK judgment may be the presence or absence of the

relevant stimulus. With the to be learned stimulus present, a subject is able to use either stimulus knowledge, process knowledge, or memory knowledge to arrive at a JK. With the to be learned stimulus absent, however, stimulus knowledge is excluded by definition as a factor in FK judgments. Nevertheless, like a JK, a FK judgment may be based on process knowledge. That is, a subject in a laboratory learning task who has been presented a trigram and is later asked to give a FK judgment when it can not be recalled in its entirety (i.e., Blake, 1973) may remember that he had rehearsed the item more than other items and, on this basis, predict future recognition. Since the greater accrued frequency for the correct alternative may lead to successful recognition (Underwood, 1971), this basis for a FK judgment may be quite accurate and not involve item specific memory knowledge.

However, a prediction of future recognition based on process knowledge, i.e., a FK judgment, would appear to require less of the subject than a prediction of future recall based on process knowledge, i.e., a JK. That is, inasmuch as recall is usually considered to require search and retrieval (cf. McCormack, 1972), a JK based on process knowledge would likely require an evaluation of the retrieval potentialities of the specific processing performed. Contrariwise, to the extent that recognition does not require search and retrieval (cf. McCormack, 1972), a FK judgment based on process knowledge need not demand as elaborate an examination.

A further distinction between a JK and a FK judgment is seen when both are considered to be based on memory knowledge. For a JK, the memory knowledge utilized is theorized to be that of the present memorial status or degree of learning of the material. For a FK judgment,

however, the memory knowledge utilized would seem to be either or both of two kinds. A first basis may be target related information which is inherent in the proffered question itself, that is, information concerning the population of potential targets (Koriat & Lieblich, 1974). With respect to the previously sought painter's name, for instance, knowledge of the painting's pointillistic style may lead one to expect that he or she will be able to select the correct alternative from among painters of different styles. In other instances, however, knowledge of the potential alternatives may cause one to give a negative FK. A second memory basis for an FK judgment may be attribute knowledge of the target (Blake, 1973). In the present example, such a basis might be the knowledge or "feeling" that the name has two syllables and begins with either the letter "S" or the letter "T". In contrast to the JK, then, the memory knowledge used for an FK judgment is sometimes merely on the tip of one's tongue.

Research relevant to memorial self-knowledge

Although direct experimentation on what has been labeled a JK is rare, a goodly number of studies may be related to memorial self-knowledge. That is, eventually the question which must be asked regarding the JK is whether it is based on stimulus, process, or memory knowledge. Research relevant to memorial self-knowledge is of much importance in this regard. However, the studies to be mentioned were rarely designed to distinguish the three types of knowledge. Thus, no further attempt will be made to specify memorial self-knowledge in these articles. Because of the relative disparateness of the individual experiments, they will be considered under the general areas of psychology rather than

under topic headings. The areas will be educational psychology, developmental psychology, physiology, and learning-memory.

Educational psychology and memorial self-knowledge. A persistent concern in educational psychology has been the value of changing answers in multiple-choice tests. Although the conventional wisdom emphasizes trust in first impressions, research has consistently found changes to be beneficial (v.g., Lowe & Crawford, 1929; Reile & Briggs, 1952; Foote & Belinky, 1972). For example, Foote and Belinky (1972) found that 55% of the changes on four introductory psychology tests were from incorrect to correct, while only 22% were in the opposite direction. Also, Reile and Briggs (1952) found that A and B students made less changes than D and F students but profited more from the conversions. Because such situations must involve considerable doubt, it may be plausibly argued that many of these correct changes are based more on "feelings" than on rational elimination of alternatives. However, the methodological difficulties of classroom research demand many limits on such hypothesizing.

A second relevant concern of educational research has been the effect of immediate knowledge of results (KR) on test performance. For example, fifth and sixth grade students who were given feedback in a multiple-choice situation were found to profit only slightly on a later recall test regarding similar information relative to students of matched ability who had not been given immediate feedback (Hanna, 1976). Immediate KR has also been found to have a slight positive effect on the final examination scores of college students but to have no significant effect on the performance of either high school or college students on a total of ten shorter within-course tests (Beeson, 1973). Likewise,

graduate students were found to perform more poorly in a KR testing situation than in a delayed feedback setting (Strang & Rust, 1973). In all such experiments, some roles would seem to be present for both the monitoring and the cueing functions of knowledge of results. More importantly, however, the general smallness of the effect would seem to substantiate the hypothesis of Ross (1933) that students' intrinsic judgments of accuracy tend to be very high. Further validation of this comes from Hanna's finding that the slight improvement due to total feedback over no feedback was present only among medium and low ability grammar school students and not among high ability students.

More direct evidence of the accuracy of subjective knowledge regarding the correctness of one's own responses in a classroom situation comes from an experiment by Blum (1935). Subjects were asked to predict their number of correct multiple-choice selections out of 100 questions. Students who scored highly tended to make very accurate estimates (a predicted mean of 82.8 correct answers versus an actual mean of 83.3 correct answers). However, students who scored low were rather inaccurate in their self-estimates (a predicted mean of 74.4 correct answers versus an actual mean of 55.8 correct answers).

A final relevant concern in the educational psychology literature is the accuracy of subjective estimates of overall academic performance, particularly in comparison to standardized tests. Binder, Jones, and Strowig (1970) correlated high school students' grade point averages (GPA) with scores on a self-concept of ability scale. Correlations from two different samples of boys were .56 and .51; correlations for two groups of girls were .71 and .67. Biggs and Tinsley (1970) used essentially the same inventory to assess college students' estimates of their

academic ability. The correlation with GPA was .54. When general academic achievement as measured by American College Test scores was held constant, the correlation remained of moderate strength at .41. Also, as has been found with students who accurately predict their performance on individual course tests, more accurate predictors of overall college grades were higher in GPA scores and, expectedly, tended to predict higher GPA scores than less precise self-evaluators (Keefer, 1971).

Taken as a whole, the results from in-class experimentation present evidence of students' rather sophisticated awareness of both their memorial contents and capabilities.

Developmental psychology and memorial self-knowledge. Developmental changes in the knowledge individuals have regarding their memorial capability have been investigated indirectly by experiments which observe children's activities during memory tasks. For example, Rogoff, Newcombe, and Kagan (1974) asked subjects to self-pace their study of 40 stimulus pictures in anticipation of a recognition task which was to occur in either minutes, one day, or one week. Eight year old children took more time to study items for the two longer retention intervals while six and four year olds studied the items for the same time regardless of the retention interval.

In a similar vein, Flavell, Beach, and Chinsky (1966) observed the spontaneous verbalizations of kindergarteners, second-graders, and fifth-graders during study and test periods of a nonverbal (pointing) serial recall task. An increase in detectable verbalization was found from kindergarten to fifth grade. One interpretation of this continuing increase in rehearsal and verbal coding is that it indicates an increasing awareness of the efficacy of such plans for memorial tasks (Flavell

et al., 1966).

Similar results were obtained by Appel, Cooper, Knight, McCarrell, Yussen, and Flavell (1972) when they asked preschoolers, first-graders, and fifth-graders either to look at or to remember a series of pictures. Preschoolers showed the same kinds of study behaviors and recalled the same number of items under the two instructional sets. Although first-graders also did not study differentially under the two conditions, recall performance in one experiment was found to be superior under the memorization instructions. Fifth-graders, however, studied differently for the two tasks--and recalled more under the memorization instructions.

A more direct approach to the study of the development of subjective knowledge about memorial capabilities compared children's predicted and actual memory spans. Flavell, Friedrichs, and Hoyt (1970) had children from nursery school to fourth grade predict their memory span by indicating whether or not they could serially recall all the items presented to them on strips of paper. From 1 to 10 items were on each strip. The papers were shown in ascending order of set size. The actual span was tested by serial repetitions of the experimenter's oral presentations of series of items.

Errors indicated that children could be separated into those who predicted a memory span of ten items and those who predicted less than ten items. Children who predicted a ten item span almost never recalled so many items and were seen as merely assenting to the highest span offered by the experimenter. Children who predicted memory spans of less than ten items, however, were found to be rather accurate. For example, nursery school children were in error by an average of only 1.7 items, while fourth-graders were off by only 0.9 items.

Another direct evaluation of children's knowledge of their memories was an experiment by Berch and Evans (1973). Kindergarteners and third-graders were compared in the accuracies of their confidence regarding items' being OLD in a continuous recognition task. A four-point scale of confidence contained photographs of "doubtful" and "smiling" children. The two photographs represented two degrees of confidence for both OLD and NEW judgments. It was found that the lower the child's level of confidence in judging an item as OLD, the lower was the probability that the item was actually a repetition. Even for kindergarteners, the probability of an item's being OLD was found to decrease with the four confidence levels (.80, .68, .30, .23). To some extent, then, even 5½ year olds can monitor their memory states.

Physiology and memorial self-knowledge. The present section will review research which has related the learning process to changes in eye movements, pupillary dilation, the galvanic skin response, electroencephalographic recordings, and heart rate. Although these processes generally occur without conscious direction, perceptible differences among levels of these correlates of states of learning may be sufficient cues for accurate judgments of knowing.

The results of an experiment by McCormack and Haltrecht (1966) are typical of research which has monitored eye-movements. Subjects learning pairs of nonsense syllables in the study-test procedure were found originally to spend more time viewing the response than the stimulus term. Within three trials, however, a cross-over occurred and subjects began devoting more attention to the stimulus term. When Haltrecht and McCormack (1966) observed the eye movements of fast and slow learners, they found that slow learners exhibited the cross-over effect. Fast

learners, however, were found to view the stimulus more than the response term from the outset and the difference increased with practice. Over trials, subjects also tended to look away from learned pairs. This was especially true for fast learners.

The experimenters also examined the three most easy and three most difficult CVC pairs for each slow learner. The cross-over point for the fixations on the difficult pairs was observed to occur on a later trial than that for the easy pairs. If subjects are somehow aware of such differential responses to items or if such response-changes are related to the more general indices of arousal which will be discussed below, a somewhat reliable cue for the JK is present.

A second physiological variable which has been observed in relation to learning and memory has been pupillary dilation. The research of Kahneman and his associates has dealt with more general concerns than stages of learning. However, a consistent finding has been that peak pupillary dilation is a function of momentary memory load, with a larger number of items to be recalled correlating with greater dilation (Kahneman & Beatty, 1966). More relevant to the present study is a finding by Colman and Paivio (1970). Pupillary dilation was shown to be greater for abstract than for concrete stimulus terms in a PA learning task. The reaction would seem, therefore, to be a rather sophisticated index of task difficulty.

The galvanic skin response (GSR) has also been studied in relation to the learning of particular items. Brown (1937) had subjects learn a list of 19 nonsense syllables of 0% association value to a criterion of one complete anticipation of the entire list. The galvanic deflections were greater on trials when the subject's response was incorrect

than on those trials on which a correct anticipation was made. Greater deflections, then, were associated with the learning of a syllable than with its correct anticipation.

A design which is less affected by the learning state of previous items than would seem to have been true for Brown's subjects is one utilized by Kintsch (1965) to measure the temporary changes in subjects' skin resistance levels during the test periods of PA learning. Twelve nonsense syllables paired with the numbers 1 or 2 were the to be learned material. On test trials before the last error on a particular item, subjects tended to react with rather large GSRs to the presentations of the stimulus term, with no difference in GSR magnitude for errors and successes. On trials after the last error for an individual item, however, habituation set in and the reactions for that item declined in magnitude, becoming much lower than the initial values. It may be suggested that such changes in the GSR resulted from some awareness of the appropriateness of subjects' own state of knowledge of the individual paired-associates.

The pattern with electroencephalographic (EEG) correlates of verbal learning appears similar to that of the above indices. Thompson and Obrist (1964) had subjects learn a list of twelve nonsense syllables by the serial anticipation method. Maximal changes in voltage reduction occurred when syllables were first anticipated correctly. Since such electrocortical changes tended to diminish with practice, they were considered to reflect variations in attention.

Although heart rate has not apparently been studied on an item specific basis, research by Kahneman, Tursky, Shapiro, and Crider (1969) indicates that heart rate peaks as a function of task difficulty in a

digit transformation task.

In summary, then, relationships have been found between physiological indices and degrees of learning. Such correlations have been found even at the level of individual items. It may be theorized that such changes in subjective arousal occur because a previous cognitive evaluation has indicated that some learning has occurred. A vague awareness of only one of the various indices would seem a good cue for many item-specific judgments of knowing.

A different approach to the role of physiological cues as they correlate with degrees of learning is the application of closed loop theory to verbal learning (Adams, 1967, 1968). The theory maintains that both a memory trace and a perceptual trace are formed during learning. The perceptual traces are seen as laid down by both response-produced stimuli and by environmental stimuli. During later recall or recognition, the matching of sensory feedback from a response with the appropriate perceptual trace serves as a discriminative cue regarding the correctness of that response.

The finding that recall is positively related to the number of sense modalities in learning (Murray, 1965) is cited as support for the theory. Similarly, an experiment in short-term memory by Adams, McIntyre, and Thorsheim (1969) found poorer recall following interference with both audio- and tactual-proprioceptive feedback than with either single modality interference or with no interference. If perceptual traces are indeed laid down during learning, their relative strength presents a potential cue for a correct JK. Such perceptual traces may even be theorized to underlie the previously mentioned correlation between learned states and physiological changes.

Learning-memory and memorial self-knowledge. Research on learning and memory may be divided into experimentation relevant to FK judgments and to JK ratings.

Some of the FK studies since the original Hart (1965) experiments have sought to extend the generality of the phenomenon. Thus, Hart (1967a) later validated the concept of memory-monitoring in the PA learning of trigrams. Similarly, Yarmey (1973) has shown the accuracy of FK judgments regarding photographs of famous persons and the names of the people pictured. On a more general level, Gruneberg and his associates have shown the FK to be a valid predictor of later spontaneous recall for presently unrecallable information (Gruneberg, Smith, & Winfrow, 1973) and of cued recall (Grueneberg & Monks, 1974).

Other researchers have attempted to discover the bases of FK judgments. Blake (1973) attempted to distinguish recalled partial information from all other retrieved attribute information which might serve as a potential basis for a positive FK. In two experiments, the Peterson and Peterson (1959) short-term memory paradigm was used in which brief presentations of consonant trigrams were soon followed by a recall attempt. If recall was incorrect, an FK rating was requested prior to a recognition selection from eight alternatives. In the first experiment, the eight alternatives were devised so as to limit the utility of partial knowledge. That is, the incorrect alternatives were similar to the target by containing up to two of the correct letters. The same alternatives were shown to all subjects. The results indicated that FK judgments in the short-term memory task were similar in accuracy to FK judgments regarding general information questions. The second experiment differed only in that the recognition alternatives given the

subject always contained whatever correct information the subject had recalled. That is, if the target were KSW and the subject had recalled LSV, all eight alternatives contained S as the middle letter. FK accuracy was reduced substantially. Further, the likelihood of a positive FK judgment was only slightly less than in the first experiment and mirrored the first experiment by increasing as a function of the number of trigram consonants which had been correctly recalled. The role of partial recall in a trigram recall-FK task, then, is rather striking. Still, the difference between Hit and Miss rates was reliable. Partial recall would thus seem to be important in a FK judgment more as a constituent of retrieved attribute information than as the sole basis for a positive FK. Although Blake did not further specify the nature of such "retrieved attribute information," the present theorizing would allow the inclusion of process and memory knowledge. For example, if the retrieved attribute knowledge included a memory for a significant amount of repetition for an item, such information may be expected to be a very good predictor of future recognition performance.

Other researchers have emphasized the role of general knowledge in FK accuracy. DaPolito, Guttenplan, and Steinitz (1968) asked subjects general information questions and, if recall was incorrect, for FK judgments. In contrast to other research, however, subjects were asked to eliminate as many incorrect alternatives as possible before they indicated their recognition choice for each question. The results offered the possibility that subjects are better at eliminating incorrect alternatives than at recognizing the correct response. However, the value of such information may be expected to vary directly with the similarity of the alternatives present as answers.

The value of general knowledge in answering specific questions has been more directly examined in terms of the "tip-of-the-tongue" (TOT) phenomenon first studied by Brown and McNeill (1966). These authors read definitions of relatively unfamiliar words. When subjects were unable to recall a defined word but felt the word on the tip of their tongue, they were found to have some knowledge of the initial and final letters in the missing word, the number of syllables in the word, and the location of the primary stress.

Koriat and Lieblich (1974) have presented a more refined methodological analysis of the TOT data. Of most theoretical importance were tests of the hypothesis that subjects in a TOT state possess information regarding unretrieved words which is different from class or population information. Subjects' answers regarding both initial-final letters and number of syllables were found to have some basis in general information regarding the probability of occurrence of such characteristics in the population of infrequent words. Nevertheless, item-specific detection was also substantiated. The general information had by subjects was not sufficient to explain their success rates.

Rubin (1975) has differed from the other investigators of TOT states by his consideration of a rather small number of definitions presented to a large group of subjects. Only words which subjects had indicated to be the experimenter's targets were scored. Initial letters and final clusters were recalled with impressive proportions when subjects claimed to be in the TOT state.

In general, then, research on retrieval from long-term memory has shown the importance of general knowledge in FK judgments. Similarly, such experimentation has strengthened a role for knowledge of item-

specific information in the FK judgment. If the JK can be shown similar to the FK, similar conclusions about the role of item-specific information in the JK may be expected.

Research in the area of learning and memory which is relevant to the JK may be categorized into studies on self-pacing and studies on confidence ratings.

Zacks (1969) has replicated the previously mentioned results of Bugelski and Rickwood (1963) regarding the total time invariance for experimenter-paced and subject-paced tasks. Further, subjects were found to devote more study time to difficult than to easy pairs. Lastly, self-reports in the experimenter-paced conditions indicated the existence of greater covert rehearsal for the pairs which were empirically defined as difficult to learn.

The role of self-knowledge is also manifest in an experiment by Royer (1973). Subjects were asked to study Turkish-English vocabulary in a PA task. One group of subjects was instructed to study to mastery while using a self-test method in which they were allowed to look at the reverse side of an index card which contained only the stimulus term. The subjects in a second group were temporally yoked to the first group but had no ability to self-test. Study by the third group was self-paced but overt self-testing was not allowed. The two self-paced groups learned more items than the yoked group--without significant differences in amount of study time. Subjects seem able to assess their present state of knowledge without any direct chance for self-testing.

Among the experiments on confidence judgments, two offer comments as to the potential basis of such judgments. Adams and Adams (1960) had subjects evaluate the correctness of the spelling of various words.

Confidence ratings were found to increase with accuracy as well as with familiarity. Another experiment by Adams and Adams (1958) found a similar monotonically increasing relationship between familiarity and confidence regarding judgments of synonymy and antonymy.

With a different goal in mind, Suboski and his associates have also investigated confidence judgments in the PA learning paradigm. An early experiment (Suboski, Pappas, & Murray, 1966a) used the study-test method with lists of ten letter-number pairs. During test periods which occurred after each of three presentations of a list, subjects also rated their confidence in their responses on a five-point scale. The probability of a response being correct was found to increase monotonically from .20 to about .90 for the five confidence levels. More important was the finding that, on the second test trial, the probability of the highest confidence rating (5) for twice recalled items was an increasing function of confidence during the first test trial. The authors felt this to support the hypothesis that confidence ratings reflect the strength of the learned association between stimulus and response terms.

Suboski, Pappas, and Murray (1966b) also replicated the experiment in a Study-Test-Test paradigm. Presentation of each list was followed after 15-sec. by two tests, each 15-sec. apart. Letter-number items correct on the first test trial but incorrect on the second test trial (CN items) were found to have been given significantly lower confidence ratings on the first trial than items correct on both trials (CC items). Likewise, items incorrect on the first but correct on the second test (NC items) were given reliably higher confidence ratings on the second trial than were given to items in the reversed situation (CN items)--

even though the ratings had been higher on the first trial for the CN items. Similar results have also been obtained by means of signal detection analysis of both recognition tasks (Pappas & Suboski, 1966) and Study-Test-Test tasks (Suboski, 1967).

However, others have argued that such confidence ratings are not mediated by an item's response strength. Thus, Grasha, Schumsky, and Elliott (1973) investigated relationships between short-term recall, intrusions, and subjective certainty ratings regarding recalled items. Mean ratings for intrusion items as a function of initial repetition condition were found not to correspond to the probability of occurrence of intrusion items. The intrusions were also described by a conditional probability measure which was meant to adjust for response availability. Response availability was assumed to be differential because there were different numbers of repetitions for items as well as different retention intervals. However, the mean rating data did not correspond to the conditional probabilities of intrusions. The authors felt that such noncorrespondence would not be expected if confidence ratings directly reflected response strength. Because the confidence ratings decreased over the initial retention interval and increased as a function of the number of initial repetitions, Grasha, Schumsky, and Elliott concluded that such ratings reflect relative certainty regarding correctness. In the present terminology, "certainty regarding correctness" would seem to refer to process knowledge because the ratings appear to be based on general principles.

The results of a continuous PA learning experiment by Bernbach and Bower (1970) also offer support for the role of nonspecific factors in confidence ratings regarding recalled items. Subjects were found to have

a bias to give higher confidence ratings for later presentations, independent of the correctness of recall.

A different source of confidence ratings was studied in an experiment by Arbuckle and Cuddy (1969), namely, subjects' "ability to discriminate differences in associative strength." In terms of the present paper, this would seem to refer to stimulus knowledge. Subjects were to learn lists of five, six, or seven letter-number pairs at a presentation rate of 3-sec., with one study and one test trial per list. On the first trial, subjects predicted whether or not they would recall the appropriate response term on the next trial.

The predictions of these well-practiced subjects were much more accurate than chance expectations. For each subject, the frequency of predictions of future recall was also very similar to the frequency of correct recalls. Other subjects were then asked to rate the perceived difficulty of some of the pairs. It was found that the modal difficulty rating of the pairs decreased with the relative frequency of predictions of recall. It may be pointed out, however, that the rated pairs were limited to ones which had previously been given the same YES or NO rating by the pair of subjects who had studied them.

A second experiment was also performed in which naive subjects were tested by the probe technique (only one item from a list was "probed" during recall). The lists contained five noun PAs. The results indicated that for the first four serial positions, recall was a decreasing function of previous predictions. Words in the fifth position were recalled almost perfectly regardless of the prediction. Because such judgments did not interfere with learning as measured by the recall of nonjudging control subjects, the authors theorize that the assessment of

information required overtly by such predictions is made covertly in standard PA learning situations. The results would seem to offer clear predictions for the present inquiry.

The Present Research

This study poses four questions relative to judgments of knowing. Because of the general lack of knowledge about predictions of future recall performance, the questions are more empirically than theoretically motivated. However, the data also offer some sense of direction about the basis for the JK.

The first question asks whether subjects can improve JK accuracy with practice. The second question examines the relationship between ease of learning judgments and JK ratings for the same items. Thirdly, subject variables are utilized to test the validity of theoretical predictions. That is, if memorial self-knowledge is involved in the JK, then subjects adept at FK judgments ought also to be proficient at the JK. Similarly, if capability at the JK is relevant to school performance, high correlations should be expected between JK performance and scholastic achievement scores. To the extent that high correlations are not found, similarity in task requirements becomes less likely. Lastly, the relationship between learning ability and the EL, FK, and JK tasks is examined.

At this point it will be helpful to outline the actual research procedure. Table 1 indicates the progress of subjects through the experiment's stages. The first stage was relevant for only half the subjects (Group EL) who made EL judgments on the word pairs they were to learn later. In the remaining stages, all subjects performed the same

TABLE 1

Outline of sessions, stages, and tasks

Session I.

- Stage One. EL ratings for half the subjects (Group EL)
- Stage Two. Five study-test trials on letter-number PAs for all subjects
- Stage Three. Four lists of word PAs for all subjects

Tasks

Trial	1	2	3
	Study	JK	Test

Session II.

- General information questions and feeling of knowing judgments for all subjects

tasks.

The second stage involved five PA learning trials on letter-number pairs. The intent of this stage was to give subjects familiarization with the PA task as well as to provide a basis for separating subjects into good and poor learners.

The third stage involved the learning of four lists of word pairs by the study-test method. A pilot study indicated that the average subject could recall about half of each list's items after one study trial. When learning is at such a level, process knowledge should not be a good guide for JK ratings. For each list, then, a single study trial preceded a trial during which judgments of knowing were made. The next trial was the test trial which evaluated JK accuracy.

In a second session, subjects were asked a series of general information questions concerning such areas as history, literature, art, and natural science. For incorrectly answered items, an FK judgment was requested. This was immediately followed by a recognition test for the correct answer from four similar alternatives.

Question 1. Accuracy, practice, and the Judgment of Knowing

A finding that the JK is an accurate predictor of recall would be to a large extent a replication. Therefore, the major concern of this question is the effect of practice on the JK. Because of the introductory nature of the present study, this concern is more data-oriented than theoretical. Specifically, information is sought regarding the presence or absence of improvement in JK accuracy with practice, regarding the relationship of any improvement to the specific type of JK prediction (positive or negative), and regarding the role of subjective bias

in any listwise changes.

To motivate subjects to learn and to recall as many items as possible, this experiment involved subjects in a game situation. Five points were gained for each recalled response term and five points lost for each nonrecalled term. Thus, if evaluation apprehension tempted a subject to withhold recall of a word previously rated as unrecallable, it was hoped that the cost-gain ratio would favor recall rather than withholding.

To encourage subjects to make accurate judgments of knowing, the game also involved bonus and penalty points on the basis of JK accuracy. These were smaller in value than the recall-nonrecall points in a further attempt to discourage either differential study or potential withholding. However, the "game-points" also increased parallel to the magnitude of the JK so that extreme JK ratings gained (or lost) more points when the predictions proved correct (or incorrect) than did moderate JK ratings. It was hoped that the competitive nature of college students, together with the exhortations of the experimenter to improve their point total relative to that on the previous list(s), would preclude subjects from using the middle points of the JK scale unless they felt their present state of knowledge to be in fact intermediate.

Question 2. The EL judgment and the Judgment of Knowing

A concern which past research has not evaluated on an intrasubject level is the extent to which subjects might base their judgments of knowing on such stimulus characteristics of the to be learned material as imagery, association value, or pronunciability rather than on an examination of their present state of knowledge of the item. To show

that the two factors are confoundable, an experiment by Lovelace (1974) may be cited. Subjects were first given three successive study trials on a 60-item list of words in preparation for later free recall. During the fourth presentation, subjects rated each item for their confidence regarding its recall on a later test. Across subjects, average recall on the next trial was a monotonically increasing function of the previous confidence ratings. Though an interpretation of the data as indicating that subjects are quite aware of the contents of their memories might be expected on the basis of previous theorizing, it is also possible that subjects were basing their confidence judgments on stimulus knowledge. That is, Lovelace's stimuli were as low in imagery value as 2.5 on the seven-point imagery scale used by Paivio, Yuille, and Madigan (1968). Given the list's frequency, concreteness, association value, and word length restrictions, possible stimuli include such words as "reaction" and "advice", together with words such as "sugar" and "ankle". To the extent that learning scores correlate with imagery values (cf. Paivio, 1969), the average subject giving judgments of knowing on the basis of this stimulus characteristic would be likely to "know" his or her present state of knowledge quite well. However, it cannot be disputed that the three 3-sec. study trials given these subjects should have led to better learning of the high than the low imagery terms (Paivio, Smythe, and Yuille, 1968). Thus, memory examination leading to judgments of knowing would possibly yield ratings which correlate with imagery values.

To examine the relationship between EL judgments and the JK, the present research utilized a PA learning task in which high frequency three-letter nouns were stimulus terms. Homogeneous stimulus terms were chosen to prevent subjects from basing their judgments of knowing more



on stimulus than on response term characteristics. The response terms were all two-syllable nouns of rather low frequency of usage in the English language. Because of the generally potent effects of imagery on learning, even on the response side (Smythe & Paivio, 1968), each list's response terms sampled the entire range of imagery values found in the Paivio, Yuille, and Madigan (1968) norms.

Prior to the JK task, therefore, half the subjects made EL judgments on the to be learned material. Specifically, they were asked to indicate whether the word-pairs appeared to be hard or easy to learn in a PA task.

It is hoped that the present lack of information about intrasubject similarity in the two ratings can be remedied. The presence of such ratings also allows their comparison with JK ratings as to predictive accuracy. Further, the present design allows an examination of the influence of prior experience with the stimuli on the nature and accuracy of later JK ratings.

Question 3. Feeling of knowing judgments, scholastic achievement, and the Judgment of Knowing

The third area of experimental concern is an attempt to heed the suggestion of Underwood (1975) to identify individual-difference variables with theoretical processes. If a theory is to survive, such variables ought to predict the direction of subjects' task performance. In the present case, FK capability and scholastic achievement have been chosen as crucibles regarding two possible facets of the JK. Thus, if the JK involves a subjective memorial examination, accuracy at the JK task ought to correlate with accuracy at the FK task because this by

definition involves some sort of memorial examination. Likewise, if scholastic success requires actual judgments of knowing, scholastic achievement should correlate with JK performance. As the correlations increase, support can be postulated for similar task requirements.

Question 4. Learning ability and the Judgment of Knowing

A third subject variable which is investigated in the present study is learning ability. It is possible that only very fast and very slow learners are adept at the JK. Such subjects can be theorized to base their JK ratings to a great degree on process knowledge regarding their capabilities. A role for item specific knowledge in the JK would be thus strengthened if improvement in JK accuracy over lists were found to occur for subjects of all levels of learning ability. Learning ability may also be investigated for its role in the relationships between EL, FK, and JK ratings.

CHAPTER II

METHOD

Design

Except for EL ratings required of Group EL in the initial stage of this experiment, all subjects had the same tasks. The second experimental stage involved five study-test trials on a letter-number PA list. During the third stage, all subjects learned four PA word lists and made judgments of knowing prior to each list's test trial. In the second session, all subjects attempted to answer general information questions. For incorrectly answered items, subjects made FK judgments and then selected the correct answer from among four alternatives.

The present design thus allows the evaluation of four hypotheses regarding judgments of knowing. The first hypothesis is that JK accuracy will improve with practice. The second is that a JK is related to an EL judgment for the same item. The third hypothesis is that strong correlations will be found when individuals' JK accuracy is compared with their FK accuracies and with their scholastic achievement. The last is that the relationships found in the first three hypotheses will be found for subjects of all levels of learning ability.

Lists

The various stages of the experiment required the creation of three different lists of items. A list of letter-number pairs was formed to separate subjects on the basis of learning ability. A set of word pairs

was constructed to serve as the material for both the EL ratings and the judgments of knowing. Thirdly, for the FK task, a list of general information questions was devised, together with a set of recognition alternatives for each question.

The word pairs for the stage one EL rating task for Group EL were the items studied by all subjects during stage three. Their selection and ordering will be described below.

In stage two PA learning, the stimuli were 16 consonants randomly paired with two-digit numbers as response terms. The numbers were selected from the Battig and Spera (1962) norms to include an approximately equal number of low, moderate, and high association value terms. All numbers differed from each other by at least one digit. A minimum number of digits was repeated. To minimize serial learning, six orders were created.

For stage three PA learning and JK tasks, 64 consonant-vowel-consonant (CVC) nouns were chosen from the Noble (1961) norms to serve as stimulus terms. Semantic, and to a lesser degree, orthographic similarity was minimized. The nouns were all high in association value, ranging from 3.42 to 4.78 on the 0 to 5 point scale used by Noble. All were names of observable entities. The nouns were next divided into four groups of 16 with an attempt to minimize the number of words in each list which had the same first letters. Phonemic similarity was also minimized within each list.

Response terms for the stage three tasks were 64 two-syllable nouns selected from the Paivio, Yuille, and Madigan (1968) norms. All words were of low frequency, occurring from 1 to 10 times per million English words (Thorndike & Lorge, 1944). Orthographic and semantic similarity

was minimized in the selection of these nouns. The entire range of imagery values was sampled so that four levels of imagery were defined, each level having 16 members. The nouns were then rank-ordered by imagery value. Each of the nouns in the first four positions became the first word in a different list, i.e., A, B, C, D. List placement was then reversed for the second set of four nouns so that the fifth ranked noun was placed into List D, the sixth ranked noun into List C, etc. This alternation of list placement for every four imagery rankings produced approximately equal lists in terms of mean imagery values. The two-syllable nouns were then randomly paired with the CVC nouns in each of the four lists under the restrictions that no pair of words begin with the same letter and that obvious associations be eliminated. Five orders were created for each of the four lists. Four of the orders served as study and JK lists and one as the constant test list. The study and JK orders were formed under the restriction that the first two pairs in each order were not present in the final two positions of the preceding order. The test order for each list was randomly determined. However, the first two items in the test order were never present in the last two positions of any study and JK list.

Four EL lists of 16 items each were formed from the four JK lists. To create an EL list, one item was randomly selected from each of the four imagery levels of every JK list. The EL lists were thus similar in imagery level among themselves and in relation to the JK lists. In each list, the first two items were always from the highest and lowest quartiles of response term imagery. On alternating lists, the first item was from the highest or lowest imagery level. Although there was only one internal order for each list, the permutations of the four lists

allowed 24 different orders of presentation.

Lastly, 52 general information questions were constructed. These were intended to have answers which would be relatively difficult for first year college students to recall but which could be recognized since the information was of the kind likely to have been experienced previously. Care was taken to insure that the questions allowed plausible distractors in the recognition task. The areas of history, literature, natural science, art, sports, geography, and popular culture were sampled approximately equally. An example of a history question is: What Spaniard extensively explored the U.S. southwest? The alternatives were: V. de Balboa, H. Cortés, H. de Soto, F. Coronado. No more than three questions from the same area followed each other contiguously. The recognition alternatives were arranged so that the correct answers appeared in each of the four positions equally often.

Procedure

As Table 1 indicates, the first task for half the subjects was the making of EL judgments on the word pairs which would later be learned as paired-associates. The ratings were made in a room different from that in which learning was later required. No slide projection equipment was present. To further the impression that different tasks would follow the making of EL ratings and hence to minimize attempts by subjects to study the items while making EL judgments, the testing room also contained stacks of mimeographed materials. The following instructions were read by the experimenter as the subject followed on his printed instructions.

You are being asked to make what are called "Ease of Learning

Judgments." That is, you will be given four sheets of paper. On each sheet are typed 16 pairs of words. You will be asked to picture a situation in which the 16 items on the list will be presented one pair at a time for study. You will also be asked to assume that you are going to study each pair of words so that later you can give the right-hand word to the cue of the left-hand word. Each pair of words will be presented for only three seconds at a time. Then, after the entire list has been presented once, you will be given the left-hand word that is paired with each. As during the study period, you will only have three seconds to give the correct word. The pairs of words will then be presented again, but in a different order, and you will be asked to study the same words again. After this, there will be another test period. After many such study-test trials, you should be able to give all the right-hand words correctly. This then is the situation you are being asked to keep in the back of your mind.

OK? Again, first you see two items and then only one and you are to give the second item.

What you will be specifically asked to do is to make judgments about the speed with which you would learn the typed pairs of words in such a situation. That is, some pairs of words are rather easy to learn, while others are much harder to learn. For example, RED - PETAL is probably much easier to learn than RED - XYLEM, even though both refer to parts of flowers. What we want you to do is to give your judgments about how easy or hard certain pairs of words would be for you to learn.

To make these judgments, you are asked to use the straight line which will be typed next to each word pair. Consider this line as a scale of learning difficulty in which the left-hand side indicates EASY to learn items and the right-hand side stands for HARD to learn items. Thus, for a pair of words like RED - PETAL, you would probably put a mark somewhere on the left-hand side of the line. For a pair like RED - XYLEM, on the other hand, you would be more likely to put a mark somewhere on the right-hand side of the line. That is, in making your ease of learning judgments, look at each word pair and try to see where they would fall along such a scale. If you think the word pair would be easy to learn, place a mark somewhere on the left-hand side of the scale; for a word pair that is hard to learn, put a mark further over to the right.

Use as much or as little of the scale as you wish.

Any questions?

The subject was then handed four sheets of paper face-down. All were labeled "Ease of Learning Judgments." On each sheet were typed 16 pairs of words downward along the left-hand margin. Next to each pair was a line which ran for six inches across the page. Above the first line were the words EASY and HARD on the left and right ends of the line,

respectively. Subjects self-paced their EL ratings, turning over one page at a time. As each sheet was completed, it was returned to the experimenter, who also recorded the time for each list.

Group EL subjects were next taken to the experimental room. For Group NoEL subjects, who did not make EL judgments, all tasks occurred in the experimental room. All subjects were tested individually.

Stage two was a PA learning task. The stimulus terms were 16 consonants and the response terms two-digit numbers. Subjects were given PA instructions for the study-test procedure. They were left to their own devices during the study trials and wrote down the appropriate numbers on unnumbered, lined answer sheets during the test trials. The 16 pairs were presented via a slide projector at a 3-sec. rate with a 3-sec. intertrial interval. There were five study-test cycles. No feedback was given.

After a two minute rest interval, subjects were told that they would next be asked to learn some pairs of words in a manner similar to that used for the letter-number pairs. Group EL subjects were informed that these were the same word-pairs they had previously rated. Subjects were also asked to consider themselves in a game situation in which five points would be awarded for each correct response on the test trial and five points lost for each incorrect response. They were therefore asked to learn as many word-pairs as possible to maximize their point total. The words were presented for one study trial at a 3-sec. rate.

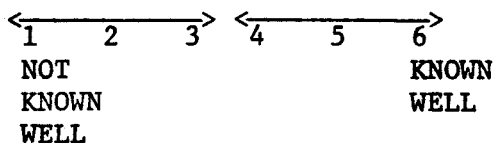
Prior to the second presentation, subjects were handed a folded sheet of paper (Table 2). The portion of the paper which the subject first viewed was labeled "Judgment of Knowing." Beneath this was a six-point scale on which NOT KNOWN WELL and KNOWN WELL were typed under-

TABLE 2

Information given subjects regarding the game-point manipulation

JUDGMENT OF KNOWING

SCALE



For each response word you recall: +5 points.

and

BONUS or PENALTY based on JK accuracy:

JK	1	2	3	4	5	6
	-3	-2	-1	<u>+1</u>	<u>+2</u>	<u>+3</u>

So.

If you predict recall for a word and you do recall that word, you can get as many as 8 points.

For each response word you do not recall: -5 points.

and

BONUS or PENALTY based on JK accuracy:

JK	1	2	3	4	5	6
	<u>+3</u>	<u>+2</u>	<u>+1</u>	-1	-2	-3

So.

If you predict that you will not recall a word and you do not recall that word, you may lose as few as 2 points.

neath the first and last three numbers, respectively. A straight line was present under the numbers 1-3 and its continuation under 4-6. Outward pointing arrows were also present on both ends of each line to emphasize further the dichotomous nature of the scale. Subjects were then read the following instructions.

You have now studied the entire list. The presentation after this one will be a test trial. However, before this test trial, you will see the list again and this time you will be asked to predict how likely you are to recall the response term on the test trial.

That is, you are being asked to indicate how well you know each pair of words. You are being asked to make what we are calling a "Judgment of Knowing." Further, you are asked to make this judgment for each pair on the basis of a six-point scale. Thus, a judgment of knowing of 4, 5, or 6 should mean that you know the pair of words rather well and will probably be able to recall the right-hand word on the next trial, while a judgment of knowing of 3, 2, or 1 should mean that you do not know a particular word-pair well enough to recall the right-hand word on the next trial. OK?

To help you to try to make these judgments as accurately as possible, remember again that you are in a game situation in which you will get five points for every word recalled and will lose five points for every word not recalled. Further, in this game, you will get bonus or penalty points for judgment of knowing accuracy. Please unfold your sheet of paper so that you may get a picture of the bonus-penalty point system.

Note first that there are no more than three bonus or penalty points. Thus, if you predict recall for a word by giving a judgment of knowing of 4, 5, or 6 and you later recall the word, you will get one bonus point for a JK of 4, two bonus points for a JK of 5, and three bonus points for a previous JK of 6. However, if you predict recall for a word and do not recall it, you will lose the same number of points--plus the automatic five points lost for not recalling a word. So, if you're not sure you know a pair well enough for later recall, your best bet is to write down 1, 2, or 3--at least then you will only lose a few points if you do not recall the word. But please remember that you are being asked to try to get as many points as possible. So try to be as exact as you can in your predictions.

Also, since these words will be presented at the same rate as previously, try to write down your JK as quickly as possible for every item.

The game-points instruction sheet was then replaced by a sheet of paper with lines and subjects were informed that they were to write down

their JK numbers on this paper. The same scale that had been in the top quarter of the instruction sheet (Table 2) was also on the top of the answer sheet. The test trial immediately followed the JK trial. Presentation rates were the same on all trials, approximately 3-sec. for each item. A tape-recorded signal connected to the slide carousel projected each slide at the same rate.

Subjects rested in the 3-min. interval between the first and second PA lists. After the experimenter calculated the subject's game-points, he informed the subject of his or her point total and asked the subject to try to improve JK accuracy on the next list and to get more points. To further motivate the subjects, the experimenter asked general questions about how subjects were doing and suggested that the task would get easier.

The same procedure was followed for the remaining three lists, with a 3-min. interval between lists. However, there were no further interruptions between study and JK trials to explain the game-points manipulation or the JK. After being informed of their scores on the fourth list, subjects were asked to make an appointment for the second session, which occurred about a week after the first session.

During the second session, subjects were given index cards on which were typed general information questions. They read each question aloud and attempted to answer it. Subjects were asked always to give an answer. However, if resistance was offered to guessing regarding a question about which the subject was totally uninformed, the point was not pressed. If the answer was correct, subjects were told so and the next card was presented.

The lower portion of each question-containing index card contained

a six-point scale. The scale was similar to the JK scale but was labeled FEELING OF KNOWING. CANNOT SELECT and CAN SELECT were typed underneath the numbers 1-3 and 4-6, respectively. After a subject's first incorrect answer, he was asked if he had any "feeling" of knowing the correct answer well enough so that he would be able to select the correct answer from among four alternatives. Subjects were asked first to think in terms of a dichotomous FK judgment and then to make finer distinctions. They were also reminded that the alternatives would be similar to the correct answer. Immediately after a subject gave a FK rating, the card with the question's four recognition alternatives was presented. The subject then indicated his selection, together with a second choice. It was emphasized that the first choice was the important answer and that the second was mainly to verify the validity of alternatives. The subjects were allowed as much time as they wished for recall, FK judgments, and the recognition task. The experimenter kept a record of subjects' responses and informed subjects both regarding the accuracy of their choices and regarding the correct answer.

Afterwards, subjects were informed of the basic details of the experiment. Subjects were also asked to sign release forms allowing the experimenter to examine their Scholastic Aptitude Test and/or American College Test scores. It was indicated that this was not necessary to get credit but would be helpful for the study. All subjects signed the release forms.

Subjects

A total of 48 Loyola undergraduates participated as subjects in this study in partial fulfillment of their introductory psychology

course requirements. All subjects were naive to FK and JK tasks.

Assignment to the two conditions was alternated on the basis of appearance in the laboratory. The data of three subjects could not be utilized due to experimenter error.

CHAPTER III

RESULTS

Question 1. Accuracy, practice, and the Judgment of Knowing

The initial question has two concerns. First, the JK must be evaluated with regard to its accuracy as a predictor of future recall performance. Secondly, the JK is to be observed over lists to determine if such predictive accuracy increases with practice. The two concerns are considered for the JK both as a dichotomous measure and as a six-point scale.

Table 3 indicates terminology from signal-detection theory which will be employed in describing the data. The columns refer to the JK ratings which were given by subjects to the paired-associate items. Ratings which predicted nonrecall (1, 2, or 3) are symbolized by \overline{JK} ; ratings which predicted recall (4, 5, or 6) by JK. The rows distinguish items as either recalled or nonrecalled. Four joint events are thus possible. If item recall followed a prediction of 1, 2, or 3, the item is scored as a Miss. That is, the recall of an item rated 1, 2, or 3 indicated a previous failure to assess an item's future recallability. However, if nonrecall followed a prediction of 1, 2, or 3, such a correct expectation of an item's future nonrecall is labeled a Correct Rejection. Similarly, if item recall occurred after a rating of 4, 5, or 6, this accurate prediction of an item's future recall is considered a Hit. However, if a JK rating of 4, 5, or 6 was followed by nonrecall, this is

TABLE 3

Terminology from signal detection theory as applied to JK accuracy

	$\overline{JK} = 1, 2, 3$	$JK = 4, 5, 6$
Item Recall	Miss	Hit
Item Nonrecall	Correct Rejection	False Alarm

called a False Alarm because the item was incorrectly evaluated as recallable.

Originally, the appropriate measure to evaluate JK accuracy and its improvement over lists was considered to be that used by Hart (1965, 1967a, 1967b) and Blake (1973) to assess FK accuracy. Recall probabilities for items given JK ratings of 1, 2, or 3 (Miss proportions) were to have been subtracted from recall probabilities for items given ratings of 4, 5, or 6 (Hit proportions). The JK was to be considered a valid predictor of future performance if recall followed positive JK ratings more than negative JK ratings.

However, this measure can be shown to be a misleading index of performance when Hit or Miss proportions are based on a small number of instances. As examples, consider the actual data in Table 4. When the accuracy measure is the difference between Hit and Miss proportions, Subject A's performance on Lists 3 and 4 is seen to vary greatly. However, the .43 increase in sensitivity from $-.07$ to $.36$ is mainly due to the one item change in the Miss proportions, from $2/2$ to $1/2$. Similarly, Subject B's change in JK accuracy from $.94$ on List 3 to $.00$ on List 4 results principally from the two erroneous predictions of nonrecall on the last list. Since 39 of the 48 subjects had at least one of their eight listwise proportions based on 4, 3, 2, or 1 instances, the Hit less Miss measure would seem to imply more listwise variability within subjects than is present in the data.

The proportion measure likewise accentuates variability between subjects with respect to overall JK accuracy. For example, Subjects A and B in Table 4 have identical Hit proportions. Yet, a difference of three Miss items places Subject A at the 11th percentile and Subject B

TABLE 4

Comparisons of relative (JK Accuracy) and absolute (JK Errors)
measures of JK sensitivity

Subject A.

	Hit	-	Miss		JK Accuracy	JK Errors
List 1.	12/12	-	4/4	= 1.00 - 1.00 =	.00	4
List 2.	11/12	-	3/4	= .92 - .75 =	.17	4
List 3.	13/14	-	2/2	= .93 - 1.00 =	-.07	3
List 4.	12/14	-	1/2	= .86 - .50 =	.36	3
Overall	48/52	-	10/12	= .92 - .83 =	.09	14

Subject B.

	Hit	-	Miss		JK Accuracy	JK Errors
List 1.	10/12	-	1/4	= .83 - .25 =	.58	3
List 2.	9/10	-	4/6	= .90 - .67 =	.23	5
List 3.	15/16	-	0	= .94 - 0 =	.94	1
List 4.	14/14	-	2/2	= 1.00 - 1.00 =	.00	2
Overall	48/52	-	7/12	= .92 - .58 =	.34	11

at the 57th percentile when the 48 overall accuracy scores are evaluated by comparisons of Hit and Miss proportions.

To examine JK accuracy and its improvement over lists while avoiding the above statistical problem, the present research will utilize measures of sensitivity and criterion change which were employed by Underwood (1974) to compare recognition memory and frequency judgment tasks. The measure of sensitivity is the sum of a subject's erroneous predictions, Misses plus False Alarms. Because this measure is based on absolute numbers rather than on proportions, it is not overly influenced by a small number of instances in either \overline{JK} or JK categories.

As can be seen in the last two columns of Table 4, number of JK errors as a measure of subjective sensitivity leads to less ambiguous descriptions of performance in the present task than does the difference between Hit and Miss proportions. Whereas the proportion measures led to accuracy scores of $-.07$ and $.36$ on Lists 3 and 4 for Subject A, number of JK errors as a measure of sensitivity interprets the two lists as both involving three JK errors. Interlist differences are similarly minimized for Subject B on Lists 3 and 4. While the measures based on proportions led to accuracy scores of $.94$ and $.00$ on Lists 3 and 4, number of JK errors describes the two lists as having one and two errors, respectively.

The measure which reflects changes in the criteria that subjects used in predicting their future recall performance will be the difference in error types divided by their sum, $\text{Miss} - \text{False Alarm} / \text{Miss} + \text{False Alarm}$. A positive proportion thus indicates a bias toward predicting nonrecall; a negative proportion a bias toward predicting recall.

The JK as a dichotomous measure. Three comparisons were performed

to establish the general adequacy of the dichotomous JK as an accurate predictor of future recall performance. To establish the overall accuracy of the JK, the total number of right and wrong predictions were compared. That is, Misses and Correct Rejections were subtracted from Hits and False Alarms for each subject. The resultant difference is the standard correction for guessing in a two-alternative, forced-choice situation. Thus, the average subject made 45.3 correct predictions. This is significantly greater than the average 18.7 incorrect predictions, $t(47, \text{two-tailed}) = 16.51, p < .001$.

A further examination of JK accuracy is the difference between mean Hit (.75) and Miss (.46) proportions. Although the two components have faults as descriptive measures for the present research, the difference is highly reliable, $t(47, \text{two-tailed}) = 8.35, p < .001$.

To evaluate the separate adequacies of positive and negative JK ratings, probabilities of recall and nonrecall conditional on a prediction of recall (Hit) or nonrecall (Correct Rejection) were compared with the actual probabilities as chance expectations. Of the 39.0 items for which the average subject predicted recall, 31.1 (80%) were actually recalled. Mean recall, however, was only 41.8 (65%) of the 64 items. The z -score difference between the conditional and expected recall proportions was 2.19, $p < .05$. Similarly, the average subject predicted nonrecall for 25.0 items. Of these items, 14.2 (57%) were correct predictions. The z -score difference between this conditional nonrecall probability and the actual nonrecall probability of .35 was 3.21, $p < .01$. Subjects would seem to have made predictions of both recall and nonrecall with an accuracy greater than the chance expectations.

The second concern regarding the dichotomous JK is the estimation

of the effects of practice on its sensitivity and bias. To examine the effect of practice on JK sensitivity, the mean number of JK errors (Misses and False Alarms) per list per subject was first analyzed by a 2 X 4 ANOVA in which the first factor was EL judgment (presence or absence) and the second factor was Lists. The means for this comparison (collapsed across the EL judgment variable) are given in the JK Errors column of Table 5. The main effect of EL judgment was significant, $F(1,46) = 6.76, p < .05$. Subjects who did not make EL judgments gave reliably less erroneous JK predictions than did subjects who had previously made EL judgments. Of more importance, however, was the significant main effect of Lists, $F(3, 138) = 4.97, p < .01$. Measured by the number of errors, subjective sensitivity at the JK task can be said to have increased with practice. The interaction was not significant.

The decrease in JK errors over lists (5.60, 4.67, 4.46, 3.96) was also analyzed for linear trend. The linear component was found reliable $F(1,46) = 13.55, p < .01$.

To determine if practice led subjects to change the criteria governing their predictions, the Miss - False Alarm/Miss + False Alarm bias ratios were examined via a 2 X 4 ANOVA in which EL judgment (presence or absence) and Lists were factors. Neither the main effect of EL judgment nor its interaction with Lists was significant, both F 's < 1 . However, the main effect of Lists was reliable, $F(3, 138) = 11.45, p < .01$. As can be gleaned from the last column in Table 5, errors on the first list tended to be underpredictions of recall. To examine the temporal course of the reversal tendency which began on the second list, each subject's number of Miss and False Alarm errors per list was analyzed via a 2 X 4 ANOVA in which JK Error Type (False Alarms and

TABLE 5

Mean performance per subject relevant to the evaluation of the
sensitivity and bias of the dichotomous JK

All Subjects:

List	Hits - FAs* = Pos. Acc.	CRs - Misses = Neg. Acc.	JK Errors	$\frac{M-FA}{M+FA}$
1	6.25 - 1.27 = 4.98	4.15 - 4.33 = -.18	5.60	.54
2	7.88 - 2.12 = 5.76	3.46 - 2.54 = .92	4.67	.15
3	8.23 - 2.38 = 5.85	3.31 - 2.08 = 1.23	4.46	-.05
4	8.73 - 2.15 = 6.58	3.33 - 1.79 = 1.54	3.96	-.09

Group EL:	JK Errors	$\frac{M-FA}{M+FA}$	Group NoEL:	JK Errors	$\frac{M-FA}{M+FA}$
List 1	6.04	.56	List 1	5.17	.52
List 2	5.37	.07	List 2	3.96	.23
List 3	4.88	-.13	List 3	4.04	.04
List 4	4.38	-.14	List 4	3.54	-.04

* FAs = False Alarms

Pos. Acc. = JK Positive Accuracy

CRs = Correct Rejections

Neg. Acc. = JK Negative Accuracy

M-FA/M+FA = Miss - False Alarm/Miss + False Alarm

Misses) and Lists were factors. The effect of major importance was the significant JK Error Type X Lists interaction, $F(3, 138) = 14.24, p < .01$. (The mean values are found under the two headings in Table 5.) When the difference between the two types of errors was next examined by simple effects analyses for each list, the difference was found reliable for the first and second lists, F 's > 10.99 , but not for the last two lists, both F 's < 1 . Thus, the decrease over lists in the overall number of errors was accompanied by a change from an initial bias towards erroneous predictions of nonrecall to an eventual impartiality of erroneous judgments of knowing.

To determine the locus of improvement, the effect of practice was also analyzed separately for predictions of recall and nonrecall. (Such analyses, of course, must be viewed with caution because repeated analyses of the same data increase the probability of a significant effect due to chance alone.) To control for the differential frequencies of use of the two predictions, the dependent measure was again formed by the standard correction for guessing in a two alternative forced-choice situation: right less wrong for each type of prediction. The means of the corrected predictions of recall (JK Positive Accuracy) and nonrecall (JK Negative Accuracy) are given in the third and sixth columns of Table 5. These are each preceded by their components.

To analyze positive JK predictions, a $2 \times 4 \times 2$ ANOVA was performed in which the first two factors were EL judgments (presence or absence) and Lists. The third factor consisted of two dependent measures, namely, each subject's number of Hits and False Alarms. It was named JK Positive Accuracy. The main effect of JK Positive Accuracy was significant, $F(1, 46) = 20.75, p < .01$. The greater number of Hits than False Alarms

per subject replicates the earlier finding of the adequacy of the positive JK as a predictor of recall. However, although the increase in Hits over lists was slightly more rapid than the increase in False Alarms, this interaction was not reliable, $F(3, 118) = 2.04$. Thus, practice cannot be said to have led to increased accuracy regarding the positive JK. The other interactions were also not significant.

The listwise course of positive JK ratings may also be expressed in terms of proportions of Hits to predictions of recall. The slight decrease over lists (.83, .79, .78, .80) would seem to indicate that the bias against positive predictions on the first list set limits on relative improvement in positive judgments of knowing.

To examine the effect of practice on negative JK predictions, a $2 \times 4 \times 2$ ANOVA was next performed in which the first two factors were EL judgments (presence or absence) and Lists. The third factor was labeled JK Negative Accuracy because it consisted of the two dependent measures Correct Rejections and Misses. The mean performance per subject is found in Table 5. The significance of the main effect of JK Negative Accuracy, $F(1, 46) = 4.35$, $p < .05$, may be interpreted to indicate that negative JK ratings are somewhat accurate predictors of nonrecall. Of more importance is the significant Lists \times JK Negative Accuracy interaction, $F(3, 138) = 4.37$, $p < .01$. Although negative predictions occurred less often with practice, the decrease in the number of Correct Rejections was much less rapid than the decrease in the number of Misses. Expressed as the ratio of Correct Rejections to the number of negative JK predictions, the listwise probabilities of accurate negative JK predictions are .49, .58, .61, .65. For the dichotomous JK, then, the major effect of practice would seem to be this improvement in

predicting nonrecall.

Consideration may also be given to the finding that the NoEL group made significantly less erroneous predictions than the EL group, 16.7 versus 20.7 JK Errors. The Group NoEL superiority over Group EL was similar for both types of errors (t 's = 1.37 and 1.39 for False Alarms and Misses, respectively).

The cause of such a difference may be theorized to be some interference stemming from the familiarization inherent in the EL task. However, other interpretations must first be discounted. Thus, the number of correct recalls in the five-trial letter-number task was compared for groups EL and NoEL. The nonsignificant difference, $t < 1$, would seem to indicate that NoEL superiority is not related to general learning ability as defined by letter-number performance. The difference was also not obviously related to the level of recall on the JK lists. When number of recalls per list were submitted to a 2 X 4 ANOVA in which EL judgments (presence or absence) and Lists were factors, neither the main effect of EL judgment, $F < 1$, nor its interaction with Lists, $F = 1.47$, was significant. The level of recall was also found not to change over the four lists, $F < 1$. Differences in sensitivity between the EL and NoEL subjects will be further observed when the JK is examined as a six-point scale.

The JK as a six-point scale. The second series of analyses of JK accuracy will focus on the full JK scale. Figure 1 contains the recall probabilities for the JK scale values on the four lists. In complement, Table 6 presents the number of recalls for each JK scale position in relation to the frequency of usage of each scale position. With only one exception in four lists, the recall probabilities can be seen to

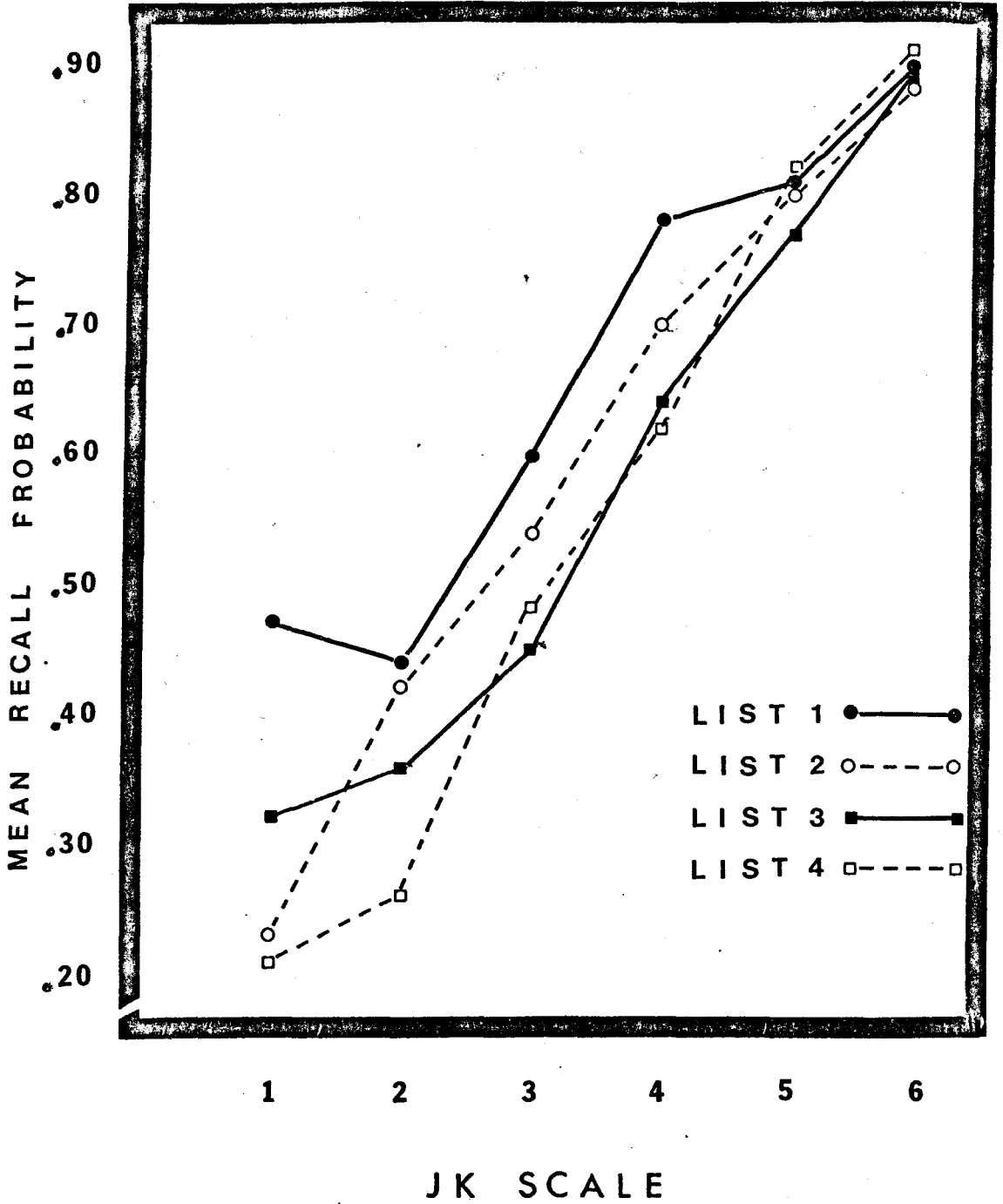


Figure 1. Recall proportions for JK scale values on four lists.

TABLE 6

Recall proportions for JK scale values on four lists

JK Values

List	1	2	3	4	5	6
1	$\frac{49}{105} = .47$	$\frac{62}{141} = .44$	$\frac{97}{161} = .60$	$\frac{113}{145} = .78$	$\frac{69}{85} = .81$	$\frac{118}{131} = .90$
2	$\frac{17}{73} = .23$	$\frac{38}{90} = .42$	$\frac{68}{125} = .54$	$\frac{135}{193} = .70$	$\frac{102}{127} = .80$	$\frac{141}{160} = .88$
3	$\frac{24}{75} = .32$	$\frac{31}{85} = .36$	$\frac{45}{99} = .45$	$\frac{109}{170} = .64$	$\frac{98}{128} = .77$	$\frac{188}{211} = .89$
4	$\frac{13}{61} = .21$	$\frac{18}{68} = .26$	$\frac{56}{117} = .48$	$\frac{96}{155} = .62$	$\frac{95}{116} = .82$	$\frac{228}{251} = .91$
Sum	$\frac{103}{314} = .33$	$\frac{149}{384} = .39$	$\frac{266}{502} = .53$	$\frac{453}{663} = .68$	$\frac{364}{456} = .80$	$\frac{675}{753} = .90$

increase as JK ratings increase.

Examination of performance on the individual scale positions indicates that the recall probabilities for items with JK ratings of 6 or 5 are initially very high and change little over the four lists. Moreover, Table 6 indicates that such accuracy occurs in the context of a considerably increased usage of these ratings. The recall probabilities for the JK scale position of 4, however, show a rather consistent decrease over the four lists. Still, on every list, recall for items rated 4 is higher than recall for items rated 3.

Scale usage predicting nonrecall is also very accurate. The recall proportions for items given JK ratings of 1, 2, and 3 generally decrease over lists. Again, Table 6 points out that such improvement occurs while the number of items receiving these ratings also declines regularly. Practice would thus seem to allow subjects to identify minimally learned items more accurately.

To determine if the six-point JK scale was an accurate instrument for predicting recall performance, the data considered were those on the bottom line of Table 6. If the first three proportions are each subtracted from unity, the results are probabilities of nonrecall conditional on predictions of nonrecall. When these proportions (.67, .61, .47) are assessed against the actual nonrecall proportion of .35, the resulting z -scores for the JK ratings of 1 and 2 (z 's = 4.32 and 3.80, $p < .001$) indicate that the first two scale points were more accurate predictors of nonrecall than the chance expectation. The difference for the JK rating of 3, however, failed to reach acceptable levels of significance, $z = 1.75$. Yet, the trends of the proportions over lists were in the appropriate direction. Likewise, the recall probabilities for

the three positive JK ratings (.68, .80, .90) were compared with the actual recall proportion of .65. Ratings of 6 and 5 were found to be more accurate than the expectation, $z = 3.59$, $p < .01$ and $z = 2.16$, $p < .05$, respectively. The JK rating of 4, however, was no more accurate a predictor of recall than the expectation, $z = .43$. Indeed, the trend over lists was for the accuracy of this prediction to decrease. Except for the JK rating of 4, then, the individual JK scale values would seem to be rather accurate indicators of future recall performance.

In testing for improvement in JK accuracy over lists, examinations were made which weighted usage of the extreme JK scale positions more than usage of the middle scale points. This reflects the approach subjects were asked to take in making their JK ratings. Thus, recall after JK ratings of 3, 2, 1 and nonrecall after ratings of 4, 5, 6 was scored as 1, 2, and 3 negative game-points, respectively. Likewise, recall after JK ratings of 4, 5, 6 and nonrecall after ratings of 3, 2, 1 was scored as 1, 2, or 3 positive game-points, respectively. Table 7 contains the mean scores per subject for negative game-points, positive game-points, and the components of each.

Analyses of JK performance in terms of game-points generally validated previous findings with the number of dichotomous JK errors. Sensitivity as measured by a decrease in negative game-points was found to increase with practice, $F(3, 138) = 5.24$, $p < .01$. Practice also led to changes in the Miss - False Alarm / Miss + False Alarm bias ratio when each type of error was weighted in terms of the game-points involved. After an initial predominance of negative game-points for underpredictions of recall on List 1, subjects were found to lose an equal number of game-points for the two types of errors on Lists 2, 3, and 4 (F 's < 2.56

TABLE 7

JK performance scored as negative game-points, positive game-points,
and their components: Mean game-points per subject.

All Subjects:

	Negative Game- points = FAs* + Misses			Positive Game- points = Hits + CRs $\frac{M-FA}{M+FA}$			
List 1	9.81	=	2.15 + 7.67	20.79	=	12.60 + 8.19	.55
List 2	7.52	=	3.44 + 4.08	22.71	=	15.88 + 6.83	.16
List 3	7.69	=	3.96 + 3.73	24.67	=	18.10 + 6.56	-.04
List 4	6.27	=	3.54 + 2.73	26.42	=	20.21 + 6.35	-.11

Group EL:	Negative Game- points	Positive Game- points	Group NoEL:	Negative Game- points	Positive Game- points
List 1	10.50	18.71	List 1	9.12	22.88
List 2	8.50	20.00	List 2	6.54	25.42
List 3	7.83	22.46	List 3	7.54	26.88
List 4	6.21	24.67	List 4	6.33	28.17

* FAs = False Alarms

CRs = Correct Rejections

$M-FA/M+FA$ = Miss - False Alarm/Miss + False Alarm

in simple effects analyses). Thirdly, separate examinations indicated that both predictions of recall and of nonrecall involved more positive than negative game-points, F 's = 91.61 and 7.30, respectively. Further, while game-points for predictions of nonrecall decreased over lists for both incorrect and correct predictions (cf. columns 3 and 6 in Table 7), the decrease was much less for Correct Rejections than for Misses, $F(3, 138) = 3.25$, $p < .05$. This replicates the earlier finding that predictions of nonrecall become more accurate with practice.

Two important differences were also found between analyses based on game-points and those based on the dichotomous JK. A first difference was discovered when the listwise course of positive JK predictions was examined. A $2 \times 4 \times 2$ ANOVA was performed in which the factors were EL judgments (presence or absence), Lists, and Positive Game-point Accuracy (Game-points for Hit items or for False Alarm items). The means are presented in the fifth and second columns of Table 7. Of most importance is the significant Lists \times Positive Game-point Accuracy interaction, $F(1, 46) = 7.26$, $p < .05$. Game-points gained for Hit items increased more rapidly than did game-points lost for False Alarm items. With practice, greater confidence was expressed in the positive predictions which were eventually followed by recall than in eventually disproven positive predictions. When the listwise progression of the four item types is considered (cf. Table 7), the magnitude of this increase in game-points for Hit items is seen as the most striking effect of practice.

A second difference between the two analyses was the finding that the lesser number of game-point errors for Group NoEL relative to Group EL was not reliable, $F = 1.55$. Thus, the smaller number of dichotomous

JK errors for Group NoEL did not lead to an appreciably lower number of game-point errors. Stated differently, the mean number of game-points lost for an erroneous JK item was slightly greater for a subject from Group NoEL than for a subject from Group EL.

It is thus possible that subjects who have had no experience with stimuli are thereby able to express more confident judgments of knowing than can subjects who have had greater familiarity with the items. This leads to the expectation that Group NoEL will prove superior to Group EL regarding the number of positive game-points gained. The number of positive game-points per subject per list was therefore analyzed by a 2 X 4 ANOVA in which EL judgments (presence or absence) and Lists were factors. The number of positive game-points was found to increase reliably over lists, $F(3, 138) = 9.89, p < .01$. More importantly for the present hypothesis, the NoEL group was found to have gained more positive game-points than the EL group, $F(1, 46) = 5.54, p < .05$. The interaction was not significant, $F < 1$, the means indicating that the difference between the two groups was present from the first list. Increments in situational frequency from the EL task may thus be interpreted to have led to both less accuracy and less confidence in judgments of knowing by the EL group.

That such Group NoEL superiority over Group EL is related to specific ratings rather than to any general response bias was strengthened when each subject's mean JK rating per list was submitted to a 2 X 4 ANOVA in which EL judgments (presence or absence) and Lists were factors. Although the increase in the mean JK over lists was reliable, $F(3, 138) = 18.93, p < .01$, neither the main effect of EL judgments nor its interaction with Lists was significant, both F 's < 1 . For Group NoEL

superiority to exist under such group equality in scale usage would seem to require a rather sensitive assignment of the JK ratings.

The examination of the full JK scale has thus confirmed the main results found with the JK as a dichotomous predictor of future performance. The previously uncovered inability of practice to increase reliably the accuracy of positive JK predictions was also clarified by the finding that practice did lead to the expression of much greater confidence in the future recall of eventual Hit items. Lastly, this analysis also has offered the possibility that involvement with to be learned stimuli prior to a JK task increases the difficulty of later distinguishing recallable from nonrecallable items.

Question 2. The EL judgment and the Judgment of Knowing

The second question examines the relationship between ease of learning judgments and judgments of knowing made on the same items. The relationship will first be studied in terms of the individual correlations between the EL and JK ratings for each subject in Group EL, both as overall and as listwise measures. This concern will also be observed with the data collapsed over subjects.

The second comparison of EL and JK ratings will involve their separate adequacies as predictors of recall performance. Both individual and group data will be included.

Lastly, two secondary analyses will be discussed. The first contrasts JK predictive accuracies when JK ratings are similar to earlier EL ratings with JK accuracies when the two ratings are dissimilar. The second inquires into the relationship between JK ratings and the stimulus characteristics of imagery and recallability.

Correlations between EL and JK ratings by individuals and by groups. For each subject in Group EL, Pearson ~~and~~ Spearman correlations were calculated between EL judgments and judgments of knowing on the same items. The EL judgment values were based ~~on~~ the distance in millimeters from the end position of the EL scale. Table 8 presents measures of central tendency for the correlations between EL and JK ratings. The four lists are considered both separately ~~and~~ as an overall list.

The mean overall Pearson product-moment correlation coefficient for the 24 subjects is .32. If this were an individual correlation coefficient for 64 pairs of items, the value would be reliably different from zero, $p < .01$. The mean overall Spearman rank-order correlation coefficient is .30. An individual correlation of this size would not be

TABLE 8

Measures of central tendency for EL and JK correlations*

	List 1	List 2	List 3	List 4	OVERALL
Mean	.34 (.32)**	.32 (.29)	.39 (.35)	.30 (.26)	.32 (.30)
Median	.42 (.32)	.38 (.34)	.37 (.29)	.35 (.30)	.34 (.33)
Range	-.15 to .84 (-.18 to .84)	-.28 to .77 (-.37 to .85)	.05 to .74 (-.08 to .71)	-.32 to .66 (-.51 to .70)	.04 to .60 (-.08 to .62)

* The data are individual correlations for 24 subjects.

** The correlation coefficient outside the parentheses is the Pearson product-moment correlation coefficient; that within the parentheses is the Spearman rank-order correlation coefficient.

For an individual subject on a single list, significance at the .05 and .01 levels requires Pearson correlation coefficients of .50 and .62 or Spearman correlation coefficients of .54 and .68, respectively.

For an individual subject on all 64 items, significance at the .05 and .01 levels requires Pearson correlation coefficients of .25 and .32 or Spearman correlation coefficients of .36 and .47, respectively.

reliably different from zero, $p < .10$. The overall median values of both coefficients are slightly greater than the mean values, .34 and .33 for Pearson and Spearman correlation coefficients, respectively. There would seem to be a moderate relationship between individuals' EL and JK ratings.

Over lists, the trend for the mean Pearson correlation coefficient is a slight decrease (.34, .32, .39, .30). The decrease is seen as more monotonic when the listwise course is described by changes in the median values of the Pearson coefficients (.42, .38, .37, .35). Spearman values are similar. Neither of these measures indicates a substantial or significant change in the EL and JK relationship over lists.

However, it is possible that the wide range present among the correlations may have led to measures of central tendency which greatly obscure listwise differences between the correlations for individual subjects. Therefore, the subjects were divided into high and low groups based on the strength of their EL-JK correlations on the first list. The correlations then served as data in a 2 X 4 ANOVA in which Correlation Level (high or low) and Lists were factors. The Correlation Level X Lists interaction proved significant, $F(3, 66) = 5.88$, $p < .01$. The initially higher correlations decreased monotonically over lists (.59, .50, .43, .38), while the initially lower correlations tended to increase with practice (.09, .14, .35, .23). It is, of course, plausible that such changes are merely a regression artifact. However, the interaction did occur while the variance of individual subjects' JK ratings decreased reliably over lists, $F(3, 69) = 5.05$, $p < .01$. The tendency for all subjects to give higher JK ratings with practice, together with the trend for fast learners to utilize the higher JK scales dispropor-

tionately over lists, seem to have been causal factors.

Because the previous correlations for each list were based on only 16 items per subject, it is possible that some were disproportionately influenced by a few extreme scores. This same influence also appears to have been present in some scatterplots of even overall data. To minimize the role of extreme ratings in EL-JK comparisons, then, a new series of analyses was performed in which the EL scale was considered as a six-point, equal-interval scale. Specifically, it was assumed that subjects' EL ratings were dichotomous judgments and their adequacy as dichotomous predictors of JK ratings was evaluated. For example, an EL judgment of 4 was considered an accurate predictor of the item's later JK rating if the subject's JK rating for that item was 4, 5, or 6. However, if the item's JK rating was 3, 2, or 1, the EL judgment was considered an inaccurate predictor. The interpretations of such relationships are similar to those of EL-JK correlations.

Overall comparisons will be considered first. For the 64 EL predictions of JK ratings, the average subject was found to make 27.4 errors. EL ratings as dichotomously predictive of JK ratings for the same items, then, involve a mean error proportion of .43. When this mean error proportion was compared with a .50 probability of success-failure, the two proportions were found to be not reliably different, $z = 1.11$. When extreme values can have no disproportionate influence on the strength of the EL-JK relationship, then, a subject's dichotomous EL rating is not a reliably accurate predictor of the later JK for the same item--at least when the items are considered without regard to list.

The number of erroneous EL predictions of JK ratings was next considered on a listwise basis. The mean number of such EL-JK errors

was found to increase with practice (5.88, 7.12, 7.25, 7.17). However, this was not reliable, $F(3, 69) = 1.62$. The nonsignificant difference was thought to have resulted from the wide listwise variability within subjects in the number of erroneous EL-JK predictions.

To examine this variability directly, the subjects were next divided on the basis of their number of erroneous EL predictions of JK ratings on List 1. This variable may be labeled EL-JK Errors and has the values high and low. The listwise progress of this variable was analyzed by a 2 X 4 ANOVA. The EL-JK Errors X Lists interaction was found to be significant, $F(3, 66) = 5.88$, $p < .01$. The change over lists was mainly due to an increase in the number of EL-JK errors by subjects whose List 1 JK ratings had been rather accurately predicted by dichotomous EL ratings (3.58, 7.17, 6.50, 6.00 EL-JK errors over lists). Subjects whose early JK ratings were poorly predicted by their EL ratings tended to have EL-JK errors at the chance level on all four lists (8.17, 7.08, 8.08, 8.08). It thus appears that the easy-hard dichotomy reflected in EL judgments is never a dominant factor in some subjects' judgments of knowing. Similarly, the strong early similarity between EL judgments and JK ratings is soon diminished for the other subjects.

A potential objection to the above analyses regarding the dichotomous prediction of JK ratings by EL ratings might question whether an EL rating of 3 or 4 is correctly considered erroneous if it is followed by a JK rating of 4 or 3. However, on the first list where there is no influence from past experience, 55% of the EL values of 3 and 4 correctly predicted their respective JK ratings. This proportion is only slightly less than the 58% correct proportion for JK values of 3 or 4 when these predicted recall performance on the first list. It may also

be mentioned that, over the four lists, EL-JK errors of only one scale position accounted for only 19%, 22%, 18%, and 14% of the total number of EL-JK errors. In sum, then, this objection does not seem valid.

The data from the predictive accuracy of dichotomous EL judgments in respect to later JK ratings do not deny the previous finding of some relation between the EL and JK ratings of the same subjects. However, the dichotomous data indicate that the relationship is stronger in terms of the full EL and JK scales than in terms of the dichotomous scales. Stated differently, some items should seem to be more influential in the EL-JK relationship than others. Such items are not allowed to have a greater import when the scales are dichotomous. Further, when the influence of relatively extreme items is eliminated by usage of a dichotomous scale, the listwise course for some subjects would seem to be no EL-JK relationship while for others the relationship becomes weaker.

The EL-JK relationship was also investigated with data from group means. When the mean EL values for each item were correlated with Group EL's mean JK rating for the same item, the correlation was found to be .80. This is much higher than the mean individual overall EL-JK correlation (.32) and is even greater than the strongest individual overall correlation (.68). The above-mentioned strong influence of some items in the correlation would seem to be validated. Further, evidence is also offered for an idiosyncratic component in the individual EL-JK correlations. Such idiosyncraticity is likewise attested to by a drop in the EL-JK correlation (from .80 to .68) when the mean JK values are those of Group NoEL.

Predictions of recall performance by EL and JK ratings. A second EL and JK comparison involves the evaluation of their individual valid-

ities as predictors of recall performance. This concern is not with EL accuracy in itself. Recall or nonrecall after only one study trial may not be an adequate test of the accuracy of an EL prediction. However, if JK ratings were made on the same basis as that underlying the original EL ratings, the JK accuracy values ought to be similar to those for the EL ratings. Such comparisons are presented for both individual and group data. The same information also allows further comparisons of the effect of an EL task on later JK accuracy.

Each subject's EL and JK ratings were first compared as predictors of recall performance by individual biserial correlations. Table 9 contains measures of central tendency for the biserial correlation coefficients which were run for each Group EL subject on both the overall data and on the separate lists.

When the overall data are considered, two relationships are prominent. First, for Group EL, the mean value of the correlation between EL ratings and predictions of recall-nonrecall (.26) is weaker than the mean value of the biserial correlation between JK ratings and predictions of recall-nonrecall (.38). The difference is greater when the central tendency of the individual correlations is described by median values, .22 for the EL-recall performance and .44 for the JK-recall performance correlations. Secondly, the mean JK-recall performance correlation value (.47) for Group NoEL is greater than either of the two correlations mentioned for Group EL.

However, such overall data involve a JK which may be influenced by feedback as to its accuracy plus other treatment differences relative to the EL judgment. To compare actual JK accuracy when feedback is not present with the expected accuracy if the basis for the JK were that

TABLE 9

Measures of central tendency for biserial correlations
between EL and JK ratings and recall performance

Group EL: EL ratings and Recall-Nonrecall performance.

	List 1	List 2	List 3	List 4	OVERALL
Mean	.23*	.23	.25	.30	.26
Median	.30	.26	.21	.35	.22
Range	-.33 to .66	-.24 to .62	-.28 to .78	-.94 to .84	-.16 to .59

Group EL: JK ratings and Recall-Nonrecall performance.

	List 1	List 2	List 3	List 4	OVERALL
Mean	.36	.46	.52	.43	.38
Median	.40	.48	.58	.43	.44
Range	-.42 to 1.11**	.00 to 1.10	-.60 to 1.04	-.24 to 1.14	-.23 to .78

Group NoEL: JK ratings and Recall-Nonrecall performance.

	List 1	List 2	List 3	List 4	OVERALL
Mean	.59	.55	.50	.56	.47
Median	.64	.67	.55	.67	.51
Range	-.26 to 1.09	-.38 to 1.01	-.33 to 1.06	-.48 to 1.05	-.26 to .78

* For an individual subject on a single list, significance at the .05 and .01 levels requires biserial correlation coefficients of .50 and .62, respectively.

For an individual subject on all 64 items, significance at the .05 and .01 levels requires biserial coefficients of .25 and .32, respectively.

** The biserial correlation coefficient may be greater than 1.00.

which underlay EL ratings, List 1 performance may be observed via biserial correlations. For such a comparison, factors which may non-differentially facilitate the JK need not be distinguished from the JK. However, differences which may have had a negative influence on only EL ratings must be considered. Such differences may be that the EL ratings involved the entire list, that they were self-paced, and that their accuracy was assessed after a longer interval than that present for JK ratings.

However, the listwise experience can be expected to have led to more sophisticated scale usage by subjects making EL judgments. Similarly, time at the EL task was only slightly related to the number of errors with dichotomous EL predictions of recall performance ($r = .16$). This contrasts with the stronger and facilitative ($r = -.29$) effect of time at the EL task and number of JK errors. Lastly, since the EL judgment was meant to elicit relatively stable subjective estimates of item difficulty, a test of EL accuracy which occurs within minutes may be considered adequate.

When the accuracy of EL and JK ratings is assessed by mean biserial correlations on the first list, the same ordering is found as with overall correlations. Accuracy of Group EL at the EL task (.23) was lower than its accuracy at the JK task (.36), while both were very much inferior to the accuracy of Group NoEL at the JK task (.59).

Two conclusions seem to follow from these results. First, the criteria utilized for the JK task are better indicators of future performance than those used in the EL task. Whether such differences are in degree or in kind are beyond the present data. Secondly, the poorer JK performance of Group EL compared to Group NoEL would seem to indicate

again that previous experience with the same stimuli has a strong negative effect on the accuracy of later JK ratings for the same items.

Table 9 also indicates listwise performance by all subjects. It was expected that EL accuracy would remain constant over lists while JK accuracy improved with practice. However, the biserial correlation may not be an adequate assessor of JK accuracy on lists after the first for the present data because of its nature as a comparison of group means in term of a dichotomous characteristic. That is, a few large-valued erroneous predictions can influence the value of the correlation very much. Since all subjects, but especially Group NoEL, used extreme ratings more frequently with practice, this may explain the failure of Group NoEL's mean JK-recall performance correlations to increase with practice (.59, .55, .50, .56). Even under such a limitation, however, the mean JK-recall performance correlations for Group EL were found to increase with practice (.36, .46, .52, .43) while the EL-recall performance correlations were more constant (.23, .23, .25, .30).

To give a more adequate description of changes over lists in the differential accuracies of EL and JK ratings, therefore, both judgments were expressed as dichotomous predictors of recall performance and their accuracies determined. On each list, the average subject's dichotomous EL ratings were found to lead to 7.28 erroneous predictions of recall performance. This was reliably greater than the 5.17 JK errors found per list, $F(1, 23) = 28.45$, $p < .01$. Because the mean number of EL errors in predicting recall performance was almost constant over lists (7.17, 7.79, 7.08, 7.08) while the mean number of JK errors decreased monotonically with practice (6.04, 5.37, 4.88, 4.38), the nonsignificant ($F = 1.33$) interaction of Error Type X Lists was unexpected. However,

individual t-tests performed for the four lists indicated that differences in the number of erroneous predictions by the two measures were reliable on the last three lists, t's (23, two-tailed) > 2.60, p < .02. The failure to find a significant interaction, then, would seem to have been highly influenced by a large degree of within-subject variability in EL accuracy over lists. For the same items, then, an individual subject's EL ratings proved much less accurate as predictors of recall performance than his JK ratings. The differences may be said to have been generally greater after the first list.

Predictions of overall recall performance by EL and JK values were also computed on group data. The mean EL value for each item was found to correlate .49, p < .001, with the number of times the PA item was recalled by Group EL. On a groupwise basis, therefore, EL ratings are rather accurate predictors of recall performance. The same prediction by the mean JK values of Group EL, however, was somewhat stronger at .64. In contrast to the results with the individual biserial correlations, this JK-recall value by Group EL was similar to the .66 correlation by Group NoEL.

Rather surprisingly, the mean EL rating for an item was also a good predictor for Group NoEL recall, $r = .53$. Similarly, the mean JK ratings by Group EL correlated .63 with Group NoEL recall while the mean JK ratings by Group NoEL correlated .57 with Group EL recall.

Because group predictions were generally superior to those by individuals, it was decided to evaluate the present PA data in terms of the finding of Underwood (1966) that group predictions of individuals' free recall performance were superior to individuals' predictions of their own future free recall performance. Thus, group EL and JK values

were used to predict individual recall-nonrecall via biserial correlations. When mean EL ratings predicted each subject's recall performance, the mean correlation was found to be .25 for Group EL. This value is rather similar to the mean value of .26 when subjects' own EL ratings predicted their performance. When the mean EL ratings were used to predict recall by Group NoEL, the mean correlation was similar, $r = .27$.

When the mean JK values were used to predict recall performance for the individual subjects, the mean biserial correlation was .32 for Group EL's predictions for its individual members and .34 for Group EL's predictions of the other group's individual performances. These are only slightly less than the mean biserial correlation (.38) for each Group EL subject's JK ratings relative to actual recall or nonrecall.

However, when the mean JK values for Group NoEL predicted individuals' performances, the mean biserial correlation was .36 in regard to Group NoEL and .26 in regard to Group EL. The mean value of .36 for Group NoEL is in contrast to a .47 value obtained when the mean biserial correlations resulted from individuals' predictions of their own performance. This difference within subjects is significant, $t_{(23, \text{two-tailed})} = 2.24$, $p < .05$. To some extent, then, the JK ratings of individual subjects in Group NoEL can predict the subjects' performance better than can those of the group.

Secondary Analysis: EL and JK congruence and JK accuracy. The relationship between EL and JK ratings may also be studied by examining the effect of EL-JK similarity and dissimilarity on the validity of the JK. The concern in such an examination is whether two similar judgments about an item (predictive congruence) have different predictive validities than two opposite judgments about the same item (predictive incon-

gruence).

As in previous analyses, the comparisons will be on a dichotomous basis. Thus, if an item with a JK of 1 was previously given an EL rating of 1, 2, or 3, this is scored as predictive congruence. However, if a JK of 1 was originally given an EL rating of 4, 5, or 6, this is viewed as predictive incongruence.

The average subject's probability of a correct JK prediction for predictively congruent items was found to be .69. This was not significantly greater than the .66 probability found for predictively incongruent items, t (23, two-tailed) = .59. Likewise, for subjects who had entries in both categories, congruent and incongruent predictions were both more accurate regarding recall than nonrecall. For predictively congruent items, the average subject was correct on 79% of the recall and on 50% of the nonrecall predictions, t (21, two-tailed) = 2.88, $p < .01$. The 67% versus 46% difference for the incongruent predictions was in the same direction, but this failed to be statistically reliable, t (22, two-tailed) = 1.89, $.10 > p > .05$. Thus, when subjects gave two similar ratings to the same items, they were as likely to predict correctly as when they reversed their original assessment.

The examination of the progression over lists of JK-EL congruence or incongruence by similar intra-subject comparisons would have involved proportions based on very different numbers of instances, plus subject loss due to a failure to have both types of items on all lists. Instead, listwise data was examined collapsed across subjects. Table 10 presents this comparison of JK accuracy as a function of predictive congruence and incongruence over lists.

TABLE 10

Predictive congruence-incongruence and
probabilities of correct JK predictions

Predictive Congruence (JK = EL)

List 1	List 2	List 3	List 4	OVERALL	JK=EL
--------	--------	--------	--------	---------	-------

P(Correct Prediction re Recall):

$\frac{88}{106} = .83^*$	$\frac{93}{119} = .78^*$	$\frac{105}{128} = .82^*$	$\frac{99}{120} = .82^*$	$\frac{385}{473} = .81^*$	
--------------------------	--------------------------	---------------------------	--------------------------	---------------------------	--

$\frac{611}{880} = .69$

P(Correct Prediction re Nonrecall):

$\frac{66}{139} = .48$	$\frac{49}{92} = .53$	$\frac{48}{81} = .59$	$\frac{63}{95} = .66$	$\frac{226}{407} = .56$	
------------------------	-----------------------	-----------------------	-----------------------	-------------------------	--

Predictive Incongruence (JK \neq EL)

List 1	List 2	List 3	List 4	OVERALL	JK \neq EL
--------	--------	--------	--------	---------	--------------

P(Correct Prediction re Recall):

$\frac{49}{63} = .78^*$	$\frac{82}{120} = .68$	$\frac{84}{131} = .64$	$\frac{88}{122} = .72^*$	$\frac{303}{436} = .70$	
-------------------------	------------------------	------------------------	--------------------------	-------------------------	--

$\frac{431}{656} = .66$

P(Correct Prediction re Nonrecall):

$\frac{36}{76} = .47$	$\frac{33}{53} = .62$	$\frac{30}{44} = .68$	$\frac{29}{47} = .62$	$\frac{128}{220} = .58$	
-----------------------	-----------------------	-----------------------	-----------------------	-------------------------	--

* $p < .05$.

In general, similarities in JK accuracy for congruent and incongruent JK predictions seem more prominent than dissimilarities. Indeed, the relation is almost one of proportional identity on the first list. Likewise, with one exception, both types of predictions lead to greater JK accuracy regarding recall than nonrecall, although the relationship is generally stronger for congruent predictions. Over lists, the probability of correct predictions of nonrecall increases for both types of predictions, with the improvement more monotonic for congruent predictions. The most striking difference between results for the two types of predictions is the greater accuracy regarding recall for congruent than for incongruent predictions and the greater accuracy regarding nonrecall for incongruent items.

In general, then, similarity or dissimilarity of the JK to the original EL judgment has little differential effect on JK validity. The two sets of data appear almost as random samples drawn from a larger population. On a dichotomous basis, then, similarity between the previous EL and the later JK may be more epiphenomenal than reflective of related underlying bases for the two judgments.

Secondary Analysis: Stimulus characteristics as predictive of the JK. The EL-JK relationship has been examined rather extensively so that some sense of direction may be obtained regarding the relationship between the JK and stimulus knowledge as expressed by EL ratings. A different approach to the role of stimulus knowledge in JK ratings is the consideration of the adequacy of actual stimulus characteristics as predictors of JK ratings.

If stimulus characteristics are related to recall, they can be expected to provide accurate bases for a JK. However, because of the

same relationship between the stimulus characteristic and recall, a correct prediction can also be expected under the hypothesis of memory knowledge.

The present experiment attempts to minimize such interpretative difficulties by focusing not on only a single correlation but on two potential manifestations of the relationship. In effect, two samples will be drawn from the set of each subject's judgments of knowing--items whose JK ratings were correct predictions and items whose JK ratings were incorrect. For both correct and incorrect predictions, specific expectations are possible for different item types (Hits, Correct Rejections, Misses, False Alarms) at different levels of the stimulus characteristic.

The general procedure will be first to establish that the stimulus characteristic has a role in recall performance and that subjects were to some extent aware of this potential relationship in their EL ratings. Secondly, each subject's predictions will be described by the proportion $\text{Hit} - \text{Correct Rejection} / \text{Hit} + \text{Correct Rejection}$, while the incorrect predictions will be represented by the proportion $\text{Miss} - \text{False Alarm} / \text{Miss} + \text{False Alarm}$. The terms of the proportions will be the number of items the individual subject has placed in the four categories at the relevant levels of the stimulus characteristic. The use of proportions is an attempt to equate the number of items in the correct and incorrect categories. The directionality of the terms in the proportions is meant to allow the expectation of an interaction if JK ratings are given mainly as a function of the stimulus characteristic.

For example, if JK ratings are theorized to be given as a function of response term imagery, it may be expected that low imagery items will

receive low JK values. If such predictions are correct, the low imagery items can be expected to be scored as Correct Rejections; if they are incorrect, as Misses. The same reasoning leads to the expectation that high imagery items will be given high JK ratings. If correct, such predictions would be classified as Hits; if incorrect, as False Alarms. When the terms for the correct predictions are combined in the Hit - Correct Rejection/Hit + Correct Rejection proportion, it must therefore be expected that the proportion will be lower in value for low imagery items (when Correct Rejections predominate) than for high imagery items (when Hits are more prominent).

In the Miss - False Alarm/Miss + False Alarm proportion, it may similarly be expected that the proportion will be higher in value for low imagery items (when Misses are greater in number) than for high imagery items (when False Alarms are in the majority). Since the proportions for correct predictions are anticipated to increase with imagery while the proportions for incorrect predictions are expected to decrease, an interaction can be expected if JK ratings are given on the basis of response term imagery.

Imagery as predictive of the JK. The first stimulus characteristic to be considered is that of response term imagery. To validate the relationship between response term imagery and recall, each subject's number of correct recalls for each of the four levels of imagery per list was determined and became the dependent variable in a 2 X 4 X 4 ANOVA in which EL judgment (presence or absence), Lists, and Response Term Imagery Level were factors. The only significant effect was that of Response Term Imagery Level, $F(3, 138) = 14.44$, $p < .01$. As response term imagery increased, the mean number of recalls for that level also

increased (2.28, 2.50, 2.78, 2.89 out of a maximum of four recalls.). When the data were considered collapsed across subjects, the correlation between an item's response term imagery and the number of subjects who recalled the item was .38.

To determine if Group EL subjects were aware of the potential relationship between response term imagery and recall, the mean ratings on the six-point EL scale were calculated for each of the four levels of imagery for each subject. The EL values (3.08, 3.28, 3.48, 3.75) paralleled the increase in response term imagery, $F(3, 69) = 14.77$, $p < .01$. The same relationship between mean EL values and response term imagery was .43.

In sum, response term imagery seems to have had a role in recall performance and subjects were somewhat aware of this potential relationship in their EL judgments.

The second part of this analysis involved the comparison of the proportions for correct and incorrect predictions of recall performance. The ideal analysis would have included eight ratios for each of four lists per subject. However, only two subjects had entries in all 32 categories. Instead, to maximize the number of items in the proportions for each imagery level, the data were collapsed over lists and proportions were formed on the basis of only two levels of imagery. Each subject thus contributed four proportions, a correct and an incorrect prediction proportion for each imagery level. The proportions for incorrect predictions of low imagery paired-associates were thus based on 10.3 items per subject and those of high imagery paired-associates on 7.9 items. A 2 X 2 X 2 ANOVA was performed on the factors EL judgments (presence or absence), Proportion type (proportion for correct predic-

tions or for incorrect predictions), and Imagery level. Of the interactions, only that between Proportion type and Imagery level was reliable, $F(1, 46) = 9.94, p < .01$. However, the interaction was not of the mirror-image type predicted. Instead, while the mean proportions for correct predictions rose with imagery level, from .23 to .50, the change for the mean proportions for incorrect predictions was minimal, from .15 to .18 over the two levels of imagery. Similar but not significant results were found by a 2 X 4 ANOVA which examined the two prediction ratios over four levels of imagery without regard to lists.

However, it is plausible that incorrect predictions were incorrect precisely because of a difficulty in determining the level of the stimulus characteristic. The generally small differences between Misses and False Alarms for all levels of response term imagery are not incompatible with such an interpretation. One should therefore expect the greatest difficulty in labeling to occur for the two moderate levels of imagery so that performance on such items should be near chance for both correct and incorrect predictions. Yet, the trend for values of both types of proportions was to increase with level of imagery-- .14, .25, .38, .49 for correct predictions and .11, .19, .22, .19 for incorrect predictions. Labeling difficulties do not seem an adequate explanation for the lack of similar JK values for items of similar imagery level.

Following the previously explained reasoning, the data from the correct predictions may thus be taken as support for the hypothesis that the JK was based on the stimulus characteristic of response term imagery. However, the failure to find comparable results for incorrect predictions casts doubt on the generality of this hypothesis.

Learnability as predictive of the JK. Because of such uncertainty, a second variable was examined as a stimulus characteristic. This was labeled "learnability" and was defined as the number of times an item was recalled. It may be taken to refer to all those characteristics of a PA which predict its future recall or nonrecall. Its correlation with response term imagery was .38. Although the definition of this variable is subjective, the rather strong correlation of .73 between recall for the two groups in this experiment suggests generalizability. More importantly, the correlation of .71 between number of recalls and mean JK rating indicates a very strong relationship. The expectation of a similarly strong relationship between mean JK and both correct and incorrect predictions is thereby even more plausible than could be expected from the .46 correlation between mean JK and response term imagery.

Four levels of the learnability dimension were determined for each list. This determination of learnability levels created an ordering of item difficulty within-lists which was slightly different from the ordering of the entire 64 item list. However, the four product-moment correlations between an item's within-list position in terms of number of recalls and the item's position in the overall list in terms of number of recalls were all very strong (from .98 to .998 for the four lists). Thus, little is lost when the data are considered without regard to list. More importantly, the lists may be said to have been counterbalanced with respect to learnability.

The four levels of the learnability dimension were analyzed in the same manner as response-term imagery had been. As with imagery, the increase in EL ratings as a function of the learnability (3.06, 3.37, 3.39, 3.76) was reliable, $F(3, 69) = 16.48, p < .01$. For group data,

the correlation between mean EL ratings and the learnability dimension was .55.

As was true with response-term imagery, potentially small proportions limited the nature of the analyses which could be performed. The data were therefore collapsed over lists to form correct and incorrect proportions for two levels of the learnability dimension. The ratios were analyzed via a 2 X 2 X 2 ANOVA in which the factors were EL judgment (presence or absence), Proportion type (proportion for a correct prediction or for an incorrect prediction), and Learnability levels. The dependent measure was the value of the proportion. The only significant interaction was that of Proportion type X Learnability levels, $F(1, 46) = 5.37, p < .05$. However, as in the previous analysis, the interaction was not that predicted by a hypothesis of JK ratings based on learnability. Instead, the difference in bias ratios for the hard-to-learn and easy-to-learn items was greater for the correct predictions (.10 versus .57) than for the incorrect predictions (.06 versus .33).

The data were also analyzed by a 2 X 4 ANOVA in which Proportion type was examined over the four levels of the learnability dimension. The interaction was again significant, $F(3, 135) = 4.49, p < .01$. The monotonic increase over the learnability levels for correct predictions (-.09, .27, .42, .67) was greater than a similar increase for incorrect predictions (.00, .14, .30, .37).

Thus, for easy-to-learn items, subjects' correct predictions were more likely to be Hits than Correct Rejections while the differences were less clear among hard-to-learn items. This may be taken as support for a stimulus knowledge basis for the JK. However, incorrect JK predictions reversed the trend regarding easy-to-learn items. JK errors

for easy-to-learn items were more apt to be Misses than False Alarms. Thus, the hypothesis of a stimulus knowledge basis for JK ratings must explain the somewhat paradoxical situation in which subjects sometimes gave easy-to-learn items a high JK and at other times gave items of the same class a low JK.

Summary. The second question has investigated the relationship between EL judgments and JK ratings. A moderate correlation was found between subjects' EL and JK ratings for the same items. It may also be mentioned that, on a dichotomous basis, the EL judgment was found to be a poor predictor of later JK ratings for some subjects on all four lists and a decreasingly accurate predictor for subjects for whom the initial predictions were very accurate. Also, as a predictor of actual recall performance, the EL judgment was less adequate than the JK, even when the two predictors were compared on the first list where no possibility of feedback was present. The negative effects of the EL task were also evident in such comparisons.

Secondary analyses were also performed. The predictive accuracy of JK ratings proved extremely similar when they were both congruent and incongruent with earlier EL ratings. Another secondary analysis evaluated the stimulus characteristics of response term imagery and learnability as predictors of JK ratings. Both dimensions were found monotonically related to JK ratings when the predictions were later correct. However, the same monotonic relationships were not present with items whose JK ratings were later seen to be incorrect.

Question 3. Feeling of knowing judgments, scholastic achievement, and the Judgment of Knowing

The third question seeks first to determine if there is a relationship between subjects' capabilities at feeling of knowing and judgment of knowing tasks. Secondly, the question examines various educational achievement measures in their relationship to the JK. If success at the JK involves knowledge of the contents of one's memory, some relationship may be expected between the JK and the FK judgment. Similarly, if scholastic achievement requires judgments of knowing, proficiency at the two tasks ought to correlate.

The feeling of knowing judgment. Prior to a consideration of its relationship with the JK task, the FK task will be examined in its own right. Of the 52 general-information questions, the average subject recalled answers to only 7.85 or 15% of the items. This contrasts with the 40% recall by subjects in the more difficult of the two Hart (1965) general-information experiments. However, the present subjects expressed a positive FK judgment for 25.8 or 59% of the nonrecalled items. These proportions are similar to the 38% and 62% for the respective judgments in the Hart experiment. Comparisons between the present experiment and the Hart research seem appropriate.

Table 11 presents the proportions of correct recognitions for the six FK scale positions pooled across subjects. The upper proportions evaluated the accuracy of the first multiple-choice selections. When the data are considered for all subjects, the increase in the probability of correct recognition parallels the increase in the FK scale (.27, .31, .35, .43, .46, .61). When the individual probabilities of correct recognition for each scale position are evaluated against

TABLE 11

Proportions of correct recognitions for the six FK scale positions

FK*

1	2	3	4	5	6
$\frac{119}{442}$	$\frac{121}{396}$	$\frac{142}{402}$	$\frac{211}{491}$	$\frac{110}{241}$	$\frac{89}{147}$
.27	.31	.35	.43	.46	.61
		$\frac{382}{1240}$	$\frac{410}{879}$		
		.31	.47		

FK2**

1	2	3	4	5	6
$\frac{229}{442}$	$\frac{237}{396}$	$\frac{248}{402}$	$\frac{342}{491}$	$\frac{163}{241}$	$\frac{115}{147}$
.52	.60	.62	.70	.68	.78
		$\frac{714}{1240}$	$\frac{620}{879}$		
		.58	.71		

* Only the first multiple-choice answer is scored in the evaluation of FK accuracy.

** Both multiple-choice answers are scored in the evaluation of FK accuracy.

the chance probability of .25, the three lower values are all not significantly different from the chance probability (z 's = .35, .79, 1.62 for ascending FK values). Similarly, when the pooled data are considered dichotomously, the .31 probability of a Miss is not reliably greater than the chance probability of .25 (z = .96). This is dissimilar to the findings of Hart (1965), Freedman and Landauer (1966), and DaPolito, Guttenplan, and Steinitz (1968), all of whom reported rather high correct recognition probabilities for items given negative FK ratings (Misses). This result may be taken as a validation of Hart's (1967a) suggestion that correct predictions of nonrecognition (Correct Rejections) will occur at chance levels when item difficulty is high.

In contrast to the results with lower FK ratings, the values for the three upper JK scale positions are all significantly greater than chance (z 's = 2.72, 3.12, 4.75). Likewise, the .47 probability of a correct dichotomous FK prediction (Hit) is significantly greater than the .25 chance expectation (z = 3.12). Although such results indicate above chance recognition probabilities of items subjects feel they know, the performance is far from perfect.

The FK recognition probabilities were next considered dichotomously on a within-subject basis. Because the mean number of items given positive FK ratings was rather similar to that for the number given negative FK ratings, the Hart accuracy technique is more appropriate with the FK data than it was for the JK data. Thus, if the probability of a correct recognition given a positive FK judgment (Hit) is greater than the probability of a correct recognition given a negative FK judgment (Miss), the subject can be theorized to have some facility in knowing the contents of his or her memory. The terminology from signal-

detection theory will be utilized in the same manner as it was for the JK.

For the average subject, the probability of a Hit (.48) was reliably greater than the probability of a Miss (.31), $t(47, \text{two-tailed}) = 8.15$, $p < .001$. When the positive FK judgments were considered separately, the probability of a Hit (.48) was found to be not significantly different from the probability of a False Alarm (.52), $t < 1$. This was anticipated since Hart (1967a) has theorized a tendency for False Alarms to be rather high on a difficult FK task. For negative FK judgments, however, the probability of a Correct Rejection (.70) was reliably higher than the probability of a Miss (.30), $t(47, \text{two-tailed}) = 6.32$, $p < .001$. For the present data, then, subjects were somewhat adept at predicting their future recognition choices. However, subjects were more accurate at indicating a lack of knowledge than its presence.

The subjects in the present experiment were also asked to select a second answer for each FK question. The second recognition choices by themselves are relatively unimportant. However, correct recognitions on second choices can be combined with correct recognitions of the first choices to provide an evaluation of overall FK accuracy which compensates somewhat for item difficulty.

The lower half of Table 11 contains correct recognition proportions for the six FK scale positions when both first and second choices are included. The variable may be labeled FK2. The recognition probabilities for FK ratings of 1, 2, 3 were all not significantly greater than the chance probability of .50 (z 's = .29, 1.58, 1.92) while those for FK ratings of 4, 5, 6 were all significantly greater than chance (z 's = 3.04, 2.73, 3.97). As with first choices, the mean individual

Hit probability (.70) was found to be reliably greater than the mean Miss probability (.58), t (47, two-tailed) = 5.00, p < .001. Unlike the results with the first choices, though, the comparison of the Hit (.70) and False Alarm (.30) mean proportions was highly reliable, t (47, two-tailed) = 13.53, p < .001. For negative predictions, however, the probability of a Miss (.58) was found to be reliably greater than that of a Correct Rejection (.42), t (47, two-tailed) = 5.50, p < .001. This was not unexpected because the instructions had asked subjects to relate their FK judgment to the probability of a correct selection on the first choice. Thus, a subject who gave a FK of 3 or 2 might be expected to have given a higher rating if he believed that the FK judgment was to be tested by two choices. Thus, when the object of concern is the accuracy of both first and second choices, the present research will consider only measures which involve the positive FK2 predictions.

. The feeling of knowing judgment and the Judgment of Knowing. Two kinds of measures are used to compare FK and JK performance: overall measures which combine positive and negative predictions and measures which consider positive and negative predictions separately. Because the appropriate descriptive statistic for comparing FK and JK performance is not obvious, three measures are used: FK Correct Predictions, FK Game-points, and the Hart accuracy measure (FK Accuracy).

The FK Correct Prediction measures considered the FK predictions as dichotomous. The FK Correct Prediction value was determined by subtracting the number of incorrect FK predictions from the number of correct FK predictions. To allow the separate consideration of negative FK predictions, the subject's number of Correct Rejections and Misses were subtracted to form a measure labeled FK Negative Accuracy. Simi-

larly, Hits and False Alarms were subtracted to form a measure allowing the comparison of positive FK predictions, FK Positive Accuracy. Although the FK Correct Prediction score involved the difference between right and wrong predictions, it can be legitimately compared with the JK Errors value because the latter was based on a constant number of 64 items. Comparisons between the JK and FK Positive and Negative Accuracy measures are appropriate because all were formed similarly. Lastly, a measure labeled FK2 Positive Accuracy was formed by the subtraction of right and wrong positive FK predictions when the correctness of both the first and second choices was evaluated.

The FK Game-points values were formed by the same scoring system as had previously been used for JK Game-points values. Every Correct Prediction measure was duplicated by a Game-points analysis.

The FK Accuracy measure introduced by Hart was also compared with JK values. Because this measure equates the number of positive and negative predictions, it has advantages over the other two measures. FK Accuracy was found to correlate .77 with FK Correct Predictions and .60 with FK Game-points. FK Correct Predictions and FK Game-points were more strongly related to each other (.86).

Prior to the actual comparisons of FK and JK scores, Groups EL and NoEL were compared on the various FK measures. No significant differences were found. Though the correlations later will be considered separately for the two groups, the present correlations are all based on 48 subjects. For this sample size, correlations of .29 and .37 are required for significance at the .05 and .01 levels, respectively.

Table 12 contains Pearson product-moment correlation coefficients for overall FK and JK ratings. The correlations consistently indicate

TABLE 12

Pearson product-moment correlations for overall JK and FK measures

	FK CP*	FK GP	FK Acc.	FK Neg. Acc.	FKGP Neg. Acc.	FK Pos. Acc.	FKGP Pos. Acc.	FK2 Pos. Acc.	FK2GP Pos. Acc.
JK Errors	-.38	-.30	-.31	--	--	--	--	--	--
JK GP	.31	.25	.26	--	--	--	--	--	--
JK Neg. Acc.	--	--	--	.32	--	--	--	--	--
JK GP Neg. Acc.	--	--	--	--	.20	--	--	--	--
JK Pos. Acc.	--	--	--	--	--	.33	--	.46	--
JK GP Pos. Acc.	--	--	--	--	--	--	.30	--	.37

* FK CP = FK Correct Predictions

FK GP = FK Game-points

FK Acc. = FK Accuracy

FK Neg. Acc. = FK Negative Accuracy

FK Pos. Acc. = FK Positive Accuracy

etc.

a moderate relationship between FK and JK scores. The correlation between FK Correct Predictions and JK Errors was $-.38$; that between FK Accuracy and JK Errors was slightly less at $-.31$. The overall correlations based on game-points measures are slightly smaller but still offer complementary support for some similarity in capabilities at the two judgments.

When the predictions of recall and nonrecall are compared separately for the two tasks, the moderate relationship is still present, with the game-points measures correlating lower than the correct prediction measures.

The last two columns of Table 12 contain correlations which included both the first and second recognition choices in the computation of the accuracy of positive FK predictions. The correlation of $.46$, $p < .001$, between the FK2 Positive Accuracy and JK Positive Accuracy strengthens the concept of a relationship between the FK and JK tasks.

Correlations were also run separately for the EL and NoEL groups. Two observations may be made. First, the general size of the correlations for the two groups did not differ greatly from those for the overall correlations. Second, overall correlations for Group EL were stronger than those for Group NoEL. For example, the FK Correct Predictions and JK Errors correlation for Group EL was $-.52$ while that for Group NoEL was $-.30$. Similar results in forthcoming analyses indicate that this difference was due to the presence of more extreme scores in Group EL than in Group NoEL.

Secondary Analysis: Listwise comparisons between JK and FK accuracy. A secondary series of comparisons involved overall FK scores and JK performance on the four lists. Performance on the FK task was not partitioned into halves or quarters because of the differing number

of recalls across subjects and because of the different locations of the recalled items. Although the relatively small number of JK errors in any one list casts doubt on potential generalizability, the comparisons seem worthwhile because any theoretical explanation for FK and JK similarities would seem to have to expect a closer relationship with practice. For example, if the basis of both judgments is some capability at cue reading, a closer relationship ought to be present between the two judgments over lists as capability at cue utilization increases in the JK task.

The previously mentioned FK overall measures (cf. Table 12) were correlated with their listwise JK complements. The listwise course of the correlations between FK Accuracy and JK Errors on the four lists are typical: .06, -.11, -.47, -.11. Thus, the measures consistently indicated a low relationship between the FK and JK values on the first two lists. For almost all comparisons, this relationship increased very much on the third list. The consistent drop on the fourth list was unexpected because JK measures showed an increase on this list. Because the JK variance on the third and fourth lists was not reliably different ($t = .16$), the lack of a relationship between the two tasks on the fourth list may indicate that subjects did not make judgments of knowing on the same basis as on the third list.

When the negative FK predictions are considered separately, the change over lists for the correlations between negative FK and JK accuracy measures was rather small. This would again seem to reflect the greater ease of a negative FK judgment compared to a negative JK. The correlations for positive FK and JK measures, however, mirrored the overall relationship. The correlations increased on the first three

lists and decreased on the fourth list.

It would seem that there is some increase in the FK and JK relationship over JK lists. However, if this is accepted, it would seem to follow that performance on the fourth JK list had a basis which is different from that on the previous lists.

Secondary Analysis: FK accuracy in comparison with EL and JK accuracy. A final examination of the FK task is a tripartite comparison which relates FK values to both the EL and JK measures. Previous findings have indicated a moderate relationship between EL and JK tasks. Should there also be a strong relationship between FK and EL measures, then some task commonality would seem to obtain between the three judgments.

Each subject's FK Accuracy measure was first correlated with both of the previously mentioned biserial correlations between EL and JK ratings and recall performance. The two biserial correlations themselves were related via a correlation of .46. The correlation between FK Accuracy and EL accuracy was .12; that between FK Accuracy and JK accuracy was .39.

Likewise, when the FK Accuracy measure was correlated with the number of erroneous dichotomous predictions of recall by EL ratings and by JK ratings, the correlation between FK Accuracy and EL errors was .02, while that between FK Accuracy and JK errors was .39. The two dichotomous measures correlated .28 with each other. The same differential relationships were also found when the FK Correct Predictions and the FK Game-points measures were correlated with the respective JK and EL measures.

The commonality between the JK and the other two predictive tasks

was anticipated by previous findings. However, the present data indicate that the FK and EL tasks are not related. It is as if the JK were similar to the EL judgment on one basis and similar to the FK on another basis. Lastly, it may be noticed that the small relationship between individuals' EL and JK capability is possibly associated with subjects' relative inexperience at the EL task. Stronger within-subject correlations may be present with more EL judgment practice.

Overall scholastic achievement and the Judgment of Knowing. The second variable examined because of its theorized relationship with the JK was educational achievement. One subdivision was labeled overall educational achievement. This included Scholastic Aptitude Test (SAT Verbal and SAT Math) scores, American College Testing (ACT) program scores, and subjects' high-school ranks (HS Rank). A second subdivision will be considered below. As the correlation **increases** between an educational achievement variable and measures of the JK, the tasks can be hypothesized to require similar capabilities.

Of the 48 subjects, 30 had SAT scores; 37 had ACT scores, and 34 had HS Ranks available. Only one subject had no score available. Although the educational achievement scores were above average, the ranges were as broad as might be expected from first year college students.

To validate that subjects with educational achievement scores available were representative of the entire sample in JK and FK performance, t-tests were performed in which the independent variables were presence or absence of the educational achievement score in question and the dependent variables were scores on the particular JK and FK tasks which were later to be correlated. The comparisons led to two

results. First, the JK and FK performance scores of subjects who did not contribute scholastic achievement scores were consistently inferior to those of subjects whose educational achievement scores were available. Secondly, the differences were significant only in three listwise comparisons which differentiated subjects on the basis of presence or absence of HS Rank, t 's < 2.25 . For correlational comparisons, however, one would have hoped that more subjects with low FK and JK scores had SAT and HS Rank values available.

Because it was not evident which JK and FK measures were optimal for comparisons with the overall educational achievement scores, the same measures which had previously examined JK and FK similarities were correlated with the four educational achievement measures. In general, the correlations were very low. Of the overall JK correlations, the strongest relationship was that between JK Errors and HS Rank, $r = -.20$. Of the overall FK correlations, the most strong relationship with the educational achievement measures was had by FK Accuracy. This measure correlated .23 with SAT Verbal, .25 with SAT Math, and .23 with HS Rank. Of all the correlations, the strongest was the .32 correlation found between the FK2 Positive Accuracy measure and HS Rank.

Separate correlations were also performed for Groups EL and NoEL where sufficient data were present. The only finding of note was the consistently strong relationship between HS Rank and JK performance for Group EL. For example, the correlation between HS Rank and JK errors was $-.43$ for Group EL but $-.05$ for Group NoEL. The greater variability present among Group EL subjects would seem to be a sufficient explanation for such differences.

College course achievement and the Judgment of Knowing. The

second educational achievement measure examined was performance in an introductory psychology course. The subjects came from three course sections. Two of the three classes were team-taught so that both received very similar instructions and identical or equivalent multiple-choice tests. To make scores comparable, values of the relevant variables were calculated as z-scores.

Four variables were examined in relation to JK and FK performance. Individual test scores involved the subject's lowest multiple-choice test score (LoTest), the mean of the scores on the two higher tests (MidTest), and the subject's highest test score (HiTest). Two subjects had only three tests and hence their MidTest scores were based on only one score. The relative standings of subjects on the three tests are rather similar. Product-moment correlations between subjects' course grades and LoTest, MidTest, and HiTest scores were .80, .83, .80, respectively.

All the final grades (A to F) in each class were scored on a 1 to 5 basis and put into z-score form to create the variable CourseGrade. Because mean performance was characterized by a z-score of .22 for the variable CourseGrade, the subjects seem rather representative of students in the introductory psychology courses.

Product-moment correlation coefficients were computed for the four college educational achievement variables paired with the same JK and FK performance variables which have been previously considered.

The first observation of importance is that LoTest scores were consistently more strongly related to JK and FK performance than were either the MidTest or the HiTest scores. The correlation between LoTest and JK Errors is $-.37$; that with List 3 JK Errors is $-.41$; that with

FK Accuracy is .35, and that with FK Positive Accuracy is .40. The strength of these LoTest relationships with JK and FK performance measures compares somewhat favorably to such predictors as SAT Verbal scores, ACT comprehensive scores, and HS Rank, which correlated .59, .30, .38, respectively, with LoTest scores.

The causality for the close relationship between JK and FK performance and the LoTest variable is not obvious. However, it may be noticed that the moderate relationship between JK performance and scholastic achievement as defined by LoTest scores is present under academic conditions in which it may be theorized that subjects have studied least adequately in relation to the study required for test success. This leads to the possibility that the JK and scholastic achievement relationship is related to level of learning. That is, even if a relationship between scholastic performance and JK accuracy exists, it is plausible that highly motivated college students of low JK ability are aware of their limitations in estimating their knowledge states and hence tend to overlearn material. If this is true, the relationship between poor capability at the JK and test performance will not be manifest when time for study is not controlled.

The second point of interest regarding the variables based on introductory psychology course performance is that CourseGrade was more weakly correlated with JK and FK measures than the LoTest variable had been. Still, there is a low relationship present between CourseGrade and the total number of JK errors (-.23), List 3 JK Errors (-.26), and FK accuracy (.27).

Because one of the three classes was treated differently from the other two by having a different teacher and different tests, the

previous analyses were repeated for only the two groups who had identical or equivalent tests. The resultant correlations were only slightly weaker than those for all subjects. The slight decrease would seem to have resulted from selective subject loss.

The correlations were also performed for Group EL and NoEL subjects separately. When the JK task and college educational achievement measures were compared, the correlations for the two groups were found rather similar in strength. However, the FK task and college educational achievement correlations were much greater for Group EL than for Group NoEL. For example, the FK Accuracy correlation with CourseGrade was .68 for Group EL but .01 for Group NoEL.

Summary. In general, the relations of both JK and FK tasks to overall educational achievement measures such as SAT and ACT scores were found to be positive but rather low. Limitations on variability in the correlates of the overall educational achievement measures may have limited the size of the correlations. However, it is also plausible that SAT and ACT scores measure more general factors than are required for accurate JK or FK ratings.

The relationships between the JK and FK variables with the introductory psychology course scores, in contrast, indicated a moderate relationship between scores on each subject's lowest test and both the number of JK errors and FK Accuracy. That such a relationship was found with data from relatively uncontrolled measures of scholastic achievement leads to the possibility of a stronger relationship in a more temporally regulated situation.

Question 4. Learning ability and the Judgment of Knowing

The effect of subjects' learning ability on the JK is studied so that conclusions may be reached regarding the role of process knowledge in the JK. That is, if improvement over lists is limited to very fast and slow learners, process knowledge would seem to be predominant in such improvement. Contrariwise, if improvement is possible without regard to learning-ability, then stimulus knowledge and/or memory knowledge would seem potential bases for the JK.

Learning ability was originally defined as the number of correct answers on five trials of the PA letter-number task. However, the correlation between this definition and the number of recalls on the four JK word lists was only .34, $p < .02$. Task identity is not strongly supported by such a correlation. Likewise, when subjects were divided into slow and fast learners on the basis of letter-number performance, differences in JK performance (of the kind to be discussed below) were rarely found between the two groups. Further mention will not be made of the letter-number definition of learning ability.

Because of such difficulties, learning ability was redefined as the number of correct recalls on the JK lists. Since Arbuckle and Cuddy (1969) found that the requirement to make predictions about future recall performance did not diminish recall, this measure seems an appropriate index of the desired construct. Still, the definition is less than optimal because learning ability is in effect defined as learning ability in a 16 item JK list. Subjects were then classified as fast, moderate, or slow learners on the basis of their rank in regard to the number of JK recalls. There were three sub-groups of eight subjects each for both the EL and the NoEL groups.

Prior to other analyses, a 2 X 3 X 4 ANOVA was performed in which the dependent variable was number of recalls on the four JK lists and in which the factors were EL judgment (presence or absence), Learning-ability (fast, moderate, slow), and Lists. Of the main effects, only that of learning-ability was reliable, $F(2, 42) = 90.29$, $p < .001$. The mean recalls per list for fast, moderate, and slow learners were 13.7, 10.9, and 6.8 recalls, respectively. Newman-Keuls analyses indicated that the three means were reliably different from each other. Because no interaction was reliable, recall for each of the three learning ability levels may be said to have been constant over the four lists.

Learning ability and practice at the judgment of knowing. The influence of learning ability on the improvement in JK accuracy with practice was analyzed first by a 2 X 3 X 4 ANOVA in which the factors were EL task (presence or absence), Learning-ability, and Lists. Since the effects of EL judgment and Lists were analyzed previously, only the effect of Learning-ability need be considered. The main effect proved reliable, $F(2, 42) = 5.88$, $p < .01$. Fast, moderate, and slow learners made 3.84, 5.03, and 5.14 erroneous JK predictions per list. Newman-Keuls analyses indicated that the difference between fast and moderate subjects was significant but not the difference between moderate and slow subjects. Fast subjects, then, tended to make less JK errors than either slow or moderate subjects.

Of the interactions, only that between learning-ability and lists was reliable, $F(6, 126) = 2.76$, $p < .025$. Figure 2 indicates mean performance over lists as a function of the three levels of learning-ability. The trend over lists for each of the three levels was analyzed separately. The linear component for the fast subjects was found to

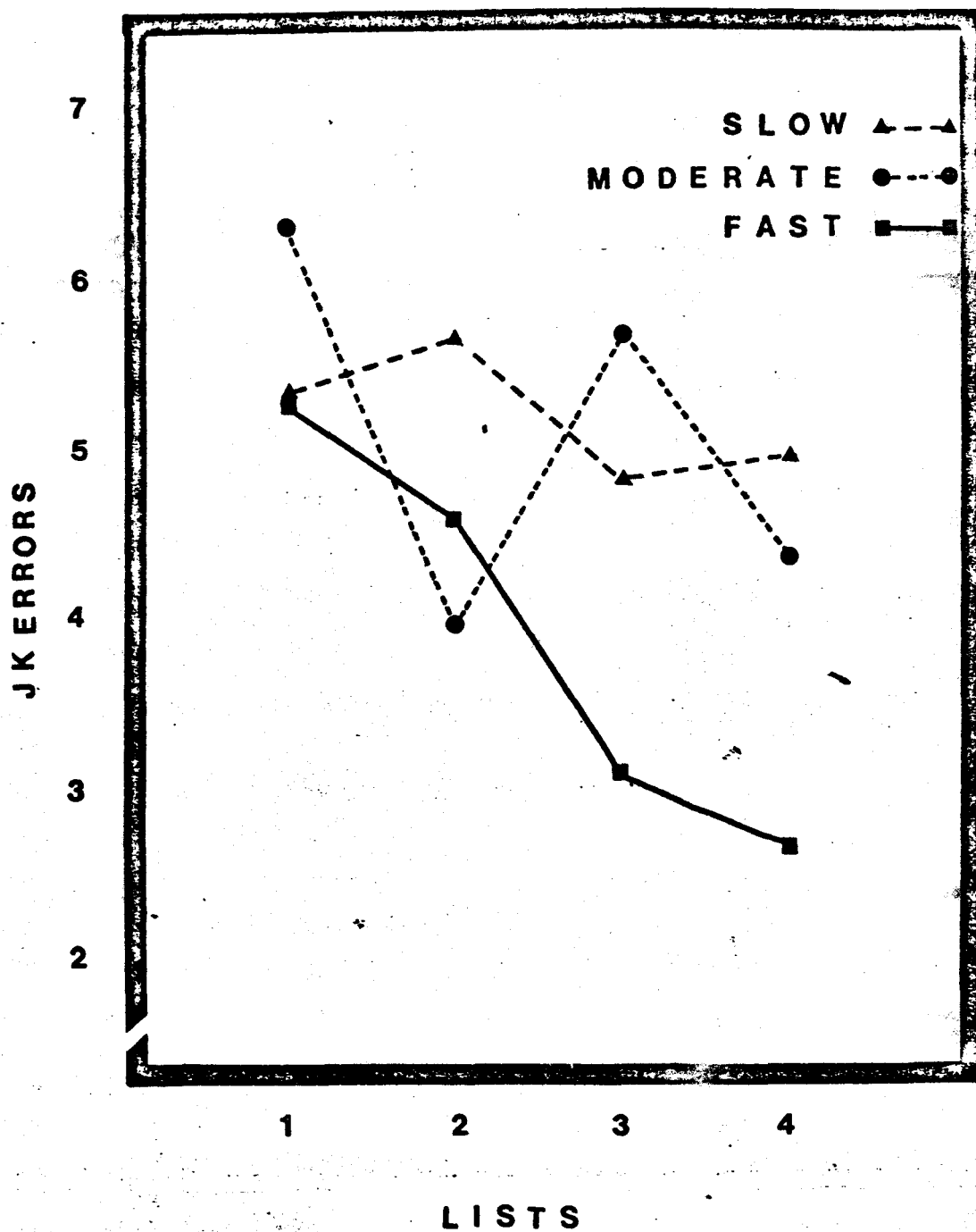


Figure 2. JK Errors over lists as a function of learning ability.

be significant, $F(1, 15) = 7.92, p < .05$. However, the linear component for moderate subjects failed to reach acceptable significance levels, $F(1, 15) = 3.31, .10 > p > .05$. Slow subjects were much less variable in JK accuracy over lists, $F(1, 15) = .49$.

The possibility is thus presented that, for situations in which only one trial is given subjects on a relatively short list, improvement in JK accuracy may somehow be strongly related to fast subjects' recall performance (13.7 items per list) being very near to the criterion of total recall (16 items).

In an attempt to clarify the role of learning-ability in the JK, the JK errors of the 32 subjects whose recall was middlemost among the 48 experimental subjects were observed over the four lists. Since mean recall was 10.8 items per list, recall performance near criterion levels was not relevant to improvement in JK accuracy for these subjects. The effect of lists was found reliable, $F(3, 93) = 2.82, p < .05$. Over lists, the mean number of JK errors decreased (5.66, 4.34, 4.72, 4.28). However, the linear component of the trend over lists was not reliable, $F(1, 31) = 2.45, p > .10$. Newman-Keuls analyses indicated that the only differences between lists in JK errors was that between Lists 1 and 2. Thus, although subjects of moderate learning ability can be said to have improved with practice relative to their initial performance, the changes over lists differed from the linear trend for fast learners. Because of such differences in trend, further comparisons between improvement rates for the two groups seem inappropriate. The clarification of the role of near-criterion recall in improvement by fast learners must await further research.

Summary. The analyses regarding the effect of learning ability

on the JK have indicated, first, that fast learners make a greater number of correct JK predictions than slow learners. However, the second finding is that only fast learners were able to improve their predictive accuracy with practice. Further research is needed to determine the role of near-criterion recall in this improvement.

The relationship between EL and JK ratings and learning ability.

Learning ability was also examined as a potential confound with respect to the relationship between EL and JK ratings. Because only 24 subjects were involved, a median split into fast and slow learners was considered preferable to three groups of eight subjects. Neither the Pearson correlations between EL and JK ratings nor the biserial correlations as indicators of the predictive accuracy of EL and JK ratings were greatly different for fast and slow learners, t 's $< .50$.

Secondary Analysis: Predictive congruence and learning ability.

Predictive congruence has previously been defined to occur when the JK and the EL ratings for the same items are similar, while predictive incongruence is said to occur when the two ratings are different. To examine predictive congruence as it is influenced by learning ability, a $2 \times 4 \times 2$ ANOVA was performed in which the factors were Learning-ability (Fast or Slow, 12 subjects at each level), Lists, and Congruence type (Congruent or incongruent predictions). The dependent variable was the probability of a correct prediction. The main effect of Congruence type was not significant, $F < 1$. However, the Learning ability \times Lists \times Congruence type interaction was significant, $F(3, 66) = 2.86$, $p < .05$. For fast learners, the probability of a correct prediction increased with practice for both congruent and incongruent predictions. For slow learners, however, the probability of a correct prediction

increased with practice only for congruent predictions (.61, .68, .72, .78). For the incongruent predictions of slow learners, the probability of a correct prediction decreased with practice (.70, .67, .62, .57). Slow learners, then, were rather accurate at predicting performance when the JK and EL ratings were in agreement but rather inaccurate when the JK was different from the EL rating.

Individual differences and learning ability. Various partial correlations were used to evaluate the relationship of learning ability to JK performance, FK performance, and scholastic achievement.

The first point of interest is the decrease in the correlation between JK Errors and FK Accuracy from $-.31$ to $-.19$ when the number of JK recalls is partialled out. However, no change occurred when the same partialling operation was performed for the JK Errors and FK Correct Predictions correlation. (The correlation remained at $.38$.) The different results are due to the much stronger relationship JK recall maintains with FK Accuracy ($.35$) than was present between JK recall and the FK Correct Predictions measure ($.08$).

Although no explanation is offered for ~~such~~ differences, a third partial correlation may be mentioned. Thus, the partial correlation between JK Errors on List 3 and FK Accuracy was seen to decrease only from $-.47$ to $-.41$ when the number of recalls on the four lists was partialled out. Similarly, the same correlation decreased only to $.37$ when the number of recalls on List 3 was kept constant. Definitive conclusions on the influence of learning ability on the JK relationship with FK capability are thus not present.

The final point regarding the relation between FK and JK performance is the rather strong correlation between recall on the four JK lists and accuracy at the positive FK predictions. For example, the correlation between FK Positive Accuracy (right less wrong regarding FK positive predictions) and JK recall was .48 when FK Positive Accuracy was evaluated by both first and second FK choices. Since the correlation with JK recall remains strong (.41) when the accuracy of positive FK predictions is defined by the proportion-based FK Hit measure, the relationship is not necessarily due to fast PA learners having a larger storehouse of general information available. The same conclusion is supported by the low correlation which the JK Miss proportion enters into with JK recall (.02). Similarly, this relationship is scarcely changed when the number of FK recalls is held constant while a new correlation is formed between JK recall and FK2 Positive Accuracy, from .48 to .44. Faster learners may simply be more aware of what they have in their memories.

Partial correlations for the JK and FK measures with educational achievement measures were also computed. The originally small correlations between overall educational achievement and JK and FK performance changed little when the number of JK recalls was partialled out. Of more importance was the maintenance of the moderate relations between LoTest scores and the JK and FK variables when the number of JK recalls was kept constant. The moderate relations which LoTest retained with JK Errors (-.31), List 3 JK Errors (-.37), and with FK Accuracy (.30) indicate further support for the presence of general factors other than learning ability in the JK.

Time at the EL task and the Judgment of Knowing. Some concern was

present because the increased study time allowed Group EL (a mean EL time of 4.8 minutes) did not lead to increased recall on the four JK lists relative to that for Group NoEL. However, the product-moment correlation between time at the EL task and number of correct recalls on the JK lists was of moderate strength, .39. The failure to find greater incidental learning may also indicate that EL judgments can be performed without the active responding operations which are important for learning.

CHAPTER IV

DISCUSSION

The judgment of knowing has been defined as the estimation of one's later memory for presently studied information. Three potential bases for the JK are stimulus knowledge, process knowledge, and memory knowledge. Although the three potential sources may not be mutually exclusive in actual learning situations, the concepts seem adequately distinct.

Stimulus knowledge was defined as knowledge which allows one to determine whether an item is easy or hard to learn. For example, an English speaker would probably rate a sentence in Turkish as harder to learn than the same sentence in Spanish--even before the two sentences have been heard. The basis for such stimulus knowledge is one's previous experience with stimuli which allows judgments of similarities and dissimilarities to be made.

The second type of knowledge was labeled process knowledge. This type of knowledge was meant to indicate a rather general factor which refers to a familiarity with the adequacy of either one's own learning operations or of learning operations in general. Thus, one who has learned that a good mark on an Organic Chemistry test requires four hours of preparation can rightly expect subpar performance if he has studied for only one hour.

The third type of knowledge was named memory knowledge. This refers to one's own knowledge of the present status of information in memory, to the knowledge that a memory has been formed. That memory

knowledge exists seems plausible in situations where associative attributes are present as aids to future recall. However, the presence of memory knowledge in learning situations where little aid to memory is offered by past knowledge is less certain.

Memory knowledge is theoretically distinguishable from stimulus knowledge and process knowledge because these two potential sources do not require an attempt to learn information while memory knowledge is otherwise not possible. A second reason for distinguishing memory knowledge from the other two types of knowledge is that subjects seem to have a general ability to indicate that a memory was laid down in the past. Positing memory knowledge thus postulates that subjects can know quite soon after a learning attempt whether or not they have formed a memory which is retrievable. Nevertheless, the possibility also exists that memory knowledge is merely a conjunct of stimulus and process knowledge such that a person may predict future recall performance because he can determine whether the amount of study given an item is proportionate to item difficulty.

The desire to increase general knowledge regarding the JK has limited the present experiment to more empirical than theoretical concerns. However, the questions were chosen to allow some sense of direction regarding the three potential bases of the JK.

Three general considerations regarding research on the JK.

Prior to a consideration of the specific questions asked in this research, three more general concerns may be discussed. First, it was feared that research on the JK may have strong artifactual components. That is, it is plausible that subjects may study items differentially

and thereby give quite accurate judgments of knowing, especially negative judgments. A related possibility is that evaluation apprehension may cause subjects to withhold recall for learned items which have been given negative JK ratings. However, the present results seem to indicate that such factors had little import on the JK.

Thus, although certitude about the presence, absence, or amount of differential study is impossible, the present finding of constant recall over lists would seem to indicate that such study would likely have been present from the first list. However, on the first list, study antedated information about the requirement of a JK. Therefore, any role for differential study would seem to be more as a constant factor than as a means for the facilitation of JK accuracy. Further, in the context of constant recall, the large increase over lists in the number of positive JK ratings, together with a decrease in the extremity of negative ratings, would seem to indicate that only a very limited number of items may be classified as having been unstudied.

The likelihood of selective withholding during recall after a negative JK seems more easily determined. Thus, selective withholding is postulated as a defense mechanism. However, subjects recalled 47% of List 1 items given ratings of 1. No great need seems to be present for subjects to avoid such theorized embarrassment.

One potential source for subjects' apparent motivation to study and recall all items is the presence of the game-points manipulation. Specifically, the pay-off matrix valued recall and nonrecall more highly than JK accuracy and rewarded or punished JK accuracy as a function of the extremity of the original rating.

A second general consideration regarding the JK task was the

adequacy of the measure of JK accuracy. The present research concerns required lists moderate in length and in difficulty. With such lists, it may be expected that there will be subjects at both extremes of recall performance. With practice it may also be expected that such subjects will be aware of their task proficiency and hence will utilize such knowledge in their predictions. A result may be many more items classified as recallable than nonrecallable or vice versa. If JK accuracy is measured by a proportion measure (v.g., the Hart Hit less Miss statistic), one of the two ratios for such subjects may be expected to be based on small numbers of instances. As has been seen, such proportions may be very deceptive. Since the sum of JK errors as an accuracy measure is not influenced by category usage, it seems a better description of the present data.

A third general concern was not considered until after the data had been collected. Specifically, later list items may either inhibit or facilitate recall performance. Thus, it is possible that the recall of a PA item which was rated as unrecallable may be facilitated by a later item serving as a mediator. Negative influences are similarly theorizable. That is, predictions based on the present state of knowledge which are nominally correct must be scored as incorrect.

However, JK predictions are meant to be made in light of a later test. Thus, "unlearning" would seem to be an indication that the PA item was less well known than had been thought. Facilitation from other list items, in contrast, can not be similarly explained. Although this may be tested by seeking an increase in the rates of Misses as the interval between JK and test increases, the present data were not designed to allow a sensitive test of such a possibility. Should it be desired

to minimize such influences, perhaps the Peterson and Peterson short-term memory paradigm can be utilized.

Accuracy, practice, and the Judgment of Knowing. That the JK was an accurate predictor of future recall performance as both a dichotomous measure and as a six-point scale replicates previous results. As has already been mentioned, the contribution of the present research in this regard is the definition of JK accuracy which allows JK performance to be considered on an individual subject basis over lists. The first new datum that this experiment offers, then, is the finding of improvement in JK accuracy with practice. Though small on an absolute scale, this improvement was found to occur in the context of only one learning trial per list.

Further, improvement in sensitivity was shown to be significant when negative JK predictions are considered as dichotomous measures. Although the trend for positive predictions was similar, this was not reliable. However, when usage of the scale positions is weighted so that extreme ratings are worth more than other ratings, the increase over lists in the number of points gained for positive ratings is significant. The magnitude of the increase over lists of such expressed confidence for positive predictions which are eventually proven correct (Hit items) is particularly striking.

A second result relative to the first question which is noteworthy is the bias change over lists from a predominance of Misses (incorrect predictions of nonrecall) to an equality of error types. It is as if subjects eventually made most of their predictions accurately and guessed randomly for the remaining predictions.

A second interpretation of the fact of improvement with practice

refers to the characteristics of the stimulus knowledge which may serve as a basis for the JK. That is, it is plausible that knowledge about the relative ease of learning of stimuli may be a permanent part of one's memory for particular words. However, the fact of increased accuracy with practice would seem to require that such knowledge include a learned component. What must be learned is not evident from the present data. For example, the requirement of deciding a single JK for a paired associate item may bear no direct relationship to separate judgments about each individual term and hence may have to be learned.

The ease of learning judgment and the Judgment of Knowing. In seeking to clarify the role of stimulus knowledge in the JK, the second question considered the relationship between ease of learning ratings and judgments of knowing for the same items. To minimize any role for process knowledge in the EL judgment, the task occurred prior to any experience with PA learning.

The first finding in this regard is somewhat ambivalent. Thus, the product-moment correlation between the mean EL and JK ratings was very strong (.80). However, no individual subject manifested so direct a relationship. The mean individual correlation was much lower (.32). Further, the consideration of the EL as a dichotomous predictor of the JK indicated that for some subjects the hard-easy to learn distinction of the EL judgment was always a poor predictor of the JK while, for others, good predictability on the first list dropped considerably with practice. It would seem that a rather strong relationship is theoretically possible between the EL and JK ratings for the same items. However, the relationship existing for individual subjects is generally only moderate. Further, because of the rather limited predictability

of the dichotomous EL regarding the dichotomous JK, any strong correlation may be based more on the relative extremity of some individual items than on the hard-easy dichotomy itself.

A second datum from this question is that the predictive validity of the JK was found consistently superior to the EL judgment. Thus, the mean EL ratings for items were somewhat accurate predictors of Group EL's number of recalls for the same items, $r = .49$. (It may be noticed that this is not as large a correlation as those around .90 which have been found in the free recall task by Underwood, 1966.) However, the same subjects' later mean JK values were better predictors of future recall performance than the EL ratings, $r = .64$. Similar differences were present with individuals' EL and JK ratings.

An attempt was next made to discover if subjects could predict their own performance better than could the group. Individual EL values were found no more accurate in predicting recall than the group EL values. However, the individual JK values of Group EL subjects were slightly better predictors of recall performance than the group JK values. More importantly, the individual JK values of NoEL subjects were significantly more accurate than the JK values of Group NoEL. Contrary to results with EL judgments (Underwood, 1966), the JK ratings of individual subjects in a PA task were better predictors of their own ratings than were the group's ratings.

The general superiority of JK ratings over EL ratings also has theoretical import. Thus, if the JK ratings of Group NoEL were made on a basis similar to the hard-easy dichotomy present in the EL judgments, their accuracy could then be expected to be similar to that of the EL ratings, especially on List 1 when feedback was not present for

Group NoEL. The differences, however, were quite substantial: the mean biserial correlation for the JK ratings of Group NoEL was .59 while that for EL ratings was .23. The strong EL inferiority makes it very doubtful that subjects in Group NoEL based their JK ratings on the knowledge manifest in the ease of learning judgments. The experimenter-paced study and JK trials given subjects may have increased their capability at determining if a PA item appears hard or easy to learn. Similarly, memory knowledge may have been involved in the JK.

However, it may also be mentioned that, while improvement in stimulus knowledge is a possibility, the means whereby such improvement can occur are not immediately obvious. That is, for the free recall of individual items, highly accurate predictions based on EL judgments, pronunciability, and association value have been found to correlate highly (Underwood, 1966). Similarly, EL judgments correlated highly with judgments regarding the degree of similarity of meaning of PA items (Richardson & Erlebacher, 1958). Thus, improvement is unlikely to occur merely by the use of different objective criteria.

A third interesting result from the examination of the EL and JK relationship is the finding that the EL task leads to poorer JK performance. This is present on all lists and is not clearly related to the type of prediction. Thus, when JK accuracy on List 1 was compared for Group EL and Group NoEL, the mean biserial correlations were .36 and .59 for the respective groups. It was also found that the JK predictions of Group NoEL involved greater confidence than those given by subjects familiar with the items. Though the performance of the EL task may have caused such differences, the causality would seem more likely to be the result of item familiarity. However, since the present

experiment was designed to observe potential effects of the EL task rather than to control its role in the JK, further research is required before even the basic datum is shown reliable.

Such experimentation may be rather similar to that necessary to offer comments on the roles of stimulus and memory knowledge in the JK. In essence, what is required is a procedure which will effect memory knowledge but allow stimulus knowledge to be constant. Better controlled familiarization is one potential means to assess this concern. However, a procedure of more value than familiarization in the disentangling of stimulus and memory knowledge would take advantage of the known effects of the AB, AC paradigm in PA learning. Thus, subjects would first learn a number of PA lists with either homogeneous or randomly chosen items. The expectation is for an improvement in JK accuracy and an increase in confidence over lists. Then, for one group, the last list may take the form of an AC relationship to the previous AB list. Further, the number of trials on all lists should be such that the AC association on the last list may be approximately equal to that for the AB association. The other group, of course, would learn a new last list.

The interpretation of causality on the trials before the last list is unimportant because stimulus and memory knowledge would be confounded. On the last trial, however, the hypothesis of a memory knowledge basis for JK ratings would seem to expect less certain JK ratings for the AC than for the CD group because the actual knowledge state can be expected to be one of less strong PA associations. The stimulus knowledge interpretation, however, should not expect any change, unless possibly more extreme ratings because of greater familiarity with the stimulus terms. The accuracy of such predictions would seem to be less important

than the confidence expressed in the JK ratings because predictions based both on stimulus knowledge and on memory knowledge may be expected to lead to less accurate ratings.

Even in such a design, however, it must be mentioned that process knowledge may serve as a confound. That is, a person to some degree fearful that the changed stimulus-response term relationship may be detrimental to future recall might thereby decrease his JK ratings. However, further research is necessary before such a potential confound need be considered.

Two results of secondary analyses regarding the JK may also be mentioned. First, it was found that items to which subjects gave dichotomously similar JK and EL ratings had probabilities of correct predictions which were very similar to those for items with JK ratings dichotomously opposite to EL ratings. Dichotomous similarity between JK and EL ratings may thus be epiphenomenal rather than reflective of similar underlying bases.

A second result was that JK ratings were found to be not simply related to levels of the stimulus characteristics of response term imagery and learnability (the number of times an item was recalled). Thus, items whose predictions proved to be accurate were positively related to the stimulus characteristics. However, items whose predictions proved erroneous did not appear to receive JK ratings in relation to their position on the relevant stimulus characteristic continuum. It is difficult to conceive why items on similar levels of a stimulus characteristic would not be given similar JK ratings if subjects made JK ratings on the basis of the stimulus characteristic alone.

Feeling of knowing judgments, scholastic achievement, and the

Judgment of Knowing. In examining the relationship between the FK judgment and the JK, the reasoning was that proficiency at the JK task ought to be similar to that at the FK task if the JK indicates a knowledge of the contents of one's memory. Moderate correlations were generally found. Since the difficulty of the questions led to negative FK judgments being much easier to make than positive predictions, many questions offered little discriminability between subjects as to FK capabilities. It is thus possible that relationships at the two tasks would be stronger if FK items required a greater degree of memorial assessment. However, while some support is present for the hypothesis that both JK and FK tasks involve a knowledge of memorial contents, the correlation does not specify whether such knowledge is inferential or more direct. More importantly, it must be noted that "knowledge of memorial contents" need not have the same meaning when it is the basis for a prediction of recognition (i.e., a feeling of knowing) as when it supports a prediction of recall (i.e., a Judgment of Knowing). Because of this, a specific interpretation of the moderate correlation would be very tenuous.

It may also be mentioned that FK proficiency was much more strongly related to JK accuracy than to EL accuracy while JK accuracy was similarly related to EL and FK accuracy. The abilities required for JK accuracy may thus facilitate both unpracticed stimulus evaluation and the knowledge of the contents of one's memory. However, the abilities required for the EL judgment may be dissimilar to those needed for the feeling of knowing.

The second individual difference variable examined was scholastic achievement. It was theorized that, if a decision similar to the JK is

required for successful school performance, high correlations should be found between scholastic success and JK ability. The relationship between JK (and FK) performance with the overall educational achievement measures (SAT, ACT, HS Rank) was low at best. Because the present achievement scores were skewed toward the higher direction, low correlations might not have been unexpected. However, it is at least as plausible that the low relationships were due to task dissimilarity.

The JK was also examined in its relationship to performance in an introductory psychology course. The correlations were at their highest (.37 with JK Errors; .35 with FK Accuracy) when the correlate considered was the subject's lowest test score for the course. Why this correlation should be stronger than others is not known. However, it is possible that a JK-scholastic performance relationship can only be manifest when study is not prolonged, as might be the case when test performance is at its worst. That is, it is possible that there is a relationship between JK ability and school grades but that highly motivated subjects of low JK ability can maintain a high level of performance by overlearning material. To test such a hypothesis, a better scholastic situation would involve a limited study time to prevent overlearning.

Learning ability and the Judgment of Knowing. In attempting to delineate the influence of learning ability in the previous three questions, learning ability was defined as the number of correct recalls on the four JK lists. The first result of note was that fast learners were found to have made more correct JK predictions than slow learners.

The second important result was the finding that improvement in

JK accuracy was related to learning ability. Only subjects of high levels of learning ability showed linear improvement in JK accuracy with practice. Subjects of slow and moderate learning speed did not show similar improvement. These results question the generality of the finding of improvement in JK accuracy with practice.

The nearness of the fast learners to total list recall also introduces the possibility that improvement is somehow an artifact of the fast learning of a relatively short PA list. That is, fast learners' increase in the number of positive JK predictions over lists (8.8, 10.9, 13.2, 14.1) may be related to their constant recall of 13.7 items per list. Specifically, for a 16 item list, fast learners' increase in positive JK predictions in effect limited the number of possible JK errors. For example, the average fast learner who recalled a constant 14 items per list could have made up to 10 JK errors on List 2 when he predicted recall for only 11 items but could only have made four JK errors on List 4 when he predicted recall for 14 items. Although the actual role of near-criterion recall in JK improvement is unclear, the possibility of such an influence would be less viable if the JK task occurred for a longer list in which total recall would be less likely.

A longer list may also clarify the role of a second potential factor in the JK improvement of fast learners. Thus, it is possible that fast learners are more aware of differences between the memorial status of to be learned materials than are other subjects. Having a better referent for a "known" item than other subjects, they may learn more accurately which items are not encompassed by the definition.

Partial correlations which kept the effect of learning ability constant were also used to examine the moderate correlations entered

into by JK accuracy with FK capabilities and with scores on subjects' lowest course tests.

The status of the JK relationship with FK capability is unclear because the relationship with FK Accuracy increased from $-.31$ to $-.19$ while that with the right-less-wrong measure of FK accuracy remained constant at $.38$. The far stronger relationship between recall performance and FK Accuracy than with FK Correct Predictions is the major cause of the decreased relationship. The relationship between LoTest scores and JK accuracy was unchanged when the partial correlation kept learning ability constant.

Lastly, a surprising finding was the rather high correlation between learning ability and accuracy at the positive FK predictions. (The relationship is slightly lower when negative predictions are included together with positive predictions. This decrease would seem to be related to the difficult questions not differentiating between subjects.) The relationship does not seem due to the fact that fast learners are likely to have more information available. The correlation changes little when positive FK performance is considered as either an absolute number or as a proportion or when FK recall is partialled out.

It is thus possible that the ability to know one's memorial contents is central to rapid learning. That is, it is possible that learning occurs more quickly when subjects are more aware of the running content of their memories. It is thus possible that reflection on the to be learned material is intimately related to the facilitation which generally occurs when subjects are asked to use mnemonics or natural language mediators to learn paired-associates and to the facilitation of free recall learning from conscious attempts at categorization. If

slow learners can be encouraged to reflect more on their knowledge by such means as frequent self-testing, improvement in memory tasks may be predicted.

Summary of the present research. The present study has dealt with the prediction of later retention for presently studied information. Such an estimation about whether study has been sufficient to allow later recall has been called the Judgment of Knowing (JK).

Three types of knowledge have been offered as potential bases for the JK. However, lack of information about the JK has led the present concern to be more data-oriented than theoretical.

The initial question concluded that the JK is a valid predictor of response term recall in a paired-associate task. Although the positive JK was more likely to predict performance accurately than the negative JK, both types were reliably more often correct than incorrect.

More importantly, the first question found that JK accuracy increased linearly with practice. Further, while listwise improvement was reliable for negative but not for positive predictions, confidence in positive predictions which were later proven correct increased much more over lists than confidence in positive predictions later proven incorrect.

The second question examined the relationship between the JK and a judgment about items' ease or difficulty of learning (Ease of Learning or EL judgment). A very high correlation was present between mean EL and JK ratings for the same items. However, the same relationship for the average subject was only moderate. Further, on a dichotomous basis, the EL judgment had little validity as a predictor of an item's later JK for half the subjects on all four lists and was a decreasingly accurate predictor over lists for subjects whose original EL and JK

correspondence was high.

The EL judgment was also found inferior to the JK as a predictor of recall performance.

Lastly, the EL data may indicate that experience with to be learned material shortly before the JK task reduces both the accuracy and the expressed confidence of the JK.

The third question compared the JK and the feeling of knowing judgment in which subjects predict their likelihood of correctly recognizing unrecallable answers to general information questions. A moderate correlation was found.

Comparisons were also made between the JK and scholastic performance. Correlations between JK accuracy and measures of overall educational achievement (v.g., SAT) were positive but low. However, a moderate correlation was found between JK accuracy and scores on the subjects' lowest test in an introductory psychology course.

The fourth question examined the role of learning ability in the JK. Fast learners were found to make more correct predictions than slow or moderate-speed learners. When JK accuracy for the three levels of learning ability was examined over lists, the only linear trend which was significant was that for fast learners. Further research is needed to determine if the fast learning of a relatively short list is a confound in JK improvement with practice.

REFERENCES

- Adams, J.A. Human memory. New York: McGraw-Hill, 1967.
- Adams, J.A. Response feedback and learning. Psychological Bulletin, 1968, 70, 486-504.
- Adams, J.A., McIntyre, J.S., & Thorsheim, H.I. Response feedback and verbal retention. Journal of Experimental Psychology, 1969, 82, 290-296.
- Adams, P.A., & Adams, J.K. Training in confidence-judgments. American Journal of Psychology, 1958, 71, 747-753.
- Adams, P.A., & Adams, J.K. Confidence in the recognition and reproduction of words difficult to spell. American Journal of Psychology, 1960, 73, 544-552.
- Appel, L., Cooper, R., Knight, J., McCarrell, N., Yussen, S., & Flavell, J. The development of the distinction between perceiving and memorizing. Child Development, 1972, 43, 1365-1381.
- Arbuckle, T.Y., & Cuddy, L.L. Discrimination of item strength at time of presentation. Journal of Experimental Psychology, 1969, 81, 126-131.
- Battig, W.F., & Spera, A.J. Rated association values of numbers 0-100. Journal of Verbal Learning and Verbal Behavior, 1962, 1, 200-202.
- Beeson, R.O. Immediate knowledge of results and test performance. Journal of Educational Research, 1973, 66, 224-226.
- Berch, D.B., & Evans, R.C. Decision processes in children's recognition memory. Journal of Experimental Child Psychology, 1973, 16, 148-164.

- Bernbach, H.A., & Bower, G.H. Confidence ratings in continuous paired-associate learning. Psychonomic Science, 1970, 21, 252-253.
- Biggs, D.A., & Tinsley, D.J. Student-made academic predictions. Journal of Educational Research, 1970, 63, 195-197.
- Binder, D.M., Jones, J.G., & Strowig, R.W. Non-intellective self-report variables as predictors of scholastic achievement. Journal of Educational Research, 1970, 63, 364-366.
- 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.
- Blum, M.L. Ability of students to estimate their grade on a multiple-choice examination. Journal of Educational Psychology, 1935, 26, 547-551.
- Brown, C.H. The relation of magnitude of galvanic skin responses and resistance levels to the rate of learning. Journal of Experimental Psychology, 1937, 20, 262-278.
- Brown, R., & McNeill, D. The "tip of the tongue" phenomenon. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 325-337.
- Bugelski, B.R. Presentation time, total time, and mediation in paired-associate learning. Journal of Experimental Psychology, 1962, 63, 409-412.
- Bugelski, B.R., & Rickwood, J. Presentation time, total time, and mediation in paired-associate learning: Self-pacing. Journal of Experimental Psychology, 1963, 65, 616-617.
- Colman, F., & Paivio, A. Pupillary dilation and mediation processes during paired-associate learning. Canadian Journal of Psychology, 1970, 24, 261-270.

- DaPolito, F., Guttenplan, H., & Steinitz, H. Accuracy of subjective judgments of information in long-term memory. Psychonomic Science, 1968, 13, 227-228.
- Flavell, J.H., Beach, D.R., & Chinsky, J.M. Spontaneous verbal rehearsal in a memory task as a function of age. Child Development, 1966, 37, 283-299.
- Flavell, J.H., Friedrichs, A.G., & Hoyt, J.D. Developmental changes in memorization processes. Cognitive Psychology, 1970, 1, 324-340.
- Foote, R., & Belinky, C. It pays to switch? Consequences of changing answers on multiple-choice examinations. Psychological Reports, 1972, 31, 667-673.
- Freedman, J.L., & Landauer, T.K. Retrieval of long-term memory: "Tip-of-the-tongue" phenomenon. Psychonomic Science, 1966, 4, 309-310.
- Grasha, A.F., Schumsky, D.A., & Elliott, L.E. Relationships among short-term recall, intralist intrusions, subjective certainty ratings, and long-term memory. Journal of Experimental Psychology, 1973, 100, 83-94.
- Gruneberg, M.M., & Monks, J. 'Feeling of knowing' and cued recall. Acta Psychologica, 1974, 38, 257-265.
- Gruneberg, M.M., Smith, R.L., & Winfrow, P. An investigation into response blockaging. Acta Psychologica, 1973, 37, 187-196.
- Haltrecht, E.J., & McCormack, P.D. Monitoring eye movements of slow and fast learners. Psychonomic Science, 1966, 6, 461-462.
- Hanna, G.S. Effects of total and partial feedback in multiple-choice testing upon learning. Journal of Educational Research, 1976, 69, 202-205.
- Hart, J.T. Memory and the Feeling-of-Knowing experience. Journal of

Educational Psychology, 1965, 56, 208-216.

Hart, J.T. Memory and the memory-monitoring process. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 685-691. (a)

Hart, J.T. Second-try recall, recognition, and the memory-monitoring process. Journal of Educational Psychology, 1967, 58, 193-197. (b)

Jenkins, J. Can we have a theory of meaningful memory? In R. Solso (Ed.), Theories in cognitive psychology: The Loyola Symposium. New York: Wiley, 1974. (a)

Jenkins, J. Remember that old theory of memory? Well, forget it! American Psychologist, November, 1974, 785-795. (b)

Kahneman, D., & Beatty, J. Pupil diameter and load in memory. Science, 1966, 154, 1583-1585.

Kahneman, D., Tursky, B., Shapiro, D., & Crider, A. Pupillary, heart rate, and skin resistance changes during a mental task. Journal of Experimental Psychology, 1969, 79, 164-167.

Keefer, K.E. Characteristics of students who ~~make~~ accurate and inaccurate self-predictions of college achievement. Journal of Educational Research, 1971, 64, 401-404.

Kellas, G., McCauley, C., & McFarland, C.E., Jr. Reexamination of externalized rehearsal. Journal of Experimental Psychology: Human Learning and Memory, 1975, 104, 84-90.

Kintsch, W. Habituation of the GSR component of the orienting reflex during paired-associate learning before and after learning has taken place. Journal of Mathematical Psychology, 1965, 2, 330-341.

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.

Landauer, T.K. Rate of implicit speech. Perceptual and Motor Skills, 1962, 15, 646.

Lovelace, E.A. Prediction during learning of later retrievability. Paper presented at the 15th Annual Meeting of The Psychonomic Society, Boston, November, 1974.

Lowe, M., & Crawford, C.C. First impressions vs. second thought in true-false tests. Journal of Educational Psychology, 1929, 20, 192-195.

McCormack, P.D. Recognition memory: How complex a retrieval system? Canadian Journal of Psychology, 1972, 26, 19-51.

McCormack, P.D., & Haltrecht, E.J. Monitoring eye movements under two conditions of presentation of paired-associate materials. Canadian Journal of Psychology, 1966, 20, 154-159.

Murray, D.J. The effect of white noise upon the recall of vocalized lists. Canadian Journal of Psychology, 1965, 19, 333-345.

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. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263.

Paivio, A., Smythe, P.C., & Yuille, J.C. Imagery versus meaningfulness of nouns in paired-associate learning. Canadian Journal of Psychology, 1968, 22, 427-441.

Paivio, A., Yuille, J.C., & Madigan, S.A. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology, 1968, 76 (No. 1, Pt. 2), 1-25.

- Pappas, B.A., & Suboski, M.D. Recognition operating characteristics as a function of prior recall confidence. Psychonomic Science, 1966, 6, 83-84.
- Peterson, L.R., & Peterson, M.J. Short-term retention of individual verbal items. Journal of Experimental Psychology, 1959, 58, 193-198.
- Postman, L. Does interference theory predict too much forgetting? Journal of Verbal Learning and Verbal Behavior, 1963, 2, 40-48.
- Reile, P.J., & Briggs, L.J. Should students change their initial answers on objective-type tests? More evidence regarding an old problem. Journal of Educational Psychology, 1952, 43, 110-115.
- Richardson, J., & Erlebacher, A. Associative connection between paired verbal items. Journal of Experimental Psychology, 1958, 56, 62-69.
- Rogoff, B., Newcombe, N., & Kagan, J. Planfulness and recognition memory. Child Development, 1974, 45, 972-977.
- Ross, C. The influence upon achievement of a knowledge of progress. Journal of Educational Psychology, 1933, 24, 609-619.
- Royer, J.M. Memory effects for test-like-events during acquisition of foreign language vocabulary. Psychological Reports, 1973, 32, 195-198.
- Rubin, D.C. Within word structure in the tip-of-the-tongue phenomenon. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 392-397.
- 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, 1-12.
- Smythe, P.C., & Paivio, A. A comparison of the effectiveness of word imagery and meaningfulness in paired-associate learning of nouns.

Psychonomic Science, 1968, 10, 49-50.

Strang, H.R., & Rust, J.O. The effects of immediate knowledge of results and task definition on multiple-choice answering. Journal of Experimental Education, 1973, 42, 77-80.

Suboski, M.D. Signal detection analysis of recall paired-associate learning. Psychonomic Science, 1967, 7, 357-358.

Suboski, M.C., Pappas, B.A., & Murray, D.J. Confidence ratings in recall paired-associates learning. Psychonomic Science, 1966, 5, 147-148.

(a)

Suboski, M.D., Pappas, B.A., & Murray, D.J. Confidence ratings in recall paired-associates: The RTT paradigm. Psychonomic Science, 1966, 5, 315-316. (b)

Thompson, L.W., & Obrist, W.D. EEG correlates of verbal learning and overlearning. Electroencephalography and Clinical Neurophysiology, 1964, 16, 332-342.

Thorndike, E.L., & Lorge, I. The teacher's wordbook of 30,000 words. New York: Teachers College, Bureau of Publications, 1944.

Tulving, E. Subjective organization in free recall of "unrelated" words. Psychological Review, 1962, 69, 344-354.

Tulving, E., & Madigan, S.A. Memory and verbal learning. Annual Review of Psychology, 1970, 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. Attributes of memory. Psychological Review, 1969, 76, 559-573.

Underwood, B.J. Recognition memory. In H.H. Kendler and J.T. Spence

(Eds.), Essays of neobehaviorism. New York: Appleton, Century, Crofts, 1971.

Underwood, B.J. The role of association in recognition memory. Journal of Experimental Psychology, 1974, 102, 917-939.

Underwood, B.J. Individual differences as a crucible in theory construction. American Psychologist, February, 1975, 128-134.

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

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

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

Zechmeister, E.B., & Shaughnessy, J.J. When you know that 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.

APPENDIX A

Stimulus and response terms for the PA lists

List A	Response Imagery*	Mean EL*	Mean JK*	Recalls*
GUN - CONCEPT	1.93	58.7	4.2	30
COW - ADAGE	2.77	75.4	3.0	30
NET - GENDER	2.90	69.1	3.8	23
GEM - DECEIT	3.30	64.1	3.9	31
TAR - NAMESAKE	3.37	61.9	3.6	36
BAY - WORKHOUSE	4.00	59.7	4.1	36
MAN - PREVIEW	4.03	60.8	3.9	37
CAT - IMPACT	4.43	51.6	4.2	33
JUG - CHAOS	4.57	62.5	4.2	28
HAM - PORTAL	5.10	76.2	3.2	35
RAG - INVOICE	5.17	72.1	3.3	26
PAN - DUMMY	5.83	52.3	3.8	29
PUB - BANDIT	5.83	54.3	4.6	35
DEN - HURDLE	6.33	57.5	3.7	32
BOX - POSTER	6.33	38.4	4.5	39
KEY - BOUQUET	6.77	58.0	4.0	34

* The Response Imagery values are on a 1-7 scale, with higher values indicating a greater degree of imageability.
 The Mean EL values are on a 1-120 scale, with higher values rated as harder to learn.
 The Mean JK values are on the 1-6 scale previously described.
 The maximum number of Recalls is 48.

List B	Response Imagery	Mean EL	Mean JK	Recalls
RAT - CONTEXT	2.13	63.0	3.4	26
BAG - PROXY	2.70	79.5	3.1	24
GUM - ABBESS	2.97	77.9	3.4	31
COT - MALICE	3.30	80.4	3.2	20
LOG - SATIRE	3.37	68.5	3.3	34
NUT - BOREDOM	3.83	46.0	4.5	30
PIN - GARRET	4.13	76.2	3.3	19
MOP - TIDBIT	4.37	66.7	3.7	31
LEG - CASEMENT	4.63	47.5	4.7	40
LIP - PHANTOM	5.03	66.5	4.4	34
FIG - KERCHIEF	5.23	60.1	4.2	28
GAS - HORSEHAIR	5.67	59.8	4.3	30
PAW - STOREROOM	5.87	65.1	3.8	22
CUP - LOCKER	6.27	48.6	4.6	38
JAW - DAYBREAK	6.43	47.2	4.7	40
VAN - SUNBURN	6.72	52.5	5.1	47

List C	Response Imagery	Mean EL	Mean JK	Recalls
HAY - FOIBLE	2.20	87.6	2.6	14
MUD - RATING	2.60	54.7	3.8	24
FAN - STEERAGE	3.00	75.3	3.1	19
BAT - HINDRANCE	3.07	73.8	3.9	32
DOT - ONSLAUGHT	3.67	68.5	3.2	33
JAM - GADFLY	3.77	59.2	3.1	30
BUG - ROSIN	4.20	68.4	3.4	16
CAR - BUFFOON	4.33	66.8	4.4	42
WEB - DEMON	4.70	38.6	5.0	44
WAX - LIMELIGHT	4.83	60.2	4.1	39
KIT - GINGHAM	5.33	77.5	2.7	20
ROD - MAMMAL	5.57	63.8	3.9	35
BOW - REPTILE	6.00	57.5	4.1	33
FUR - TRIPOD	6.23	76.8	3.8	30
HEN - SPINACH	6.47	60.4	4.5	34
MAP - SHOTGUN	6.60	54.8	4.2	40

List D	Response Imagery	Mean EL	Mean JK	Recalls
TIN - ESSENCE	2.33	63.3	4.1	34
DOG - OUTCOME	2.40	51.6	4.4	32
JET - SAVANT	3.07	80.9	3.1	30
CAN - UPKEEP	3.07	55.5	4.2	37
FOX - OFFSHOOT	3.67	54.4	3.7	23
SAW - PRESTIGE	3.67	68.8	3.6	32
PAD - CHLORIDE	4.20	74.0	3.8	22
TUB - VENOM	4.23	74.9	4.3	40
TOY - REFLEX	4.73	48.5	4.5	39
WIG - TRACTION	4.77	73.3	3.9	36
SUN - COWHIDE	5.40	51.6	4.8	39
RIB - LEAFLET	5.47	64.8	3.1	25
RUG - GOBLET	6.03	62.4	4.1	36
BAR - HAIRPIN	6.13	55.0	4.2	37
PEN - BLISTER	6.53	37.7	4.8	36
HAT - LOBSTER	6.57	51.3	4.3	27

APPENDIX B

Feeling of knowing questions

The first number beneath each question indicates the number of times the answer was recalled by the 48 subjects; the second number indicates the mean FK judgment for the item when the answer was not recalled; the third value is the probability of a correct recognition given a positive FK (Hit), the last value the probability of a correct recognition given a negative FK (Miss). The answers appear on a separate sheet which follows the questions.

1. Who was the Democratic head of the House Judiciary Committee during last year's impeachment hearings?

2 3.7 10/27 = .37 8/19 = .42

- | | |
|----------------------|-----------------|
| a. Edward Hutchinson | c. James Rhodes |
| b. Peter Rodino | d. Carl Albert |

2. Who wrote the James Bond novels?

18 3.0 13/15 = .87 7/15 = .47

- | | |
|------------------|--------------------|
| a. John le Carre | c. John Cheevers |
| b. Ian Fleming | d. Peter MacDonald |

3. Who painted "Afternoon at La Grand Jatte"?

1 1.9 0/3 = .00 5/44 = .11

- | | |
|------------------|-------------------|
| a. Edouard Manet | c. Pierre Bonnard |
| b. Henri Matisse | d. Georges Seurat |

4. What is the capital of Saudi Arabia?

0 3.0 5/20 = .25 6/28 = .21

- | | |
|-----------|------------|
| a. Medina | c. Bahrain |
| b. Riyadh | d. Qatar |

5. Who won an Academy Award for the female lead in the movie "The Prime of Miss Jean Brodie"?

3 3.4 $9/22 = .41$ $5/23 = .22$

- a. Maggie Smith c. Vanessa Redgrave
b. Glenda Jackson d. Susanna York

6. Who was the first man to step on the moon?

23 4.0 $14/17 = .82$ $6/8 = .75$

- a. Edwin Aldrin c. Neil Armstrong
b. Frank Borman d. Alan Shephard

7. Who wrote Go Tell It on the Mountain?

2 2.4 $4/10 = .40$ $11/36 = .31$

- a. Richard Wright c. Leroy Jones
b. James Baldwin d. Ralph Ellison

8. For whom was "Summer in the City" a hit record?

12 3.2 12/17 = .71 6/19 = .32

- a. Beach Boys c. Lovin' Spoonful
b. Fifth Dimension d. Carpenters

9. Who was the first European to explore the length of the Mississippi River?

2 3.5 $8/26 = .31$ $4/20 = .20$

- a. Louis Jolliet c. Jacques Marquette
b. Jean Nicolet d. Sieur de la Salle

10. Who sang "The Look of Love" in the movie Casino Royale?

4 2.9 $12/17 = .71$ $14/27 = .52$

- | | |
|--------------------|-------------------|
| a. Roberta Flack | c. Dionne Warwick |
| b. Aretha Franklin | d. Diana Ross |

11. Whose creation is Holden Caulfield?

6 2.0 $6/6 = 1.00$ $14/36 = .39$

- | | |
|------------------|------------------|
| a. J.R. Tolkien | c. John Updike |
| b. Joseph Heller | d. J.D. Salinger |

12. Who immediately succeeded Stalin as the head of state in Russia?

0 1.8 $8/31 = .26$ $4/17 = .24$

- | | |
|--------------------|-----------------------|
| a. Georgi Malenkov | c. Nicolai Bulganin |
| b. Lavrenti Beria | d. Vyacheslav Molotov |

13. Who founded Hull House in Chicago?

9 2.9 $15/19 = .79$ $7/20 = .35$

- | | |
|-----------------|----------------------------------|
| a. Dorothea Dix | c. Mother Frances Xavier Cabrini |
| b. Jane Adams | d. Mother Katherine Drexel |

14. What boxer defeated Joe Louis for the heavyweight boxing title when Louis attempted to return from retirement?

0 2.9 $1/18 = .06$ $0/30 = .00$

- | | |
|-----------------------|-----------------------|
| a. Rocky Marciano | c. Sugar Ray Robinson |
| b. Jersey Joe Walcott | d. Ezzard Charles |

15. Who wrote the poem "Leaves of Grass"?

7 2.0 $3/5 = .60$ $8/36 = .22$

- | | |
|-------------------------------|-----------------|
| a. Henry Wadsworth Longfellow | c. Robert Frost |
| b. James Russell Lowell | d. Walt Whitman |

16. Who painted "Aristotle contemplating the bust of Homer"?

3 2.1 $3/5 = .60$ $7/40 = .18$

- | | |
|-----------------------|-----------------------|
| a. Rembrandt van Rijn | c. El Greco |
| b. Vincent van Gogh | d. Peter Paul Reubens |

17. Who played the role of the airport manager in the movie "Airport"?

12 3.6 $12/20 = .60$ $5/16 = .31$

- a. Rod Taylor c. Charlton Heston
b. Richard Burton d. Burt Lancaster

18. Who was the man Capone sought to kill on St. Valentine's Day?

10 3.2 $15/20 = .75$ $11/18 = .61$

- a. J. McGurn c. G. Moran
b. D. O'Banion d. J. Colosimo

19. Who was the boatman on the river Styx who ferried the souls of the dead to Hades?

0 2.9 $9/18 = .50$ $8/30 = .27$

- a. Jason c. Charon
b. Nestor d. Pluto

20. Who was the last horse to win horse racing's triple crown before Secretariat?

1 2.6 $9/14 = .64$ $14/33 = .42$

- a. Whirlaway c. Count Fleet
b. Man O' War d. Citation

21. Who was the official (administrative) head of the White House plumbers?

0 3.0 $1/22 = .05$ $5/26 = .19$

- a. E. Howard Hunt c. Egil Krogh
b. Charles Colson d. G. Gordon Liddy

22. Who wrote the play "She Stoops to Conquer"?

1 2.1 $4/7 = .57$ $8/40 = .20$

- a. Oliver Goldsmith c. Alexander Pope
b. Richard Sheridan d. Christopher Marlowe

23. What city is the capital of Canada?

4 4.5 $8/36 = .22$ $3/8 = .38$

- | | |
|-----------|-------------|
| a. Ottawa | c. Montreal |
| b. Quebec | d. Toronto |

24. Who is the present Illinois Secretary of State?

13 3.8 $17/19 = .89$ $8/16 = .50$

- | | |
|--------------------|---------------------|
| a. Neil Hartigan | c. Alan Dixon |
| b. Michael Howlett | d. Lewis Carpentier |

25. What Spaniard extensively explored the U.S. Southwest?

1 4.1 $3/30 = .10$ $2/17 = .12$

- | | |
|--------------------|-----------------------|
| a. Vasco de Balboa | c. Hernando de Soto |
| b. Hernando Cortés | d. Francisco Coronado |

26. Who is considered the inventor of the radio?

7 3.0 $7/16 = .44$ $9/25 = .36$

- | | |
|------------------------|----------------------|
| a. George Westinghouse | c. Enrico Fermi |
| b. Thomas Edison | d. Guglielmo Marconi |

27. What is the name of the monster in the English epic Beowulf?

2 3.1 $18/22 = .82$ $9/24 = .38$

- | | |
|-------------|-------------|
| a. Grendel | c. Naegling |
| b. Hrothgar | d. Breca |

28. What is the name of the longest river in the world?

22 4.2 $12/18 = .67$ $4/8 = .50$

- | | |
|-----------------|----------------------|
| a. Amazon River | c. Yangtze River |
| b. Nile River | d. Mississippi River |

29. Who is the present governor of New York?

0 3.1 $10/18 = .56$ $7/30 = .23$

a. Abraham Beame c. Louis Lefkowitz

b. Malcolm Wilson d. Hugh Carey

30. What was the name of the movie in which Sidney Poitier played a vacationing northern law officer in Mississippi?

12 3.4 $13/17 = .76$ $6/19 = .32$

a. They Call Me Mister Tibbs c. ...Tick...Tick...Tick

b. In the Heat of the Night d. The Defiant Ones

31. With what singer is the song "Bad, Bad, Leroy Brown" most closely associated?

35 3.5 $5/7 = .71$ $2/6 = .33$

a. Elton John c. James Brown

b. Jim Croce d. Neil Diamond

32. Who was Abraham Lincoln's immediate successor as U.S. president?

15 3.2 $4/13 = .31$ $6/20 = .30$

a. Andrew Johnson c. James Buchanan

b. Ulysses S. Grant d. Rutherford B. Hayes

33. What was the name of the first British settlement in the U.S.?

19 3.6 $7/15 = .47$ $8/14 = .57$

a. Roanoke c. Massachusetts Bay

b. Plymouth d. Jamestown

34. Who wrote "The Rhyme of the Ancient Mariner"?

1 2.6 $3/13 = .23$ $2/34 = .06$

a. William Wordsworth c. Walter Scott

b. Samuel Coleridge d. Thomas Gray

35. Who was the Chicago mayor immediately before Richard Daley?

0 2.7 $5/16 = .31$ $3/32 = .09$

- | | |
|-----------------|---------------------|
| a. Anton Cermak | c. William Thompson |
| b. Daniel Ryan | d. Martin Kennelly |

36. Who is famous for the painting "Water Lilies"?

8 1.8 $2/2 = 1.00$ $23/38 = .61$

- | | |
|--------------------------|------------------------|
| a. Pierre-Auguste Renoir | c. Claude Monet |
| b. Paul Cezanne | d. Jean-Auguste Ingres |

37. Who is credited with the discovery of penicillin?

4 3.9 $13/28 = .46$ $5/16 = .31$

- | | |
|----------------------|----------------------|
| a. Marie Curie | c. Alexander Fleming |
| b. August Wassermann | d. Louis Pasteur |

38. Who was the first European to reach India by water?

4 3.1 $7/20 = .35$ $10/24 = .42$

- | | |
|-----------------------|------------------|
| a. Bartholomeo Diaz | c. Vasco da Gama |
| b. Ferdinand Magellan | d. Prince Henry |

39. Who wrote "Ode on a Grecian Urn"?

3 2.7 $6/15 = .40$ $12/30 = .40$

- | | |
|-------------------------|-------------------------|
| a. Percy Bysshe Shelley | c. John Keats |
| b. Lord Byron | d. Alfred Lord Tennyson |

40. Who was the painter of "Blue Boy"?

1 3.0 $5/19 = .26$ $9/28 = .32$

- | | |
|------------------------|------------------------|
| a. Frederic Remington | c. Joshua Reynolds |
| b. John Singer Sargent | d. Thomas Gainsborough |

41. Whose ghost returns to haunt Macbeth?

1 3.8 $9/27 = .33$ $5/20 = .25$

- | | |
|------------|------------|
| a. Macduff | c. Duncan |
| b. Banquo | d. Fleance |

42. "Born to be Wild" was popularized by what musical group?

17 3.5 $8/19 = .42$ $5/12 = .42$

- | | |
|-----------------|-------------------|
| a. Led Zeppelin | c. Guess Who |
| b. Steppenwolf | d. Rolling Stones |

43. Who founded the Standard Oil Company?

27 2.4 $5/5 = 1.00$ $10/16 = .62$

- | | |
|-------------------------|-----------------------|
| a. John D. Rockefeller | c. J. Pierpont Morgan |
| b. Cornelius Vanderbilt | d. Edward Harriman |

44. Who was voted the most valuable baseball player last year in the National League?

5 3.2 $3/19 = .16$ $4/24 = .17$

- | | |
|-----------------|-----------------|
| a. Steve Garvey | c. Hank Aaron |
| b. Lou Brock | d. Johnny Bench |

45. What is the name of the tallest mountain in the continental U.S. (excluding Alaska)?

1 3.6 $17/26 = .65$ $4/21 = .19$

- | | |
|----------------|---------------|
| a. Mt. Whitney | c. Mt. Evans |
| b. Mt. Rainier | d. Mt. Shasta |

46. Who sculpted "The Thinker"?

3 2.6 $5/12 = .42$ $6/33 = .18$

- | | |
|----------------------------|------------------------|
| a. Auguste Rodin | c. Leonardo da Vinci |
| b. Michelangelo Buonarroti | d. Gianlorenzo Bernini |

47. Who was Richard Nixon's first Secretary of State?

1 3.0 $3/21 = .14$ $10/26 = .38$

a. Clark Clifford c. William Rogers

b. Henry Cabot Lodge d. Dean Rusk

48. Who wrote Gulliver's Travels?

9 3.7 $21/23 = .91$ $11/16 = .69$

a. Samuel Johnson c. Jonathan Swift

b. Charles Lamb d. James Boswell

49. Who discovered the neutron as a separate entity?

0 2.5 $4/10 = .40$ $11/38 = .29$

a. John Dalton c. James Maxwell

b. James Chadwick d. Ernest Rutherford

50. What is the name of the first full scale battle of the Civil War
(i.e., not Fort Sumter)?

5 3.8 $7/15 = .47$ $7/28 = .25$

a. Bull Run c. Gettysburg

b. Antietam d. Fredericksburg

51. Who was the only U.S. Vice-President to resign from office before
S. Agnew?

1 2.8 $5/14 = .36$ $14/33 = .42$

a. John Calhoun c. James Knox Polk

b. Daniel Webster d. Henry Clay

52. What state is called the "Lone Star State"?

40 3.8 $5/5 = 1.00$ $3/3 = 1.00$

a. California c. Maine

b. Texas d. Vermont

Answers to the general-information questions.

1. b	14. d	27. a	40. d
2. b	15. d	28. b	41. b
3. d	16. a	29. d	42. b
4. b	17. d	30. b	43. a
5. a	18. c	31. b	44. a
6. c	19. c	32. a	45. a
7. b	20. d	33. d	46. a
8. c	21. c	34. b	47. c
9. d	22. a	35. d	48. c
10. c	23. a	36. c	49. b
11. d	24. b	37. c	50. a
12. a	25. d	38. c	51. a
13. b	26. d	39. c	52. b

APPROVAL SHEET

The dissertation submitted by Stanley J. Pasko has been read and approved by the following Committee:

Dr. Eugene B. Zechmeister, Chairman
Associate Professor, Psychology, Loyola

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

Dr. John R. Crocker, S.J.
Associate Professor, Psychology, Loyola

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

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

11-1-76
Date

Eugene B. Zechmeister
Director's Signature