DOCUMENT RESUME

ED 159 211 TM 007 517

AUTHOR Andre, Thomas
TITLE On Productive Knowledge and Levels of Questions.
PUB DATE [77]
NOTE 77p.

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS Advance Organizers; Cognitive Processes; *Complexity Level; Concept Formation; Concept Teaching; *Learning Processes; *Learning Theories; Literature Reviews; *Memory; *Questioning Techniques; Retention; *Taxonomy; Teaching Methods

ABSTRACT

A model is proposed for memory that stresses a distinction between episodic memory for encoded personal experience and semantic memory for abstractors and generalizations. Basically, the model holds that questions influence the nature of memory representations formed during instruction, and that memory representation controls the way in which stored knowledge can be used. The paper is divided into three major sections. In the first section various proposed typologies of questions are described, to identify the dimensions along whcih questions may vary and to describe both the utility and inadequacies of current typologies. It is argued that an instructionally useful taxonomy must be based upon consideration of the empirical effects of questions, and that development of such a taxonomy is a fundamental goal of research on levels of questions. In the second section the limited research on levels of questions is reviewed, to integrate what is known and not known about the effects of asking different types of questions during instruction. Both deficiencies in the existing studies and promising directions for future research are indicated. The third section briefly presents the information-processing model of knowledge acquisition and attempts to relate question-level effects to the model. (Author/CTM)

On Productive Knowledge and Levels of Questions

Thomas Andre

Iowa State University

Send Proofs to: Thomas Andre

Department of Psychology

Iowa State University

Ames, Iowa 50011

On Productive Knowledge and Levels of Questions

Abstract

The paper is concerned with the effects of asking students questions at different levels of cognitive complexity during learning. The first section reviewed taxonomies of question levels and pointed out their deficiencies as research tools; the second section reviewed research on levels of questions. It was shown that higher level questions can have facilitative effects on both reproductive and productive knowledge, but that the conditions under which such facilitation occurs are not well understood. The third section outlines an information processing model of human cognition that can account for question level effects and that serves to integrate previous research on question level and to to provide direction for future research in this area.

On Productive Knowledge and Levels of Questions

Thomas Andre

Iowa State University

Educators clearly believe the types of questions teachers ask are important. The first series of lectures on teaching methods offered at Cambridge devoted considerable attention to the art of examining (Fitch, 1879). Most teaching method books since have discussed questioning techniques (e.g., Burton, 1962; Holley, 1923; Hough, et al., 1970; Lancelot, 1929; Miller, 1922; Otto, et al., 1960; Ruediger, 1932; Shipley, 1972; Thomas, 1927). Teacher training programs dutifully introduce prospective teachers to the varieties and vagaries of question-asking and exhort teachers to ask questions that require thinking. Recent critics of American Education have condemned the question procedures of the typical teacher (Holt, 1964).

Despite a century of proletyzing for higher-level questions, the effects of questions on student learning are not well known. Empirical research examining the effects of different types of questions on achievement has been rare. The conditions under which questions influence learning and retention remain unexamined and unspecified. Although various typologies have been proposed, no general taxonomy of questions that meets criteria of inclusiveness and operationality has been widely accepted. Given this state of affairs, educational claims for the goodness of higher-level questions remain wish-fulfilling myths.

It might be argued that current psychological ideas about the level of depth of processing of information (Anderson, 1970; Craik & Lockhart, 1972) provide empirical support for a level-of-question effect. Clearly research on depth of processing has shown that, within the types of processing that have been varied, a deeper level of processing produces better retention.

However, such research provides at best circumstantial evidence for a level of question effect. In all of the level of processing literature, comparisons have been made between very primitive levels of processing and processing information for meaning. Learning in subjects who are asked to rate sentences for pronounciability or imagery might be compared, for example. In general, this research has found that the more the information is processed for meaning the better it is retained. But the verbal learning-memory research has not compared different levels of semantic processing and it is this latter comparison that is central to the levels of questions effect as virtually all questions that would be educationally relevant would involve some processing of meaning. The educators' belief that different levels of questions differentially influence learning is based on the belief that processing more-or-less deeply for meaning will produce differential effects on learning and retention. This latter issue has not been examined in the verbal learning-memory research.

The overriding purpose of this paper is to deal with the question: What are the effects on learning of different levels of questions asked during instruction? From this perspective the question is seen as a potential instructional tool that can be manipulated by a teacher or instructional designer in order to produce certain learning outcomes. The paper is conceptualized as a status report on what is known about questions as an instructional tool. It is hoped that the paper will provide an overview of current knowledge and a set of guidelines for future research on the instructional use of questions.

The paper is divided into three major sections. In the first section various proposed typologies of questions are described. The purpose is to identify the dimensions along which questions may vary and to describe both the utility and inadequacies of current typologies. It is argued that an instructionally useful taxonomy must be based upon consideration of the empirical effects of questions and that development of such a taxonomy is a fundamental goal of research on levels of questions.

In the second section the limited research on levels of questions is reviewed. The intention is to integrate what is known and not known about the effects of asking different types of questions during instruction. Both deficiencies in the existing studies and promising directions for future research are indicated.

Finally, the third section briefly presents an information-processing model of knowledge acquisition and attempts to relate question-level effects to the model. Basically the model holds that questions influence the nature of the memory representations formed during instruction and that the memory representation controls the way in which stored knowledge can be used.

Some Definitions

As used in the present context, a question is a direction to a learner to examine instructional material or his memory of it and to produce some response. Under this conception, both the statements, "Memorize this poem and recite it to me from memory." and, "What is the capital of Iowa?" would be considered questions. This conception of questions is necessary because directive statements (imperatives) and formal interrogatory sentences refer to equivalent cognitive and behavioral activities. There is little difference, for example, in asking "Who killed Cock Robin?" and saying "Memorize the name of Cock Robin's assassin and tell me it when I say Cock Robin." The latter is not a formal question but would be included within the subject matter of this review. The questions considered herein are relatively limited in scope. Thus differences between inquiry, discovery, and expository teaching are not considered in this paper. Rather the paper is concerned with the instructional effects of asking students to make relatively short responses in the process of acquiring new information, skills, and knowledge.

The terms cognition, cognitive process, mental process, and thought process are used somewhat loosely and synonymously. They refer to the set of mental steps (i.e., program) that must be performed in order to make a response. Thus they represent loose terms for the mental operations a human carries out in answering questions, or engaging in cognition.

Level of question refers to the nature of cognitive processing required to answer a question. A question may ask a learner to repeat or recognize some information exactly as it was presented in instruction. Such a question is typically referred to as a knowledge, factual, or verbatim question. Factual questions are believed to involve less complex cognitive processing than questions requiring more than direct memory (Briggs & Reed, 1943; English, Welborn & Killian, 1934; Ausubel, 1963; Anderson, 1972). Questions that require more than simple direct memory are believed to involve more complex cognitive processing. Describing the nature of such more-than-memory (higher-level) questions has been difficult, although a variety of question classification schemes have been proposed.

Taxonomies of Questions

Substance Questions

In one early classification scheme, Briggs and Reed (1943) and English, et al (1934) distinguished between substance and factual questions. Factual questions could be answered on the basis of one text sentence; substance questions required a learner to combine information from more than one sentence of the instruction. While this scheme had the advantage of objectivity, in practice it failed to discriminate high and low level questions. Given the text, "The ball is red. The ball is round." The question, "What is red and round?" would be classified as substance even though it intuitively seems fairly factual. The problem with Briggs and Reed's and English, et al's scheme is that separate nominal sentences may be combined into one underlying proposition when they are read. Kintsch (1974) and Greeno and Noreen (1974) have presented evidence that mature readers normally combine sentences this way.

Meaningful Questions

Ausubel (1963, 1968) distinguished between rote and meaningful learning. Rote learning involved memorization of material in a verbatim form. No attempt was made to relate the new information to previous experiences. Meaningful learning involved non-arbitrarily relating to-be-learned information to previous knowledge. Knowledge was contained in a hierarchically organized cognitive structure; new knowledge was meaningfully learned when its place in the cognitive structure was determined and it was subsumed into the structure. Unfortunately, Ausubel failed to provide objective guidelines for the generation of meaningful or rote questions. In fact, in much of his research on supposedly meaningful learning, the questions used seem to have been primarily factual (Ausubel & Blake, 1958; Ausubel, Robbins, & Blake, 1957; Ausubel, Stager & Gaite, 1968; 1969).

Taxonomy of Educational Objectives: Cognitive Domain

A variety of more complex classification schemes have been proposed. The best known is probably the Taxonomy of Educational Objectives: Cognitive Domain, developed by Bloom and his associates (Bloom, Engelhart, Furst, Hill, and Krathwohl, 1956). This Taxonomy described six general classes of behavior that might result from instruction. The six classes were:

Knowledge: This class esentially involved repetition of information in the form it was presented. This kind of behavior is called for in what I have labeled factual questions.

Comprehension: This class essentially involved recognition or production of some paraphrase of material presented in instruction.

Application: This class involved use of presented information in some new situation. It could include recognizing new examples of a concept or using a principle in a problem solving situation.

Analysis: Behaviors in this class involve taking a given situation and decomposing that situation into its component parts and their relationships. Typically this involves using some previously taught scheme to decompose the whole. For example, the student may be told to analyze a short story. He would produce descriptions of the main characters, minor characters, plot, theme, and literary style.

Synthesis: Behaviors in this class involve production of some product given appropriate elements. Writing a short story might be an example.

Evaluation: This category of behavior involves judgments about the value of information, concepts, ideas, relative to some goal or purpose.

Stated in such general terms, the classes may not seem particularly useful. However, Bloom _et al_ identified subclasses within each general class and gave examples of each type of behavior. The comprehension class, for example, was divided into: translation of presented material, interpretation of presented material, and extrapolation from presented material. However, there was no empirical verification of the categories and no attempt to demonstrate that the cognitive processes presumed to operate at each level were psychologically as well as logically distinct.

The taxonomic levels were assumed to be hierarchical; cognitions at the higher levels presupposed the ability to carry out lower level cognitions; one could not _comprehend_ specific content unless one _knew_ it, etc. The levels of the taxonomy varied only along a dimension of complexity. However, Madaus, Woods, and Nuttall (1973) examined the hierarchical nature of the Taxonomy and concluded that the hierarchical structure was questionable and that performance on each of the levels strongly reflected general ability.

Despite their detailed elaboration and wide use, the Bloom _et al_ Taxonomy, and others based on it, are not satisfactory classification schemes. One major problem is that the distinction between the categories tends to become blurred in practice. Consider the following: a student is given a verbal description in which a cat is given a food pellet every fourth time the cat presses a lever. He is asked a multiple-choice question about the schedule of reinforcement. The correct answer is fixed ratio schedule. This question might be classified at the application level as it is asking the student to recognize an example of a presented concept. However, if the example was presented to the student during instruction, then the question is factual or at the knowledge level in the Taxonomy. Bloom et al (1956) recognized this relationship between the level of a question and the instruction material, but many users of the Taxonomy have not. More seriously, assuming an unfamiliar example, the questions above could also conceivably be classified at the analysis level. Is not the student asked to identify the elements of the reinforcement situation and their relationships to determine the nature of reinforcement? Many questions so defy unique classification within the Taxonomy.

The basic problem with Taxonomy-like schemes is that the procedures for assigning questions to a level are not sufficiently detailed to be entirely objective. Assignment of a question to a level very often requires a professional judgment, and there is little evidence on the reliability of such judgments. Evidence from classroom observation studies suggests that interjudge reliabilities tend to be moderate (Dunkin & Biddle, 1974). However, such studies have examined only the reliability of classification of questions that teachers actually ask, not the classification reliability of all possible questions. Since teachers' questions are predominately factual the reliabilities may not be representative for classifications of higher level questions. One sequence of this lack of detailed rules for assigning questions is the di iculty in writing questions to match a particular level. No operational, that unambiguous and mechanical, rules exist that would allow researchers (and teachers!) to write questions for a given level.

Structure of Intellect

The problem of classifying questions is equivalent to the general problem of classifying cognitive behavior. Any taxonomy of cognition can be used to classify questions. One influential taxonomy of cognitive behaviors is described by Guilford's Structure of Intellect model (1966, 1967). Guilford's model differs from the Bloom Taxonomy in two major respects. First the SOI model postulates that cognitions differ in three dimensions (contents, products, and operations) instead of the single dimension of complexity. Secondly, even though logically derived, the model enjoys some empirical support from the factor analytic studies of Guilford and others (Guilford, 1967). According to the model, an intellectual activity or cognition can be carried out on one of four contents: figural (imaginal, spatial), symbolic, semantic, or behavioral. The activity can involve one of five types of operations (or processes): cognition, memory, divergent production, convergent production, and evaluation, and can result in one of six types of end products: units, classes, relations, systems, transformations and implications. The resultant three dimensional array yields 120 hypothesized mental abilities. Since the Guilford model is well known, its further elaboration is not necessary here. From the present perspective the value of the SOI model is that one could presumably write questions that would engage a given mental ability. Thus, the model could serve as a question taxonomy.

Aschner and Gallagher (Aschner, et al 1965; Gallagher, 1965) attempted to develop a classification system based on Guilford's model and to use it to study teacher's in-class questions. Gallagher (1965) reported considerable difficulty in classifying questions reliably within the categories and Dunkin and Biddle (1974) characterized the reliability of question classification instruments based on Guilford's scheme as weaker than those based on Bloom's. One practical difficulty may be that questions typically asked by teachers call into play several mental operations and abilities. If so, unique classification of questions into one of the 120 categories would be impossible. Thus, like the Bloom Taxonomy, no scheme based on Guilford's model unambiguously maps questions to categories.

Conditions of Learning Typology

Gagne's hierarchy of types of learning (Gagne, 1970) has relevance for the problem of classifying questions. Gagne's scheme distinguished between eight types of learning: signal learning, stimulus-response learning, chaining verbal association, discrimination learning, concept learning, principle learning, and problem-solving. These types of learning differ in the conditions necessary for them to occur. Such conditions were identified by examining the experimental paradigms in which the various types of learning had been studied. Since Gagne's scheme is well known it will not be described in greater detail here.

Hypothetically, it would be possible to write questions to assess different types of learning. Clearly, it would be possible to write a question to assess a learner's ability to make a discrimination or to recognize a new instance of a concept. Thus, a question taxonomy scheme based on Gagne's model would have certain advantages. The fact that levels in Gagne's model are defined in terms of the environmental conditions of learning might be helpful in developing objective criteria in classifying questions. In addition, it would be possible to relate instructional questions to the expected outcomes of the instructional system, since the Gagne model presumably encompasses the universe of cognitive instructional outcomes. Some disadvantages of the Gagne model are that some types of questions that do not readily fit within the model, e.g., questions Bloom would call evaluative for example. Also it does not seem possible to write questions for the simplest levels in Gagne's hierarchy.

Although the author is unaware of any classification scheme that makes use of Gagne's full model, a classification scheme proposed by Anderson (1972) attempts to operationalize at least two of Gagne's categories. Anderson (1972) discussed the problems of measuring student comprehension and presented a possible classification scheme of items used to assess comprehension. Anderson argued that the first factor to be considered in determining the classification of questions was the relationship between the question and the presented instructional material. Questions could be presented in either the same lexical form as in the original instruction (verbatim questions) or in a changed lexical form (paraphrased questions). Anderson argued that verbatim questions could often be answered without the student understanding the material. Verbatim questions correspond to what I have called factual questions and would correspond to Gagne's verbal association learning. Paraphrased questions involved changing the content words of the instruction in presenting the question. Bormuth (1970) made a similar distinction. Anderson clearly recognized that the verbatim-paraphrased distinction formed a continuum, any given question has near and distant paraphrases. Anderson argued, however, that a correct answer to any paraphrased question ensures that the student has processed the meaning of the presented information. In Anderson's words, answering a paraphrased question presupposes "semantic encoding."

Anderson further argued that for concepts and principles it was possible to write application questions. Such application questions required the student to recognize unfamiliar new examples of concepts or to use the principle to solve an unfamiliar problem. Only if a student could answer such questions could he be said to understand the concept or principle (Anderson, 1972; Anderson & Faust, 1973, Chap. 10). Gagne (1970) presents similar arguments.

Within its limited universe of applicability, Anderson's classification scheme seems to provide an objective means of categorizing questions. It is certainly possible to objectively determine whether content words are repeated both in instruction and in questions. Similarly, it is possible to tell if an example or problem used in an application question was also used in the instructional materials. Estimates of the closeness of instructional and test-item examples can be empirically determined (e.g., Anderson, 1973). This makes Anderson's scheme very useful for many research purposes. The major fault of Anderson's scheme is that it does not encompass many of the seemingly wide varieties of questions. Questions at Bloom's et al analytical synthetic, or evaluative levels are not included, for example. Certainly these levels appear to represent important behaviors that might result from instruction. In fairness, it should be noted that Anderson's scheme was developed in the context of a paper describing means for assessing comprehension of concepts and principles, not as a universal scheme for representing instructional goals or questions. Its lack of breadth does not obscure its valuable contribution.

Towards an Adequate Taxonomy

This brief overview of question classification schemes should make it sufficiently clear that students can be asked to do a bewildering variety of things with instructional materials. This wide variety makes it difficult to agree on any particular organization scheme or taxonomy. Thus, while a wide variety of question taxonomies have been proposed, all seem to suffer deficiencies of either objectivity or breadth. As will be seen in the next section, this diversity of taxonomies has led to little continuity between various empirical studies of question level.

In general, question taxonomies have been developed on a logical a priori basis. The taxonomist has postulated one or more dimensions along which questions vary and then attempted to develop descriptive categories within those dimensions. Different taxonomists have emphasized different dimensions and thereby produced different taxonomies. Such taxonomies have been characterized by the use of subjective judgment in the classification of questions and have typically in practice been at best moderately reliable means for classifying questions (Dunkin & Biddle, 1974).

Until some degree of uniformity and agreement is reached on a taxonomy of questions, research on the instructional effects of questions will remain fragmented. The development of an agreed upon taxonomy is extremely important and has impact beyond the mere description of questions. An adequate taxonomy of questions would also serve as a taxonomy of learning outcomes and would provide a mechanism for assessing various outcomes of instruction.

One alternative approach to developing a useful taxonomy of questions might be to attempt to identify as many possible dimensions along which questions may vary and then to use empirical methods to form categories or clusters of questions. The current overview of question taxonomies suggests a number of such dimensions: complexity of processing; nature of product, operation, or content; type of learning involved; and, perhaps, subject matter. A set of questions differing along such dimensions could be developed and numerical taxonomy and clustering methods used to form groups of questions. Such groups could then be used to determine if they in fact produce varying instructional outcomes. That is, the instructional reality of the taxonomic divisions could be assessed. Obviously this suggestion envisions a long-term program of parametric research on questions. Such a program would be both expensive to carry out and frought with practical problems. However, until such research is available, an adequate question taxonomy and an adequate description of instructional outcomes is not possible.

What would an adequate taxonomy of questions consist of and provide? Such a taxonomy would have three major features. First, it would provide for the relatively objective classification of questions and would provide an objective mechanism by which questions of various types could be constructed. By "objective" is meant that the procedures for classifying questions would be as mechanical as possible and would involve a minimum of judgment. Ideally the procedure for classifying questions (or constructing questions) could be computerized. In the absence of such automation, the interjudge reliabilities for the classification of questions should be very high. Inherent in this characteristic is the necessity for the taxonomy as a whole to be inclusive, while the individual categories should be mutually exclusive. Any given instructional by relevant question should be uniquely categorized within the system. Given the ambiguities and complexities of English, it may be that some questions would need to be decomposed into subquestions which could be uniquely classified or that context must be used as a clue to the nature of the questions. Ideally, the procedures for decomposing questions or for utilizing context would also be mechanical.

Secondly, in an adequate taxonomy the categories of questions would be shown to have psychological and instructional validity. Empirical evidence would be available that the psychological processes involved in answering the questions and/or the instructional outcomes of asking the question during learning would differ from other categories of questions. Unless some psychological or instructional difference between categories can be observed, classification of questions into different categories is meaningless. One consequence of classifying questions by their effects would be that the taxonomy would be instructionally useful. Such a taxonomy would indicate what sorts of instructional outcomes might occur if various types of questions were used during instruction. The taxonomy should indicate the probable instructional consequences of asking varying types of questions during instruction.

Finally, an adequate taxonomy would indicate how questions relate to other instructionally relevant variables. Such other variables might include individual difference variables such as age, cognitive style, intelligence, etc.; and instructionally manipulable variables such as feedback, pacing, reinforcement, media, etc. As will be indicated in the subsequent section, there is evidence that the effect of questions are modified by such variables.

Currently available taxonomies fall far short of this description of an adequate taxonomy. Taxonomies such as Bloom's, Guilford's and their deriatives are best regarded as heuristic in value. They remain sources of ideas for ways in which questions may vary and a starting point for research on question taxonomies and the effects of questions on learning. The subsequent section describes currently available research on the effects of questions. The majority of such research has used the taxonomies described above as the starting point of the investigations.

Empirical Studies of Question Level

Questions may be used in at least four different situations to guide student learning. In one situation questions are used in classroom recitation or discussion. A second use involves questions inserted in text or other instructional media. A third use, in some ways similar to the second, involves the questions used on examainations. Finally a fourth situation involves questions students ask of themselves while studying.

The literature on oral recitation questions has received a number of excellent reviews recently (Gall, 1970; Dunkin & Biddle, 1974; Rosenschine, Note 8). In general these reviews have concluded that the degree to which teachers use higher-level questions has little demonstrated relationship to student achievement. While this conclusion is discouraging to those who believe in efficacy of higher-level questions, its impact is tempered by methodological and conceptual problems which beset such observational research. One serious problem is that, as noted above, the reliabilities with which teacher's questions can be classified are at best moderate (Dunkin & Biddle, 1974, Chap. 8). A second problem is that teacher's oral questions are typically answered by only one student at a time. Anderson and Faust (1973, Chap. 6) have argued that a requirement for a student to make a response will not have an instructional effect unless the student actually makes a response. To the extent this generalization is valid, a question in an oral recitation situation provides benefit to only one pupil (perhaps with some additional benefit to those students who respond covertly). A related problem is that the classroom research has not typically separated the effects of questions on the acquisition of material covered in the questions and material not covered by questions. As the research below will indicate, the effect of higher-level questions seems to be specific to the concepts, principles or ideas covered by the questions.

These problems lead me to conclude that observational research on teachers in class questions is not an appropriate vehicle for systematically examining the effects of higher-level questions on student learning. For this reason, and also because of the quality of the recent reviews, this line of research is not considered further here.

Also not discussed is research on the effect of student generated questions or examination questions. These areas are excluded because research is not available, although Frase and Schwartz (1975) have initiated some interesting research on student generated questions.

The present paper focuses on experimental studies which use the adjunct questions technique popularized by Rothkopf (1966). The adjunct questions technique involves placing questions either in, before, or immediately after prose passages and asking students to answer such questions while studying the passage. Later students are given a posttest which may repeat the adjunct questions and may also include new questions. When the adjunct questions are interspersed in the text, they are often called <u>inserted questions</u>; I will adopt that usage here. No special term has evolved for <u>adjunct questions</u> massed either before or after a passage. Rothkopf was not the first to use adjunct questions, but most contemporary research on questions derives from his 1963 and 1966 papers. Earlier work by Germane (1920), Distad (1927), Washbourne (1929) and Holmes (1931) did not generate sustained interest and in fact became generally known only after current interest in Rothkopf's work had bloomed.

Rothkopf's seminal studies have engendered a large number of experiments examining adjunct questions. Anderson and Biddle (1975) provide an excellent review of this general literature. A less detailed review is provided in Faw and Waller (1976). Relatively few of the adjunct questions studies have examined question level; it is those studies that are reviewed here. In the typical study examining question level, students are asked to read a passage and to answer one or more questions about the passage while studying it. Subjects in different groups are given questions at different cognitive levels, the levels being defined and chosen by the experimenter. After subjects complete the passage they are given a posttest on the passage. In various studies this posttest has contained items asking for factual recall, items asking for higher-level processing, or both. Comparison of performance on the posttest has been the variable of main interest in this research.

Research using the adjunct questions technique to examine question level effects can be divided into two classes on the basis of the type of posttest employed. About half the studies employed a posttest that asked only for factual or verbatim recall of the passage. The remaining studies employed a posttest that contained higher-level questions. It is argued below that the first class of study can only yield results that are of best minor interest to instructional theory. Moreover, it seems likely that different psychological mechanisms are needed to explain question level effects in the two types of studies. For these reasons the two classes of studies are discussed separately.

Studies Using Factual Recall Tests

In an early study Rothkopf and Bisbiscos (1967) varied the nature of the content material required to answer questions. While not directly concerned with the effects of question level, the study was interesting because it demonstrated that questions could influence how the students processed the materials. Some subjects received questions that required proper names or numbers as answers; other subjects received questions that required either common words or technical terms as answers; a third and fourth group of subjects received either all types of questions or no questions respectively. All groups took a posttest containing all question types after reading. The subjects given proper name and number questions did well on that class of material on the posttest, but poorly on other material. Subjects who received common and technical word adjunct questions, or mixed adjunct questions, did well on all types of posttest questions. Rothkopf and Bisbiscos attributed their findings to the influence of the questions on the student's attention to the material. Subjects who received name and number adjunct questions could limit their attention to material of that type. Subjects who received mixed question types had to attent to all types of material in the passage, since common word questions is a broadly defined class.

Frase (1968) demonstrated a similar effect of questions on attention. Frase varied the specificity of the questions. Consider the following taken from Watts (1973).

> John is a painter and is 25 years old. Bill is a carpenter and is 64 years old. Sam is a mechanic and is 30 years old. Ed is a writer and is 70 years old.

The question "How old is John?" is more specific than "How old were the men in the story?" Comparing specific to more general questions, Frase found that the broader questions led to poorer posttest performance; a result seemingly at odds with the Rothkopf and Bisbiscos findings. Watts (1973) resolved the apparent paradox by demonstrating that Frase's broader questions did not require the student to attend to the associative relationships that were tested on the posttest. On the posttest, Frase asked questions like "How old was John? How old was Sam?" Subjects given the broader posttest questions did not have to relate names to ages during learning and thus did poorly on the posttest. When Watts (1973) gave fifth graders broad questions that forced them to attend to posttest relevant relationships, achievement was greater than in a group given simple factual questions on the material. Watts (1973) used broader questions like "Which man is too old for his job?" Such a question clearly seems higher level since it requires the student to make inferences and to compare the ages of the men to expectations about the amount of labor each job requires and the physical condition of the men. (Bloom et al, 1956, would probably classify the question as an evaluation level question.) Unfortunately the Watts (1973) study does not allow us to properly assess the effects of these higher level questions since the posttest merely tested for factual knowledge. About all that can be said is that the higher-level questions led subjects to remember more facts from the passage.

A similar problem exists in an ingenious study by Rickards and DiVesta (1974). Subjects read passages containing a topic sentence that asserted a general quality about a fictional country called Mala and three sentences which supported that general assertion. The relationship between the topic sentence and the supporting sentences was not made explicit. The paragraph below provides an example.

> The southern area of Mala can best be described as a desert. Rainfall is less than two inches per year. The soils in the southern area of Mala are either rocky or sandy. In the summertime temperatures have been recorded as high as 135 degrees in southern Mala (Rickards & DiVesta, 1974, p. 355).

Readers of these paragraphs were either asked rote factual questions which required repetition of one of the three supporting sentences, rote idea questions which required repetition of the topic sentence, or meaningful learning questions which required the use of the supporting sentences to justify the general assertion. An example of the latter type of question is "Why can it be said that southern Mala is a desert?" A fourth group received inserted questions that were irrelevant to the passage. When questions were inserted after every two paragraphs, meaningful learning questions led to superior performance; however, this effect disappeared when the questions were asked after every four paragraphs. Performance in this case refers to verbatim recall of the passage. Subjects given meaningful learning questions recalled more. This result is congruent with Frase (1968) and Watts (1973), since in order to answer the meaningful learning question the subject would have had to attend to more of the passage, having attended to more, he recalled more. In a series of studies using similar methodology, Richards and his associates have confirmed these basic results (Richards, 1976a, 1976b; Richards & Hatcher, 1976, Note 7).

Allen (1970) also reported an apparent attentional effect. Allen compared groups given either memory level or higher level questions during reading. His posttest apparently consisted of factual questions dealing either with material refered to in the memory level or higher level adjunct questions. The basic result was that subjects did better on questions related to the information they were asked about in the adjunct questions. This result is consistent with the hypothesis that questions serve to focus the students' attention on particular material. Like the Frase (1968) and Watts (1973) studies, neither the Richards and DiVesta (1974) nor Allen (1970) studies demonstrate any effect for higher level questions beyond directing student's attention to more of the information.

A series of studies by Frase and his co-workers made this directed attention hypothesis explicit. Frase (1969a,b, 1970a,b, 1972) had people read passages that asserted a series of class inclusions. The structure of the passages was: A is a member of B, B is a member of C, C is a member of D, D is a member of E. Heading each passage was an assertion that could either be proven or disproven from the information in the passage. The subjects were told to read the passage and either verify or disprove the assertion. Later they were asked to recall the passage and to draw any conclusions they could from the passage.

Frase varied the level of the to-be-verified assertion. An assertion given in text was the lowest level (Is A a member of B?), the next highest level involved relating two classes separated by one class (Is A a member of C?) and so on. Frase predicted that factual recall of the passage would increase as the level of to-be-verified assertion increased, but that the sentences recalled would be those needed to verify the given assertion. Over the series of studies this hypothesis was confirmed.

Frase had also hoped that verifying higher level assertions while studying would lead readers to draw such inferences during testing. This prediction was weakly confirmed. However, the overall level of inference drawing during recall was quite low across the series of studies, never exceeding 15% of the possible inferences. Additionally, the inferences drawn during recall tended to be those needed to verify the adjunct assertion. Similarly, Frase and Silbiger (1970) led subjects to read a greater or lesser number of text sentences to solve an adjunct problem. Subjects who read more of the sentences remembered more. However their increased retention was specific to the sentences needed to solve the problem.

Two methodological problems may have influenced Frase's results. One problem involved the directions given to readers on the recall tests. Frase simply told subjects to draw as many inferences as they could from the material. Assume a subject in Frase's experiments wrote: As are Bs. Bs are Cs. The inference that As are Cs is obvious, so obvious that it may have not been worth the subjects' effort to write it in the absence of clear directions to do so.

In addition, the specific class-inclusionary relationships used by Frase were not made clear in his passages. Subjects could have interpreted the relationships as equivalence instead of class-inclusionary relationships. For example, one passage asserted that the Fundalas were outcasts. What was meant was the Fundalas were one of several tribes of outcasts, but it would be easy to interpret the passage as asserting that Fundalas is the name given to all outcasts. Griggs (1974) provides evidence that subjects did in fact make such interpretations of materials like Frase's. To the extent that subjects did assume equivalence relationships, Frase's scoring procedure underestimated the degree of inference drawing.

While the extent to which these methodological problems may have influenced Frase's data cannot be ascertained, it should be noted that other researchers have found that subjects have difficulty correctly drawing the kinds of class inclusion implications Frase required (Griggs, 1974). Thus, the low level of performance in Frase study may simply reflect the inherent difficulty of the task.

In any case, Frase's data strongly support the contention that one effect of higher level adjunct questions is to direct students to attend to more of the material and, thus, to recall information directly related to the information needed to answer the questions. Such an effect of questions may be labeled the <u>directed attention effect</u>. Underlying the directed attention effect is the principle of least effort. As applied in the present case, the principle of least effort suggests that when students are confronted with more information than they can comfortably assimilate, they will selectively attend to those aspects of the information needed to complete their perceived task of 'getting through' the material. Adjunct aids such as inserted questions, serve to direct the students attention by altering the nature of his perceived task. This explanation of the effects of questions may be called the <u>directed attention model</u> (DAM).

DAM makes it possible to relate the question level literature discussed to this point to the general literature on adjunct instructional aids. A variety of such aids have been shown to have effects that can be accounted for by DAM. For example, Anderson & Faust, (1967) and Anderson, Faust & Roderick, (1968) demonstrated that a formal prompt in programmed instruction could lead students not to attend to associative relationships that were tested on the posttest. Subjects who received the formal prompt performed more poorly than subjects who did not. This result is similar to the Frase (1968) result involving breath of questions. Like Frase's questions, Anderson and Faust's prompts served to alter the subject's perception of his task and limit the material to which he attended. Similarly Anderson, Kulhavy, and Andre (1971) found that when subjects could examine feedback in programmed instruction prior to reading the frames, they tended to simply copy answers into the blanks and not read the frames. The students changed their task from one of reading frames and generating answers to one of reading feedback and copying it; the presence of feedback directed student attention away from the relevant instructional experience.

By influencing attention, adjunct instructional aids can facilitate performance as well. The positive effects of highlighting, underlining, and notetaking on factual learning are easily handled by DAM, as are the facilitative effects of behavioral objectives on recall of objective-related information. This analysis suggests that the effects of higher-order questions on learning are not unique; rather higher-order questions are simply a type of adjunct instructional aid. As with all instructional aids, questions which direct attention to relevant information enhance performance; questions which direct attention to irrelevant information degrade performance. In the research discussed above, higher-order questions have generally facilitated posttest performance. The questions have had this effect by directing attention to more information. Having attended to more, subjects receiving higher order questions recalled more.

It should be noted that the effects of instructional aids including questions are not directly causal. Questions, and other instructional aids, serve as one type of input to a complex information processing system--the reading student. DAM suggests that questions and other adjunct aids will have maximum effect when the subject perceives his task to be one of getting through the instructional materials with a minimum of effort. Under such conditions, questions influence the subject by changing what he perceives to be the minimum task. If however, the subject perceives his task to be one of learning the maximum amount from the material, questions and other adjunct aids may not facilitate his performance. If the subject chooses to attend to and learn all the information the directed attention effect of questions will not operate. For example, Royer (1969 Note 9) found that overprompting in programmed instruction did not have deletrious effects when the subjects were highly motivated and attempted to learn as much from the program as they could. This analysis suggests that questions will have a maximum effect on performance when the subject is operating under conditions of low motivation or ability (Anderson & Faust, 1973).

The Nature of Attention in DAM

The nature of the attention refered to in the directed attention model should be specified. In psychological usage, attention has many referents, e.g., orientation to particular stimuli in a field, processing of particular aspects of stimuli, even concentration and/or arousal. In the present case, attention refers to particular aspects of the reading (listening) process. Reading (listening) is conceptualized as a multi-staged process which involves translating visual (auditory) stimuli into encodings of meanings in the working memory of the reader. Stages in the process may include: a) recognizing the visual patterns of letters and/or words, translating the visual patterns to sounds, translating the sound and/or visual patterns to meaning representations, constructing structures which encode the meanings of sentences, paragraphs and passages, relating new meanings to prior knowledge, etc. Psychological controversy has raged over the precise nature of these stages (Gibson & Levin, 1975). The fine distinctions between the various theorists are not relevant here. What is relevant is that given a multistate conception of the reading (listening) process, we can define attention as used in the directed attention model, as the processing of presented information so as to form a unified mental structure that combines the elements attended to. In other words, a student has attended to a particular portion of an instructional communication when he has encoded some aspect of that portion.

This conception of attention is necessary to account for the effects of different levels of questions. To demonstrate this, compare the Frase (1968) and Watts (1973) studies discussed above. Frase found that subjects given more general questions did more poorly on the posttest. Watts (1973) demonstrated that Frase's results were produced because his more general questions led students not to attend to the particular associations needed for the posttest. Remember that the subjects were given short prose paragraphs which contained information like the names and occupations of men. It does not seem reasonable to me that Watts was saying that the subjects simply looked at only the words in the passage that refered to occupations when given a question like: What were the occupations of the men in the story? The other words in the passage had to be looked at, if for no other reason, than simply to identify which words were occupation words. By attributing the poorer performance of the general question group to directed attention, Watts was arguing that the subjects in the general question group failed to construct a memory representation, or encoding, which related the men to their occupations. Rather their memory encoding listed the occupations as a separate "chunk."

Consideration of the results of other studies of instructional aids similarly suggests that the effect of the aid is to lead subjects to construct different encodings of the information. A clear demonstration of how differences in encoding can influence subsequent test performance is provided by Andre and Sola (1976). Andre and Sola had subjects study a list of sentences. Subjects received either verbatim or paraphrased test questions on each sentence while studying. When tested with new paraphrase sentences questions on a posttest, the group receiving paraphrased study questions did better. Drawing on results reported by Green (1975), Andre and Sola argued that the group given verbatim questions tried to memorize the sentences as an independent series of words; they did not encode the words into a unified cognitive structure. The groups given paraphrased study questions were led to create a unified encoding in order to relate the sentences to the study questions. In contrast, the verbatim group apparently memorized the words in the sentences as words. Construction of a unified encoding increased performance on the posttest.

This conception of directed attention and encoding can be made more explicit. What I am arguing is that by directing the reader to particular kinds of information, the questions lead the reader to set up a strategy or program for processing the information in the text. Different questions lead to different strategies. Figures 1 and 2 illustrate the kinds of strategies that might be set up in response to the Frase (1968) and Watts (1973) type general questions. In Figure 1, the general question is: "What are the occupations of the men in this story?" This question leads to a reading strategy in which occupation becomes a key or criteria for selecting information for encoding. As the reader encounters occupational information, he takes further action upon it by storing it in an "occupation list." Figure 3 illustrates the kind of encoding that results from applying the Figure 1 strategy. In Figure 3 the men's occupations are associated to a retrieval cue of "occupation of men in story." Of course, this retrieval cue is context specific in that applies to occupations read about in a particular passage in a particular experimental situation.

Figure 2 illustrates the effects of the Watts (1973) general question: "Which man is too old for his job?" In response to this question the reader sets up a different set of criteria to select information for encoding and encodes more information. Figure 4 illustrates the kind of structure that results. In this second encoding, the retrieval cue is story characters, associated to this cue are the names of the men and associated to the names of the men are occupations and ages. When the subject is asked to give specific information about specific men, we would be able to do so given the second encoding (Figure 4), but not the first (Figure 3).

Two qualifications on this description should be noted. First, the probability with which a particular occupation, man's name, or age, etc. is encoded is less than unity. Thus the reader won't encode or remember all the information related to the question. Obviously the reader's posttest performance will be reduced by the extent to which his original encoding is incomplete. One general task for the kind of model being suggested here is to describe what the probabilities of encoding are under different instructional conditions.

A second related qualification is that the subject may be using more than one strategy while reading. In response to the inserted question, he may be keying on and encoding particular kinds of information; but because of his background and his expectations about what information is important and likely to be on a posttest, he may also be attempting to encode other kinds of information. Certainly most mature readers assess the importance of components of texts they read; the sales of highlighters to college students attest to that fact. Moreover, Duell, (1974) has shown that students do remember better what they consider important from a passage. Since DAM assumes that only a finite amount of processing capacity is available to the reader, the attempt to use multiple strategies will degrade the success of any given strategy and should lower the probability with which information related to a particular strategy is encoded.

Insert Figures 1, 2, 3, 4 about here

DAM and the Instructional Value of Studies Assessing only Verbatim Recall

If DAM provides a full explanation of the level of question effect, then question level does not seem to be an instructionally interesting variable. It is hardly surprising that information learners pay attention to (encode) is learned and retained better than information they do not attend to (encode). If a higher level question merely directed the reader's attention to more of the passage, it would be possible to simulate its effect by asking a greater number of lower level questions. Most educators and theorists who have dealt with levels of questions would not be comfortable with that notion. The general belief has been that higher order questions exert a qualitative, not quantitative effect. It has been believed that higher order questions lead the learner to acquire something else in addition to simply more.

Unfortunately the designs of studies discussed above have precluded the demonstration of anything except quantitative effects. These studies clearly fail to distinguish between the learning <u>of</u> a prose passage and learning <u>from</u> a passage. In laboratory verbal learning research, subjects are typically asked to learn arbitrary lists of nonsense or real words. In such experiments, the list represents the universe of to-be-learned material. Most prose research has treated prose passages in the same way. Researchers have treated prose passages as complex word lists; the passage and its words have become the material to be learned. The Rickards and DiVesta (1974) study provides a typical example. After ingeniously varying the type of adjunct question, Rickards and DiVesta simply required verbatim recall of the passage. This type of testing procedure makes it impossible to demonstrate anything other than a quantitative effect for higher level questions. When a verbatim recall or recognition procedure is used on the posttest, only the quantity of verbatim recall can be assessed.

While such recall procedures have some theoretical and pedagogical interest, they also represent a clear misunderstanding of the instructional use of prose. When a teacher assigns a passage he seldom regards learning of the passage as the end of instruction. Rather the teacher expects the passage will communicate important ideas that students will employ in the future. Such ideas may include concepts, principles, skills, or problem-solving techniques, etc. Regardless of the specific nature of the ideas, educators want students to learn from passages, not to memorize passages. From the educator's perspective, a prose passage is a device for altering the reader's knowledge of the world. Different levels of adjunct questions may facilitate the ease with which such alterations occur, but studies which require only verbatim recall will never reveal these effects. Such effects become apparent only when the reader must demonstrate his knowledge in a transfer situation (Anderson, 1972).

Studies Assessing More Than Verbatim Recall

Fortunately, a few studies have investigated the role of adjunct questions in producing transferable knowledge. Although only a handful in number, these studies demonstrate an effect of higher level questions which DAM cannot handle.

Anderson (1972) had argued that paraphrased questions induced a higher level of processing than their verbatim counterparts. Anderson and Biddle (1975) attempted to show that paraphrased questions would enhance learning from a prose passage. Over a series of 4 studies they were unable to show any convincing direct facilitative effect of such questions. Andre and Sola (1976) criticized Anderson and Biddle's work on two grounds. Anderson and Biddle's paraphrased adjunct questions were given only after subjects had read the passage. Used in this way the questions could not induce subjects to process the information more deeply. In addition, Anderson and Biddle's posttest was composed of questions used in instruction and, as such, did not provide a pure test of their hypothesis. As noted above, Andre and Sola (1976) were able to demonstrate that when adjunct paraphrased questions could guide semantic encoding of the presented information and when the posttest measured such semantic encoding of the information, the paraphrased adjunct questions led to greater performance.

Hunkins (1969) had 6th grade students study over a four week period social study materials containing either "knowledge" level or "evaluative" level questions. At the end of the period they took a posttest containing questions at all six of the levels in the Bloom taxonomy. The only significant main effect for treatment questions occurred for evaluative level questions--students who received evaluative level questions during instruction did better on new evaluative questions on the posttest. Type of adjunct question also entered into two three way interactions with sex and reading level. Unfortunately, it is difficult to interpret the Hunkins' results as the paucity of methodological details in his report makes it difficult to know exactly what was done. For example, it is not clear how long the experimental materials were, how many adjunct questions were used, whether the same number of factual and adjunct questions were given to the two treatment questions, whether the adjunct questions were inserted in the text near the material that provided answers for them or massed at the end of the sections, and whether students received feedback on the questions. In addition, the materials were used in ongoing classrooms, but the relationship between the experimental materials and the ongoing instruction was not made clear. Moreover, Hunkins reports initial interrater reliabilities for the assignment of adjunct questions to categories. These are quite low (mean = 52.2%) suggesting that the treatments were not reliably different. Finally, the data appear to have been analyzed incorrectly. Hunkins assigned intact classes to treatments, but used individuals as the unit of analysis. Had the proper hierarchical analysis been performed, it is unlikely any effects would have been significant. Given these problems, it is best to regard the Hunkins results as providing only weak support for a level of questions effect on higher-order learning. However, Hunkins is to be commended for attempting to assess more than simple factual learning.

Perhaps the most intriguing study to examine higher order questions is that of Watts and Anderson (1971). Watts and Anderson wrote passages which consisted of brief descriptions of psychological concepts and principles. Each of five passages contained, (1) an example of a psychological concept or principle, (2) a verbal statement of the concept or principle, (3) the name of a psychologist associated with the principle, and (4) a second example. Five questions were written for each passage. The Name questions asked the reader to identify the psychologist associated with the principle. Each of two Repeated Example questions asked the student to recognize one or the other of the given examples as an instance of the principle. Two Application questions asked readers to recognize a new, unfamiliar instance of the concept. Each reader read all five passages. After each passage appeared one of the five questions. Thus, there were five groups and subjects in each group answered either name, one or the other of the two repeated examples, or one or the other of the two application adjunct questions while studying the material. Subsequent to reading the passages, each student completed a test consisting of all 25 questions. The important result was that the groups asked Application questions during reading performed much better on new Application questions and about as well on the other types of questions as did the Name and Repeated Example groups. That is, the application questions groups were better able to transfer their knowledge of the concepts and principles to the new examples.

A study by Dapra and Felker (Note 2) supported Watts and Anderson's (1971) findings. Dapra and Felker had subjects study materials on basic conditioning concepts and principles. Subjects answered either comprehension or verbatim questions while studying. Verbatim questions demanded memory of the exact words in the text; comprehension questions required the subject to understand a paraphrased version of the message presented in the text or to apply the concept or principle. These distinctions were based on the Anderson (1972) paper. Subjects were given two posttests in the following order: a problem-solving test containing descriptions of situations to which conditioning principles could be applied and a multiple-choice posttest containing new comprehension items followed by verbatim items. The readers given comprehension adjunct questions scored higher on the problem solving test, but did not score higher on the new comprehension multiple-choice items.

Anderson and Kulhavy (1972) presented subjects with the definitions of concepts with which the subjects were unfamiliar. While studying the concepts, subjects either tried to make up a sentence using the defined concept (word) or repeated the definition aloud three items. On a posttest that asked subjects to recognize new instances of concept, subjects who used the concept in a sentence performed better than subjects who repeated the concept aloud. Anderson and Kulhavy argued that using the concepts required the subjects to semantically encode the definition and hence be able to use it.

In a long-term study, McKenzie (1972) had eighth grade students take weekly quizzes that required either recall of given facts or drawing inferences about the political interests of groups discussed in the material. After eight weeks subjects took a posttest which required recall of facts, new inferences about the groups considered in the weekly quizzes, and new inferences about groups not previously considered. Type of quiz did not influence either recall of facts or new inferences about new groups, but did influence performance on the new inferences about the old groups. Subjects given inferential quizzes did better than subjects given factual quizzes.

Moore (Note 3) reported apparently disconfirming results. Subjects in Moore's study were given either verbatim, paraphrased, or application adjunct questions while studying a passage. Control groups received no questions. [illegible] and eight days after reading the passage the subjects took tests on the material. No significant difference between groups were found. However, Moore's study was designed to test the general facilitatory effects of inserted questions predicted from his interpretation of Rothkopf's (1963, 1966) mathemagenic notions. For this reason none of the posttest questions were directly or indirectly related to the adjunct questions. Most importantly in the present context, none of the posttest questions asked students to apply the concepts and principles that had been involved in the adjunct questions. Moore's data therefore demonstrate the effect of adjunct application questions is specific to the content with which they are concerned. As Moore (Note 3) noted, such effects are still educationally important if they facilitate learning of specific concepts and principles.

McConkie, Rayner, and Wilson (1973) attempted to use different levels of questions to alter the reading strategies of subjects. Subjects read six passages on diverse topics and received 1 of 5 types of questions while reading. The 5 question types were: (1) factual questions with word answers, (2) factual questions with number answers, (3) structure questions, which involved telling the order of things in the passage or the amount of passage devoted to subtopics, (4) far-inference questions, which asked for things as the best title for a passage or the authors purpose in writing the passage, and (5) recognition questions, which asked students to pick out words and phrases from the article from among a list of distrators. Two additional groups received one of each type of question after each passage or two particular inference questions after each passage. Except for the latter group, each group received 5 questions after each passage. After completing the passage, subjects took a retention test which contained factual word, factual number, structure, far inference, and recognition items. Whether these items were the same as those used in the passage was not made clear, although McConkie et al indirectly suggest that the posttest contained both new and repeated items. Type of adjunct question influenced both reading speed and posttest performance. As McConkie et al suggest the differences in posttest performance are not easily interpretable. The far inference adjunct questions, which come closest to the kind of question we have been calling higher order, did not facilitate performance on far inference posttest questions. Since McConkie et al imply that the posttest questions and the adjunct questions referenced at least partially different information, this result may support the Moore (Note 3) findings.

Part of the problem in interpreting McConkie's et al's results is the ill-defined nature of the higher-order questions. I don't think it would be possible to even quasi-replicate this study from the information given in the report. One simply could not draw up sets of structure and far inference questions that would match the characteristics of McConkie's et al's questions. A second problem is that posttest data were analyzed using a repeated measures ANOVA in which posttest question-type is a within-subjects factor. They report a significant interaction of Adjunct question type and Posttest question type; but do not report the simple main effects of Adjunct question at each level of Posttest question type. But it is the latter effects that are of major interest. From an instructional viewpoint, it would be important to know if different types of adjunct questions make differences in instructionally meaningful kinds of outcome measures. The overall significant interaction merely indicates that there may be differences of interest within the data but does not isolate them. In the case where posttests contain instructionally meaningful distinct types of questions, separate between subjects ANOVAS on each question type would probably be more informative about the relationship in the data. If the researcher deems it necessary to use an overall analysis, an appropriate MANOVA which would yield univariate Fs would be a better choice than a repeated measures ANOVA. Doing the latter analysis without providing follow-up tests will typically produce only confusion about the data.

Shavelson, Berliner, Ravitch, and Loeding (1974) compared the effects of higher or lower order questions given before or after the passages that answered them. A control group received no questions. Higher order questions placed after the relevant material tended to produce better performance than did other question-position combinations. However the no-question control group did about as well as the higher-order questions group. Differences between conditions were not generally significant. A number of problems make clear interpretation of this study difficult. The authors indicate that the low-level questions were at the knowledge level in Bloom et al Taxonomy and that the higher-order questions were at the comprehension, application, and analysis levels of the Taxonomy. However, the examples provided in their article were not obviously at different levels and the specific relationships between the text and the adjunct questions were not made explicit. For example, the authors did not indicate if the knowledge level questions were verbatim repetitions of text or involved paraphrasing. More importantly, the relationship between the higher order adjunct questions and the higher order transfer questions on the posttest was not made clear. Whether these tested the same or different concepts or principles was never made clear.

From the article, I infer that different concepts and principles were tested, but this is only a guess. If the concepts and principles tested were different, then the results support the Moore (1975) study discussed above.

The results of these studies suggest that when students are given adjunct application questions about concepts and principles, as compared to adjunct factual questions, their ability to use knowledge of the concepts and principles to recognize new examples or solve problems involving the concepts and principles is enhanced (Watts & Anderson, 1971; Dapra & Felker, Note 2). The effects of the questions appear to be specific to the concepts and principles asked about in the adjunct questions; the acquisition of other concepts and principles discussed is not faciliated (Moore, 1975; Shavelson, et al, 1974; McConkie et al, 1973). The effects of other types of higher level questions are much less clear.

Some recent results originating in my laboratory raise questions about these generalizations (Andre, 1976, Note 1). In a series of three studies based upon the Watts and Anderson paradigm, I asked subjects to read prose passages explaining psychological concepts and principles and to answer either factual or application questions while reading. The materials were the Watts and Anderson (1971) passages plus additional passages constructed to match the original materials in style and format. Questions were were placed either before or after the relevant parts of the passage; hence the basic design was a 2 (Question Position) X 2 (Question Level) factorial. Over the three studies, and considering only the new application items on the posttest, the results were as follows. In Experiment 1, college student subjects given factual questions before the relevant portions of the passage did best on answering new application questions. In Experiment 2, again using college students, no significant differences were found. In Experiment 3, high school students given factual questions either before or after the relevant passage portions did significantly better on new application items than did students given application adjunct questions.

These results are inconsistent with each other and with the Watts and Anderson findings. This inconsistency raises questions about the generalizability of the effects of adjunct applicaion questions. The results suggest that the effects of questions are moderated by other variables.

One possible such variable may be difficulty level of the adjunct questions. The absolute level of performance on the adjunct questions was poorer in the Andre studies than in the Watts and Anderson study. It is reasonable to suppose that if the students are unable to answer the adjunct questions, the questions will not facilitate their performance. In fact, if the adjunct questions prove particularly difficult they might reduce overall performance by increasing the students frustration and lowering his motivation. Another variable may be the ability level of the students. Shavelson et al (1974) reported that higher-level questions had less of an effect on higher-ability students. (Although this pattern of interaction does not seem to be confirmed in the Hunkin's 1969 data.) The college students used in the Andre Note 1 research would represent a population selected for ability; the high school students came from a high school whose students' mean ability levels average above national norms. Clearly such speculations cannot account for the differences between the Andre Note 1 and Watts and Anderson (1971) results. However the speculations do suggest that future research on question level should examine possible trait-treatment interactions. The inconsistencies in the extant research underscore the need for subsequent research.

The importance of the inconsistencies in this research should not be overrated. Only a very few studies have been performed and only a very limited number of variables considered. Given the wide differences in method and procedure across studies, such inconsistencies should not be entirely unexpected. If, as I believe, questions exert only an indirect influence on the learner, then questions may well have different effects on different learners in different situations. The effects of questions will be determined by the total input influencing the learner, the prior habits and reading strategies of the learner, and how the learner combines and transforms this information into a perception of his task and goals. Only when the configuration of the system is such that the questions lead the learner to process the materials in ways he would not otherwise have done will questions influence learning and retention and transfer. The effects of questions will be conditional on other aspects of the learner's processing system. This view of the learner and of the influence of questions on the learner is pursued in the section below which proposes a model of learner and his comprehension/retention/transfer processes.

Higher-Order Effects from Higher Level Questions

It does seem clear that under appropriate, but as yet not fully specified, conditions application questions produce better transfer to new situations. As developed above, DAM does not seem adequate to account for this effect. Under DAM higher order questions had their effect by leading subjects to semantically encode more of the information in a passage. If DAM were sufficient to account for the transfer effects that have been observed, then those effects could be simulated by asking subjects several low order questions about the passage instead of asking one higher order question. On the face of it, it appears difficult to account for the observed transfer effects in this way. In the Watts and Anderson (1971) study, for example, the higher order application questions required that subjects make use of the presented information in a new way. The questions required that subjects go "beyond the information given", in Bruner's fortunate phrase. How such an affect could be duplicated by asking several factual questions about the passage is not clear. Thus, the DAM model is not sufficient to account for such higher order or transfer effects of inserted higher-order questions. A model which is designed to account for such effects is described in the final section below. Before turning to that description, it is approrpriate to note some general problems with the existing research on levels of questions.

Some Methodological Criticisms

One major problem with research on question level lies in the specification of the levels of questions in the studies. Anderson (1972) criticized educational researchers and journal editors for printing research reports which were not replicable because essential features of the materials were not fully described. The problem still exists. In several of the studies reviewed herein, essential information such as whether the adjunct questions were verbatim or paraphrased, whether the adjunct questions were repeated on the posttest, and whether the posttest tested over concepts and principles also tested in the adjunct questions or on additional concepts and principles was simply not given (e.g., Allen, 1970; McConkie, 1973; Shavelson, et al., 1974). Part of the reason for this ambiguity has been the lack of an agreed upon system for classifying questions. However, the lack of a system should imply that researchers should be especially detailed in describing the procedures by which questions of various levels were developed, because a reader cannot refer back to a system for objectively generating similar questions.

A second criticism is that many of the studies failed to provide data on the subjects' performance on the adjunct questions. Significant exceptions are the Anderson and Biddle (1975), Watts and Anderson (1971), and Andre and Sola (1976) studies. Holland (1965) demonstrated that answering relatively easy and relevant questions during instruction could facilitate performance but that answering difficult but still relevant questions would not. Where data is available, it is clear that higher and lower level adjunct questions do differ on difficulty. Thus, in the absence of data on the subjects performance on the adjunct questions, interpretations of studies varying level of questions becomes exceedingly tenuous.

A third problem, also noted above, deals with the nature of output or posttest measures. Instructional communications could have a number of potential outcomes; knowledge gained from such communications could be assessed in a variety of ways. It seems appropriate that if levels of adjunct questions are to be varied in instruction then various levels of questions should also be assessed on the posttest. Studies which rely only on one output measure, such as factual recall, reveal too little. Virtually all question level research should use multiple output measures.

Finally, it seems appropriate that investigators interested in the instructional effects of different levels of questions should devote more interest to the individual characteristic of the students. Under the model of the learner informally presented thus far, questions influence the learner by changing his perception of the task and the strategies he uses to pursue the perceived task. This conception suggests that characteristics of subjects that are related to how they perceive and act in prose learning (reading) situations will also interact with questions. Some evidence for such interactions was provided by Shavelson, et al. (1974). Like most experimentalists, prose learning researchers have for the most part ignored individual difference variables. The inconsistency in results noted herein suggests that researchers do so at their peril.

Most of these criticisms relate to the nature of scientific research and how such research should be reported and described. Good research and good science begin in careful observation and description. All of the sophisticated techniques of modern educational research are useless without these characteristics. Careful observation requires that the conditions of a study be noted as fully as possible. Such conditions include the nature of the subjects, and the nature of the experimental materials. It behooves researchers to find out as much as they can about subjects, materials, and so forth. As an example, how many prose researchers know how many males and females participated in their last study or what the reading difficulty level of the material was? Careful description requires that such descriptive data about the conditions of the study be reported. Unless such data is reported it is almost useless to report the article. Journal editors and reviewers bear a major portion of the responsibility for insuring that such information about research is available.

The point is not that such data is necessary for understanding a particular study in isolation. I agree that such information may be irrelevant to the hypothesis of a particular study. In fact, the hypotheses of an experimental study usually can be tested without knowing such information. Rather the point is that to intelligently relate, compare, and integrate different studies investigating similar but different hypotheses with different subjects and materials, such descriptive data is essential. In general, educational and psychological researchers in the prose learning area have done a bad job of reporting such information. Certainly integration of the level-of-question research would be facilitated if such information were known.

Towards a Model of Productive Learning From Prose

The purpose of this section is to present a model of productive learning. The model seeks to organize what is known about question level effects and to describe how information presented in prose is assimilated and stored in cognitive structure. In its current form, the model is primitive and somewhat speculative. I envision it as a guide for future thinking and research rather than a complete formal description. Research on prose learning and the use of questions in prose has been mostly empirical and non-theoretical in nature. One value of the proposed model is that it can serve as a theoretical guide for future research in this area. The model is certainly not a unique contribution on my part as it draws upon ideas and representations from a variety of sources. These contributions are noted below. Like most current models of memory, the present model views the learner as a complex information processing system. The basic structure of the model is illustrated in Figure 5. The now familiar features of such models: sensory registors, short term memory, an executive, long term memory, etc. are present. These structural features are highly interrelated, connections are shown between long term memory and sensory registor since schemata store in long term memory may be assessed and used in the pattern-recognition procedure. The short term memory is conceptualized somewhat differently than in the original Waugh and Norman (1965) and Atkinson and Shiffrin (1968) models. The short term store is considered to be the place where currently thought about information is stored and limits are not placed, as in the Jamesian notion of consciousness (James, 1892; see also Shiffrin & Schneider, 1977), upon the nature or type of information in STM.

The model postulates that the human information processing system continually receives information from sensory receptors and carries out various programs of processing upon such information. The way in which such information is processed is determined by both the nature of the incoming information and the context in which such information is received. Context is used here to designate the other information received along with a specific item of information and/or temporally or spatially contiguous or adjacent with such information or information retrieved from memory as a result of such information. Such other information can include sensory information received slightly before the specific item of information to be processed, internal information activated in response to incoming information, internal information related to intentions, beliefs, attitudes, and goals of the information-processer.

A specific example involving reading may clarify this notion. Consider a case where I am reading a section of an encyclopedia in order to obtain the scientific name of a certain animal. I look up a selection using a common name I believe correct, but find that I have an article on a related animal. In reading this article to discover this, I find that the article presents the common name of the animal it concerns in the first sentence. The second sentence contains the scientific name. Other sentences in the article describe the animal and also include references to other related animals. In reading these subsequent sentences in the first article, I find I have the wrong article and turn to a second article in the encyclopedia. While reading this second article, I, as an information processer, have several complexes of information influencing my interpretation of what I am reading. To illustrate, I am influenced by (1) my intention to find a scientific name, (2) my previous knowledge of scientific names such that they are often Latin and italicized, (3) the information about the order of information in the initially read article; I probably expect the scientific name to follow the common name, (4) my uncertainty that the common name I have looked up is correct. In reading the article I will probably scan to find the latinized italized scientific name, and further scan to verify that the animal I am reading about is the animal I want. Were any of the above items of information to change, my processing of the article would change.

The information processing system is further conceptualized as containing at least two distinct memory stores, an episodic memory and a so-called semantic memory (Tulving, 1972). As Tulving has argued, the episodic memory contains memories for the personal stream of events encountered by the information-processer. These memories are organized spatially/temporally; that is they are tied to places and sequences. Such events are not represented in this memory in the form in which they were input, rather episodic memory contains a record of events the system has interpreted (encoded in current cognitive jargon).

Like Paivio, I believe that the system can represent events either imaginally, that is through the visual information processing system, or verbally through the auditory information processing system (Paivio, 1969, 1971, 1974). (In fact I would probably go further and suggest that the representation may be in terms of any sensory feature depending upon the original interpretation by the system, but this is not crucial to the development of the model.) The notion that episodic memory contains stored interpretations is not new and in a sense represents one problem for the model. One task for model is to describe just when incoming information is interpreted enough to be stored in episodic memory. What I think happens is that the system stores information that it has made sense of (and perhaps the notation that some information was not interpretable). Of course this belief means that I must describe precisely what "making sense" is in terms of the model. Since this problem is not central in the present context of developing an account of learning through reading, I am going to say that episodic memory contains a record of stored perceptions and temporarily leave the problem of perception in the hands of theorists whose central interest it is (see for example, Lindsay and Norman (1973). Considerable progress towards an information-processing account of perception has been made.)

Semantic memory contains the system's abstracted or generalized knowledge. The contents of semantic memory are concepts, principles, rules, skills, etc. that are broader than specific episodes. Under the present model semantic memory is broader than what many theorists have held to be semantic memory. It is the repository of the knowledge that makes comprehension of input, not simply linguistic input, possible. This conception is congruent with recent descriptions of semantic memory (Anderson, 1977, Rumelhart & Ontony, 1977).

Specifying the nature of the content of semantic memory has been a problem to which cognitive psychology has devoted considerable attention. A number of different models of semantic memory have been proposed (Anderson & Bower, 1973; Collins & Loftus, 1975; Collins & Quillian, 1972; Kintsch, 1972; Rumelhart, et al 1972; Rumelhart & Ortony, Note 10). In general, such models represented semantic memory as a network of interconnected ideas. Networks contain nodes and connections between nodes; Figures 3 and 4 contain visual representation of networks. One major set of differences between proposed models of semantic memory seem to be primarily in the nature of the 'elements' contained at a node and the nature of the connections between nodes. A second controversial area involves the extent to which concepts are hierarchically organized and redundancy of information associated with related nodes is permitted. Since this second controversy is not essential to the current presentation, it is sidestepped here.

The question of how knowledge is represented in semantic memory (What is at a node?) is very important for understanding the effects of higher-level questions. In my estimation, questions influence the nature of the representation formed when subjects acquire new information in semantic memory. Previous models have presented two general means of representing knowledge within semantic memory. The most common scheme is to have each node represent a concept to which are associated various features (other concepts). The connections between a given concept and other concepts are of different types called "labeled directed relations." Examples of systems possessing this type of organization are Rumelhart, et al, 1972; Kintsch, 1972; Collins and Quillian, 1972; and Smith, et al., 1974.

Within this type of feature system one of the debatable issues is whether the features are entirely linguistic or whether they can include sensory-imaginal information. Anderson (1975), Anderson and Ortony (1975) and Walker (1975) have argued for the necessity of sensory information being available to semantic memory. The present model similarly adopts that position. To handle sensory information the present model argues (similarly to Lockhart, et al, 1975) that there can be connections between semantic memory and episodic memory. In the present model, sensory information is represented as being stored as episodes in episodic memory. As part of the process of acquiring new concepts connections are established between episodes and the labels given them. The labels form an entry to semantic memory. In this way certain kinds of sensory information can be available or accessed from semantic memory. How and when this information is accessed or used is discussed below.

The other general mechanism that has been used to represent a concept in semantic memory is the schema (Anderson, 1976; Rumelhart & Ortony, 1976; Schank, 1975). As used by various authors, a <u>schema</u> represents a "plot" which contains variables. The <u>plot</u> represents the constant features of a particular concept while the variables represent the changing aspects. In recognizing a concept the variables are instantiated. The schema contains information about the typical values that variables in the schema may take. These typical values may be used in the schema when the actual values are unknown. In computer terminology a schema is a kind of catalogued procedure.

To get an idea of how a schema operates to produce understanding consider the following story: <u>The man entered the door. He examined the loaves and selected one. He realized he had left his change, so he returned and got it.</u> At a surface level the story is disconnected and the sentences are not necessarily related to each other. Yet most readers quickly recognize the story as one about a man buying bread in a grocery story. In comprehending the story it is likely that readers fill in details like shelves, cash registers and clerks. In the view of schema theorists, readers have a "buying" schema which coordinates the separate information from the sentences of the story and also supplies the missing details. The schema supplies a theme that makes sense of the stimulus story.

The importance of such integrative schemata is demonstrated by Dooling and Lachman's (1971) research on ambiguous stories. Subjects given the theme recall the stories better. As an interesting demonstration that subjects do search for integrative schemata, I have used the Dooling and Lachman stories, which are very ambiguous, as class demonstrations. Very often at least one or two students (out of 150 to 200) generate the appropriate theme on their own.

It is this sense of schema that Bartlett (1932) and other classical schemata theorists were describing. This conception of schema also is congruent with Neisser's notion of reconstructing dinosaurs (Neisser, 1967, p. 285). The example also illustrates what current theorists mean by schemata (Anderson, 1977; Rumelhart and Ortony, 1977; Bobrow and Norman, 1975).

A major difference between the associated feature and schema approaches has been the extent to which the concept possesses particular defining features. Under the feature view, the concept is believed to have a set of features associated with it that are true of every instance and are used to determine if a particular instance is an instance of the category. These features can be called defining features (Smith, et al, 1974). The contrasting schema position is that concepts do not have defining features, rather instances of concepts possess a family relationship. Anderson describes this family relationship by saying that concepts possess only characteristics features, not defining ones (Anderson, et al. 1976, p. 668). This latter conception of how a concept is represented in semantic memory bears certain similarities to the concept of a fuzzy set (Zadeh, et al, 1975).

In the present author's estimation, there is no real conflict between these positions. I argue below that the nature of the representation of a concept in memory is determined by the kinds of tasks that the subject is called upon to perform with that concept. Certain kinds of tasks lead to the development of either relational features to other concepts or to the development of catalogued procedures. A particular concept, principle, idea may have multiple representations depending upon the subjects past experiences with this concept. The present model adopts the general network representation adopted by most authors. This type of representation is almost necessitated by the obvious facts of the interconnectedness of ideas. Each node in the network represents a concept, principle, or skill or other idea. Each node in the network can be connected through labeled associations to other nodes, to instances of the node in episodic memory, or to catalogued procedures that represent particular processing operations that may be carried out on the information related to the node. In addition the system is assumed to possess some general reasoning capabilities that can be applied to information connected with a node. Whether a particular node contains any given kind of connection will be determined by the kinds of tasks the learner has performed in the development of that node. Thus semantic memory is a kind of historical record of processing the system has carried out. If the learner has been asked to relate a concept to other concepts or to produce verbalizations of defining characteristics, then the representation in semantic memory will contain those type of features. If the learner has been asked to recognize unfamiliar instances of concepts, then he will have a schema (catalogued procedure) for doing so associated with the concept node. If the learner is called upon to perform a task with a particular node and the necessary information is not available in semantic memory, then the system will apply its general reasoning powers to the episodes associated with the concept and will attempt to carry out the task. Basically this model argues that the contents of semantic memory for a particular node contain the results of the various kinds of processing operations that have been conducted with that node. These contents of semantic memory partially determine the kinds of tasks the subject can do with the node; but the subject is also able to fall back on general reasoning powers to attempt new tasks.

To make this more explicit, let us trace through the development of a particular node in a semantic memory. Let us choose the principle of intermittent reinforcement as the to-be-learned information. Prior to acquiring this principle, the learner has information about prerequisite concepts stored in memory. He knows what reinforcement is, what intermittent means, what extinction is etc. What he has to acquire is (1) the verbalization the intermittent reinforcement leads to resistance to extinction, and (2) the ability to apply this principle to predict or explain particular situations. In teaching the concept, the instruction presents a verbal definition of the concept, and some examples of the concept. These events are stored as episodes in episodic memory. A node is also initialized in semantic memory, but all the node contains at this point are pointers to the episodes. The node can be accessed by the name "principle of intermittent reinforcement" and also from the examples. Figure 6 illustrates this state of affairs. (This discussion assumes the learner is successful in acquiring all presented information, of course he may fail.) The model holds that storage of epiosdes tied to a retrieval cue is the necessary precursor to the development of a node in semantic memory. However the episodes may be linguistic (presentation of a rule), descriptive (verbal description of an example), or imaginal-sensory (presentation of actual experience). For most real life concepts the initial episodes are probably some combination of these.

Now let us assume that the instruction asks the student acquiring the principle of intermittent reinforcement a question about the definition of the principle. The student must state the rule. To accomplish this, the student accesses the appropriate node through the name, he then attempts to locate an episode that contains the verbalized rule, and then attempts to repeat it. Note that the episode does not contain an exact copy of the rule, rather it contains what the student has encoded when first given the rule. Assuming the student locates an appropriate episode, he operates upon the content of the episode to produce a verbalization. The operation upon some contents of episodic memory initializes a change in semantic memory--

a representation of the verbalization is entered into semantic memory. If the subject were unable to retrieve an episode in episodic memory that contained a verbalization of the concept, but was able to retrieve one or more episodes containing presented examples of the concept, then the learner might apply reasoning to the examples in order to construct a verbalization. In this case, a representation of the verbalization based on analysis of the examples would be stored in semantic memory. In this latter case, however, the learner would have had to try to isolate some set of common features across the examples in order to develop a generalization. This analysis might also lead to the storage in semantic memory of a procedure for testing new instances to determine if they are examples or not. This latter process is a form of what is called discovery learning. Figure 7 illustrates the situation after the learner has produced a verbalization.

The development of a procedure for testing new examples can also be facilitated asking the learner to recognize new examples, as in the Watts and Anderson (1971) study. Presented with an application question the learner must retrieve examples, isolate common features, develop a system for testing new examples, and apply it to the presented new examples. This process also leads to the storage in semantic memory of a procedure for testing new examples, and if successful would increase the learner's ability to recognize other new examples. Figure 8 illustrates this state of affairs. It can be noted that in this case the learner has acquired an ability to acquire the principle through expository learning.

The procedure for testing instances that the subject develops will be as simple or as complex as necessary to integrate the presented examples and non-examples. If a simple rule will relate the examples the learner has stored in his episodic memory, then a simple rule will be stored in semantic memory. If the given examples are disparate, and bear at best what Anderson et al (1976) call a family relationship, then the catalogued testing procedure will reflect this complexity. This notion that it is procedures for doing an act that are stored in semantic memory is not new. Neisser (1967) has argued for such a position. Moreover the position has some important relationships to Osgood (1957) notion of meaning as r_m if we take n_m to be the process that produces a response instead of a copy of a response. Finally of course the idea relates back to Bartlett's (1932) and James' (1892) ideas about memory.

If a learner is asked to engage in other types of tasks while learning a concept, principle, skill or other information the kinds of features laid down in semantic memory will reflect what the student has done with the information. If the student is asked to relate the information to more inclusive information (i.e., as with advanced organizers) then subsumatory connections will be formed. If the learner is required to evaluate certain kinds of information, then evaluative features (or schemata) will be formed.

The model that is being proposed can now be summarized. (1) The learner is conceptualized as a complex information processing system. (2) The information processing system contains two types of long term memory, episodic and semantic. Episodic memory contains representations of the events encountered by the system while semantic memory contains its generalized knowledge. There are intimate relations between semantic and episodic memory. (3) Most educational tasks require that changes be made in the learner's semantic memory. (4) Semantic memory is best represented by a network model. The network contains nodes, labeled directed connections between nodes, and catalogued procedures. (5) A new node (and capability) is developed in semantic memory through a two step process in which: a) relevant episodes are laid down in episodic memory and associated with a node label, and b) the learner attempts to perform some task involving the node and the results of the processing performed in the task are connected to the node in semantic memory. (6) Because the results of operations performed on a node are stored in semantic memory, semantic memory is in a continual state of evolution as the learner uses nodes in new situations or contexts.

It seems to me that a model along the lines suggested here serves to organize and lend insight into a number of disparate areas within learning and instructional psychology. First, the model does a good job of handling the effects of questions. The model easily incorporates DAM by arguing that the effects of broader questions during instruction lead to an enriched trace in episodic memory. When the final retention task involves simple recall of the presented information, this enriched trace provides for more recall. Such an enriched episodic trace would, of course, occur only if the subject limited processing to that material needed to handle the current task.

The model can also handle the effects of higher level questions when higher level outcome measures are employed. Such questions serve to develop the semantic memory representation along certain lines. In addition the model is not embarrassed by failures of higher level questions to produce higher level outcomes. The model argues for a large number of stages that must be completed if an adequate semantic memory representation of the concept is to be formed. A failure of processing can occur at any of these stages. Beyond simply allowing for failure, the model allows prediction of the kinds of failure that can occur. While these are numerous, they are also assessable providing a means for testing predictions from the model.

Since the model argues that the semantic memory trace is dependent on the nature of the instructional task, it can handle effects like those discussed by Mayer (1975, 1976) in which different instructional treatments that emphasize different kinds of tasks were able to lead to differential superiority on near and far transfer. The model also handles the kinds of meaningfulness effects described by Royer and his associates (Royer & Cable, 1974; Perkins & Royer, 1976; Note 5). Meaningfulness from the viewpoint of the model refers to the learner's prior knowledge about the information being presented. What Royer's work has shown is when the learner has little prior knowledge about a subject matter (Royer's complex abstract passages) and the term's used in the passage access relatively empty nodes in semantic memory, presenting prior information about these nodes enhances learning. According to the model, this should happen since the encoding of the information presented in episodic memory will be strongly influenced by how well the subject can encode or interpret the presented language. The model would handle other kinds of meaningfulness effects (Johnson, 1975) similarly.

Another variable which should influence encoding of episodes is imagery. Imagery has been shown to enhance retention of presented information (Paivio, 1969, 1971, 1974; Anderson & Hidde, 1971). However the effects of imagery have typically been weaker when prose materials have been used. Moreover no study known to the author has even examined the effects of imagery when more than reproductive recall has been required. Under the model being outlined here, imagery would have its effect by influencing the ease with which episodes can be remembered. Episodes which are more vivid are remembered better. Thus, the model predicts that the effect of imagery on productive learning (learning that goes beyond the information given) and retention will be indirect. Imagery effects episodic memory, but not semantic.

The model has some interesting things to say about discovery and expository teaching which were alluded to above. It says that by appropriately choo ng tasks both expository and discovery methods can lead to the same instructional outcomes. The methods differ in the timing of certain kinds of processing, but both methods can produce the kinds of processing that lead to identical representations in semantic memory. More specifically, in order to learn a concept by discovery or by exposition, the learner must still process examples so as to construct a schema that allows him to test new possible instances. Comparisions between specific discovery and expository procedures in terms of outcomes are not relevant to any general comparison of the methods, since an outcome difference merely means that the particular procedure used did not lead subjects to do the necessary processing (that finding would have practical meaning however). A test of the methods, in general, would first have to show that the compared procedure produced identical cognitive outcomes, and then would compare the efficiency of the methods.

I think the model is quite congruent with laboratory research on concept identification-construction. Current conceptions of that research emphasize the roles of hypothesis testing and changing during learning (Kintsch, 1971; Levine, 1966). In terms of the present model the hypothesis can be considered the output of the inferential-reasoning process applied to presented examples. This hypothesis is modified as the system receives more information in a feedback episode, more examples, and has to make a response again.

The model also relates to the literature on feedback and an information-processing analysis of feedback effects in meaningful learning (Kulhavy, 1976, Note 3 ; Phye, 1977, Note 6). The model suggests that the feedback episode will have an effect only if the subject attempts to use it in subsequent responding to the problem. Very often this is true, certainly in most concept identification studies it is true. However in many educational situations and in much of the DRE research the learner is certainly not instructionally led to re-respond after the feedback episode. The current model suggests that the effect of feedback will be greater if the learner were.

The connections between episodic and semantic memories in the model give the system access to sensory features and make analysis on the basis of sensory features possible. Anderson (1975), Anderson and Ortony (1975) and Walker (1975) have provided evidence of the necessity for such capabilities in the system. However the model makes some predictions that those authors did not consider. Walker (1975) for example asked subjects to assess whether or not particular objects were in fact reasonable category members. Thirty pound turkeys were one of the examples used, which were on the borderline of reasonableness. Subjects took longer to make judgments about items near the borderlines than items distant from them. Anderson (1975) has argued that this result supports an exemplar conception of judgments. This is compatible with the present model. However, the present model goes on to predict that subjects might not use an exemplar basis to perform the task a second time. The first task might have effected changes in semantic memory such that they encoded that 30 pound turkeys were unreasonable (or reasonable).

The model that has been presented is certainly not a completed theory. Rather it is better conceptualized as a set of generalizations that need to be formalized and made specifically operational. In addition the relationship between the current model and other current conceptions need to be explored more fully (e.g., Anderson, 1975; Pask, 1975; Scandura, 1977; Wittrock, 1974). The principle value of these assumptions is to serve as a set of guidelines for research on prose learning. The model leads to some general expectations about the kinds of effects that instructional manipulations can have. These expectations can be formalized and tested in particular cases. In a sense, the model I have presented is similar to a schema as discussed by Anderson, et al (1976), Rumelhart and Ortony (1976), and Schank (1975). It is really a kind of abstract plot which can be particularized in specific situations. As such it can serve as a guide to future research, but that future research should also serve to flesh out and make more specific the more vague areas of the model.

Reference Notes

1. Andre, T. Final report grant 406-21-01. Learning concepts from prose, Ames, IA., Iowa State University, 1976.
2. Dapra, R. A. & Felker, D. B. Effects of comprehension and verbatim adjunct questions on problem-solving ability from prose material: extension of the mathemagenic hypothesis. Paper presented at the Annual Convention of the American Psychological Association, New Orleans, 1974.
3. Kulhavy, R. W. Feedback in written instruction, Laboratory for the Study of Human Intellectual Processes, Technical Report #1. Department of Educational Psychology, Arizona State University.
4. Moore, W. D. The control of learning through the use of different types of testlike events. Paper presented at the Annual Convention of the American Educational Research Association, Washington, D. C., 1975.
5. Perkins, M. R. & Royer, M. M. Knowledge Structure Retention and Facilitory Transfer in Prose Learning. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April, 1976.
6. Phye, G. D. The role of informative feedback in productive learning. Paper presented at the Annual Meeting of the American Educational Research Association, New York, April, 1977.
7. Richards, V. P. & Hatcher, C. W. The role of meaningful learning: postquestions in the comprehension of disabled readers. Paper presented at the Annual Convention of the American Educational Research Association, San Francisco, April, 1976.
8. Rosenschine, B. Teacher Compentency Research. Bureau of Educational Research, University of Illinois, Urbana, 1974.

9. Royer, J. M. Inspection behavior during verbal learning. Unpublished Master's Thesis, University of Illinois, 1969.

10. Rumelhart, D. E. & Ortony, A. The representation of knowledge in memory. San Diego: Center for Human Information Processing, University of California, 1976.

References

Allen, D. I. Some effects of advance organizers and level of questions on the learning and retention of written social studies material. Journal Educational Psychology, 1970, 61, 333-339.

Anderson, J. R. & Bower, G. H. Human associative memory. Washington, D. C.: V. H. Winston & Sons, 1973.

Anderson, R. C. Control of student mediating processes during verbal learning and instruction. Review of Educational Research, 1970, 40, 349-369.

Anderson, R. C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42, 145-170.

Anderson, R. C. Knowledge of variability and the contextual determination of meaning. Paper presented at the Meeting of the Epistemic Society, 1975.

Anderson, R. C. The notion of schemata and the educational enterprise. In R. C. Anderson, R. J. Spiro, and Montague, W. E. (Eds.) Schooling and the acquisition of knowledge. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1977, in press.

Anderson, R. C. & Biddle, W. B. On asking people questions about what they are reading. In Bower, G. (Ed.) Psychology of Learning and Motivation, 9, New York: Academic Press, 1975.

Anderson, R. C. & Faust, G. W. The effects of strong formal prompts in programmed instruction. American Educational Research Journal, 1967, 4, 345-352.

Anderson, R. C. & Faust, G. Educational Psychology. New York: Dodd Mead, 1973.

Anderson, R. C., Faust, G. W. & Roderick, M. C. "Overprompting" in programmed instruction. Journal of Educational Psychology, 1968, 59, 88-93.

Anderson, R. C. & Kulhavy, R. W. Imagery and prose learning. Journal of Educational Psychology, 1972, 63, 242-244.

Anderson, R. C. & Kulhavy, R. W. Learning concepts from definitions. American Educational Research Journal, 1972, 9, 385-390.

Anderson, R. C., Kulhavy, R. W., & Andre, R. Feedback procedures in programmed instruction, Journal of Educational Psychology, 1971, 62, 148-156.

Anderson, R. C. & Ortony, A. On putting apples into bottles--A problem of polysemy. Cognitive Psychology, 1975, 7, 167-180.

Anderson, R. C., Pickert, J. W., Goetz, E. T., Schallert, D. L., Stevens, K. V., & Trollip, S. Instantiation of general terms. Journal of Verbal Learning & Verbal Behavior, 1976, 15, 667-680.

Andre, T., & Sala, J. Imagery, verbatim & paraphrased questions and retention of meaningful sentences. Journal of Educational Psychology, 1976, 68, 661-669.

Aschner, M. J., Gallagher, J. J., Perry, J., Afsar, S., Jenne, W. & Faar, H. A system for classifying the thought processes in the context of classroom verbal interaction. Urbana, IL, University of Illinois, 1965, reported in Dunkin and Biddle, 1974.

Atkinson, R. C. & Shiffrin, R. M. Human memory: a proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.) The psychology of learning and motivation. Vol. 2 New York: Academic Press, 1968.

Ausubel, D. P. The psychology of meaningful verbal learning. New York: Grune & Stratton, 1963.

Ausubel, D. P. Educational Psychology: A cognitive view. Holt, Rinehart, and Winston, 1968, New York.

Ausubel, D. P. & Blake, E. Proactive inhibition in the forgetting of meaningful school material. Journal of Educational Research, 1958, 52, 145-149.

Ausubel, D. P., Robbins, L. C., & Blake, E. Retroactive inhibition and facilitation in the learning of school materials. Journal of Educational Psychology, 1957, 48, 334-343.

Ausubel, D. P., Stager, M. & Gaite, A. J. H. Retroactive facilitation in meaningful verbal learning. Journal of Educational Psychology, 1968, 59, 250-255.

Ausubel, D. P., Stager, M., & Gaite, A. J. H. Proactive effects in meaningful learning and retention. Journal of Educational Psychology, 1969, 60, 59-64.

Bartlett, F. C. Rembering (Cambridge University Press) London, V, 1932.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.) Taxonomy of Educational Objectives: Cognitive Domain, New York: David McKay, 1956.

Bobrow, D. G. & Norman, D. A. Some principles of memory schemata, in Bobrow, D. G. & Collins, A. (Eds.) Representation and Understanding, Studies in Cognitive Psychology. New York: Academic Press, 1975, 131-150.

Bormuth, J. R. On the theory of achievement test items. Chicago: University of Chicago Press, 1970.

Briggs, L. J. & Reed, H. G. The curve of retention for substance materials. Journal of Experimental Psychology, 1943, 32, 513-517.

Burton, W. H. The guidance of learning activities, New York: Appleton-Century Crafts, 1962.

Collins, A. M. & Loftus, E. F. A spreading activation theory of semantic processing. Psychological Review, 1975, 407-428.

Collins, A. M. & Quillian, M. R. How to make a language user. IN Tulving, E. & Donaldson, W. (Eds.) Organization of Memory, New York: Academic Press, 1972.

Craik, F. I. M. & Lockhart, R. S. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 671-684.

Distad, H. W. A study of the reading performance of pupils under different conditions on different types of materials. Journal of Educational Psychology, 1927, 18, 247-258.

Dooling, D. J. & Lachman, R. Effects of comprehension on retention of prose. Journal of Experimental Psychology, 1971, 88, 216-222.

Duell, Orpha, K. Effect of type of objective, level of test question, and the judged importance of tested materials upon posttest performance. Journal of Educational Psychology, 1974, 66, 225-232.

Dunkin, M. J., & Biddle, B. J. The study of teaching. New York: Holt, Rinehart, & Winston, 1974.

English, H. B., Welborn, E. L., & Killian, C. D. Studies in substance memorization. Journal of General Psychology, 1934, 2, 233-260.

Faw, H. W., & Waller, T. G. Mathemagenic behaviors and efficiency in learning from prose materials: review, critique, recommendations, 1976, 46, 691-720.

Felker, D. B. & Dapra, R. A. Effects of question type and question placement on problem-solving ability from prose material. Journal of Educational Psychology, 1975, 67, 380-384.

Fitch, J. G. Lectures on teaching, New York: Kellogg, 1879.

Frase, L. T. Some unpredicted effects of different questions upon learning from connected discourse. Journal of Educational Psychology, 1968, 59, 197-200.

Frase, L. T. Cy etic control of memory while reading connected discourse, Journal of Educational Psychology, 1969a, 60, 49-55.

Frase, L. T. Structural analysis of the knowledge that results from thinking about text. Journal of Educational Psychology Monograph Supplement, 1969b, 60 (6, Pt.2)

Frase, L. T. Boundary conditions for methemagenic behaviors. Review of Educational Research, 1970a, 40, 337-349.

Frase, L. T. Influence of sentence order and amount of higher level text processing upon reproductive and productive memory, American Educational Research Journal, 1970b, 7, 307-319a.

Frase, L. T. Maintenance and control in the acquisition of knowledge from written materials, in Carroll, J. B. & Freedle, R. O. (Eds.) Language Comprehension and the Acquisition of Knowledge, Washington, D. C.: V. H. Winston & Sons, 1972.

Frase, L. T., & Schwartz, B. J. Question production and answering as an aid to prose learning. Journal of Educational Psychology, 1975, 628-634.

Frase, L. T., & Selbiger, F. Some adaptive consequences of searching for information in a text. American Educational Research Journal, 1970, 7, 553-560.

Gagne, R. M. The conditions of learning. New York: Holt, Rinehart & Winston, 1970.

Galagher, J. J. Productive thinking in gifted children. Cooperative research project no. 965, Institute for Research on Exceptional Children. University of Illinois, Urbana, IL. 1965, reported in Dunkin and Biddle, 1974.

Gall, M. D. The use of questions in teaching. Review of Educational Research, 1970, 40, 707-721.

Germane, C. E. The value of the controlled mental summary as a method of studying. School and Society, 1920, 12, 591-593.

Gibson, E. J., & Levin, H. The Psychology of Reading. Cambridge, MA: MIT Press, 1975.

Green, D. W. The effects of task on the representation of sentences, Journal of Verbal Learning and Verbal Behavior, 1975, 14, 275-283.

Greeno, J. G., & Noreen, D. L. Time to read semantically related sentences. Memory and Cognition, 1974, 2, 117-120.

Griggs, R. A. Logical errors in comprehending set inclusion relations in meaningful text. Bloomington, IN: Indiana Mathematical Psychology Program, 1974.

Guilford, J. P. Intelligence: 1965 model, American Psychologist, 1966, 21, 20-26.

Guilford, J. P. The nature of human intelligence. New York, McGraw-Hill, 1967.

Holland, J. G. Response contingencies in teaching-machine programs. Journal of Programmed Instruction, 1965, 5, 474-482.

Holley, C. E. The Teacher's Technique, New York: The Century Co., 1923.

Holmes, E. Reading guided by questions versus careful reading and re-reading without questions. School Review, 1931, 39, 361-371.

Hough, J. B., & Duncan, J. K. Teaching: description and analysis, Reading, MA: Addison-Wesley, 1970.

Hunkins, F. P. Effects of analysis and evaluation questions on various levels of achievement. Journal of Experimental Education. 1969, 38, 45-58.

James, W. Psychology, Briefer Course, New York: Holt, 1892.

Johnson, R. E. Meaning of complex learning. Review of Educational Research, 1975, 45, 425-460.

Kintsch, W. Learning, memory, and conceptual processes, New York, Wiley, 1970.

Kintsch, W. Notes on the structure of semantic memory. In Tulving, E., & Donaldson, W. (Eds.) Organization of Memory. New York: Academic Press, 1972, 249-305.

Kintsch, W. The representation of meaning in memory. Hillsdale, NJ: Lawrence Erlbaum Associates, 1974.

Lancelot, W. H. Handbood of teaching skills, New York, Wiley, 1929.

Levine, M. Hypothesis behavior by humans during discrimination learning, Journal of Experimental Psychology, 1966, 71, 331-338.

Lindsay, P. H., & Norman, D. A. Human Information Processing. New York: Academic Press, 1973.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. L. Depth of processing in recognition and recall: some aspects of a general memory system. In Brown, J. (Ed.) Recognition and Recall, London: U.K. Wiley, 1975.

Madaus, G. F., Woods, E. M., & Nuttall, K. L. A causal model analysis of Bloom's Taxonomy, American Educational Research Journal, 1973, 10, 253-262.

Mayer, R. E. Forward transfer of different reading strategies evoked by testlike events in mathematics text. Journal of Educational Psychology, 1975, 165-170

Mayer, R. E. Different problem-solving competencies established in learning computer programming with and without meaningful models. Journal of Educational Psychology, 1975, 725-730

Mayer, R. E. Some conditions of meaningful learning for computer programming, Advance organizers and subject control of frame order. Journal of Educational Psychology, 1976, 143-148

McConkie, G. W., Rayner, K., & Wilson, S. Experimental manipulation of reading strategies. Journal of Educational Psychology, 1973, 65, 1-8.

McKenzie, G. R. Some effects of frequent quizzes on inferential thinking. American Educational Research Journal, 1972, 9, 231-240.

Miller, H. L. Directing study. New York, Scribner's, 1922.

Montague, W. E., & Carter, J. F. Vividness of imagery in recalling connected discourse. Journal of Educational Psychology, 1973, 64, 72-75.

Neisser, U. Cognitive Psychology. New York: Appleton-Century-Crofts, 1966.

Osgood, C. E. A behavioristic analysis of perception and language as cognitive phenomena, in President and Fellows of Harvard College (copyright holder), Contemporary Approaches to Cognition. A symposium held at the University of Colorado, Cambridge, MA., Harvard University Press, 1957.

Otto, H. J., Floyd, H., & Rouse, M. Principles of Elementary Education. New York: Holt, Rinehart, Winston, 1960.

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

Paivio, A. Imagery and Verbal Process. New York: Holt, Rinehart, Winston, 1971.

Paivio, A. Language and knowledge of the world. Educational Researcher, 1974, 3, 5-12.

Pask, G. Conversation cognition and learning. Amsterdam, Elsevier, 1975.

Richards, J. P. Interaction of position and conceptual level of adjunct questions on immediate and delayed retention of text. Journal of Educational Psychology, 1976a, 68, 210-217.

Richards, J. P. Stimulating high-level comprehension by interspersing questions in text passages. Educational Technology, 1976b, November, 13-17.

Richards, J. P., & DiVesta, F. J. Type and frequence of questions in processing textual material. Journal of Educational Psychology, 1974, 66, 354-362.

Rothkopf, E. Z. Some conjectures about inspection behavior in learning from written sentences and the response made problem in programmed self-instruction. Journal of Programmed Instruction, 1963, 2, 31-45.

Rothkopf, E. Z. Learning from written materials, An exploration of the control of inspection behavior by test-like events. American Educational Research Journal, 1966, 3, 241-249.

Rothkopf, E. A., & Bisbicos, E. E. Selective facilitative effects of interspersed questions on learning from written materials. Journal of Educational Psychology, 1967, 58, 56-61.

Royer, J. M., & Cable, G. W. Illustrations, analogies, and facilitative transfer in prose learning. Journal of Educational Psychology, 1974, 205-214.

Ruediger, W. C. Teaching procedures. Boston: Houghton Mifflin, 1932.

Rumelhart, D. E., Linsay, P. H., & Norman, D. A. A process model for long term memory. In Tulving, E. & Donaldson, W. (Eds.) Organization of Memory. New York: Academic Press, 1972, 198-248.

Rumelhart, D. E. & Ortony, A. The representation of knowledge in memory. San Diego: Center for Human Information Processing, University of California, 1976.

Scandura, J. M. Structural approach to instrumental problems. American Psychologist, 1977, 32, 33-53.

Schank, R. C. The structure of episodes in memory. In Bobrow, D. G. & Collins, A. (Eds.) Representation and Understanding. New York: Academic Press, 1975.

Shavelson, R. J., Berliner, D. C., Ravitch, M. M., and Loeding, D. Effects of position and type of question on learning from prose material: Interaction of treatments with individual differences. Journal of Educational Psychology, 1974, 66, 40-48.

Shiffrin, R. W., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 1977, 84, 127-190.

Shipley, C. M., Conn, M. M., Hildebrand, J. F. T., & Mitchell, G. T. A synthesis of teaching methods. Toronto: McGraw-Hill-Ryerson, 1972.

Smith, E. E., Shoben, E. J., & Rips, L. J. Comparison processes in semantic memory. Psychological Review, 1974, 81, 214-241.

Thomas, F. W. Principles and techniques of teaching. Boston: Houghton Mifflin, 1927.

Tulving, E. Episodic and semantic memory. In Tulving, E., & Donaldson, W. (Eds.) Organization of Memory. New York, Academic Press, 1972.

Walker, J. H. Real world variability, reasonableness judgements and memory representations for concepts. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 241-252.

Washbourne, J. N. The use of questions in social science material. Journal of Educational Psychology, 1929, 20, 321-354.

Watts, G. H. Effects of prequestions on control of attention in written instruction. The Australian Journal of Education, 1973, 18, No. 1.

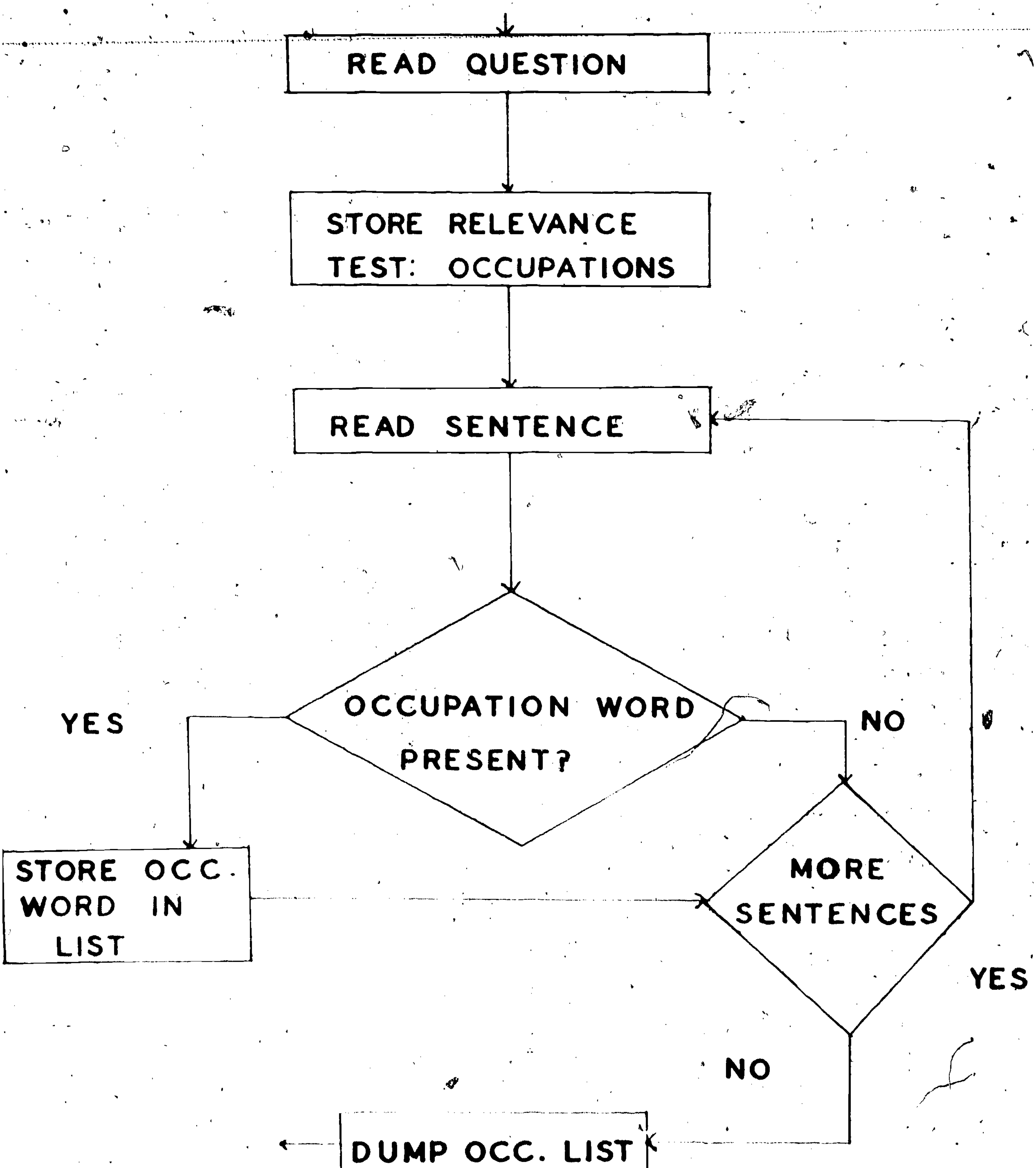
Watts, G. H., & Anderson, R. C. Effects of three types of inserted questions on learning from prose. Journal of Educational Psychology, 1971, 62, 387-394.

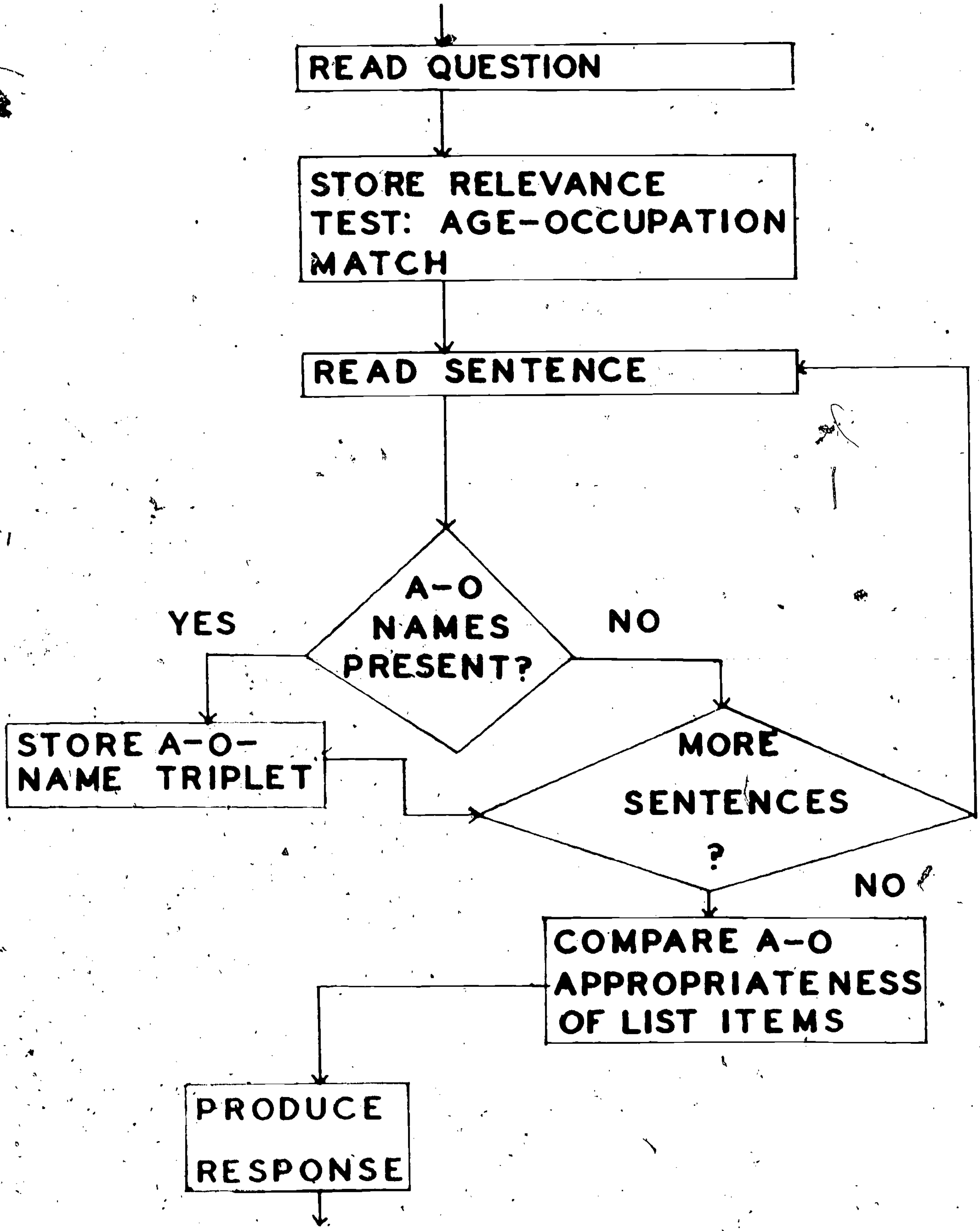
Waugh, N. C., & Norman, D. A. Primary memory. Psychological Review, 1965, 72, 89-104.

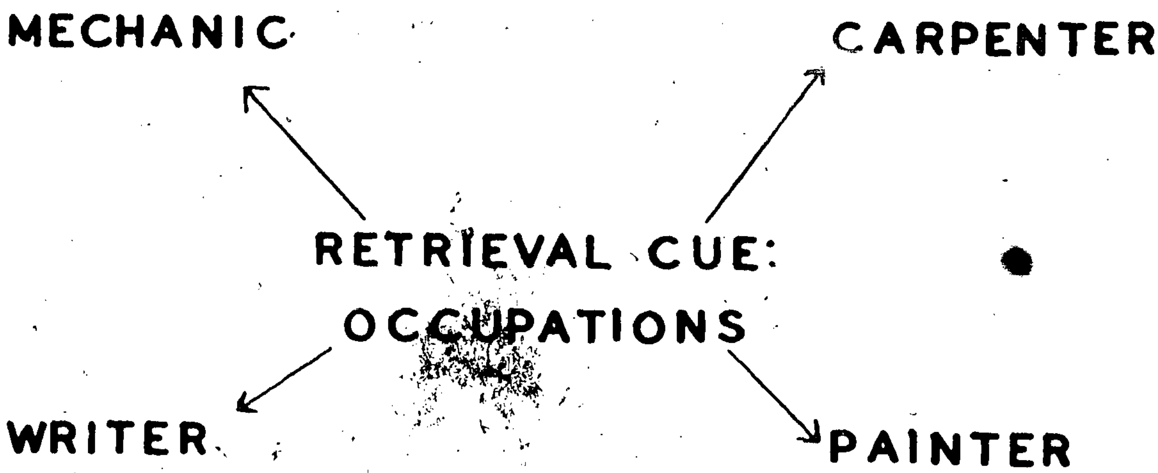
Wittrock, M. C. Learning as a generative process. Educational Psychologist, 1974, 11, 87-95.

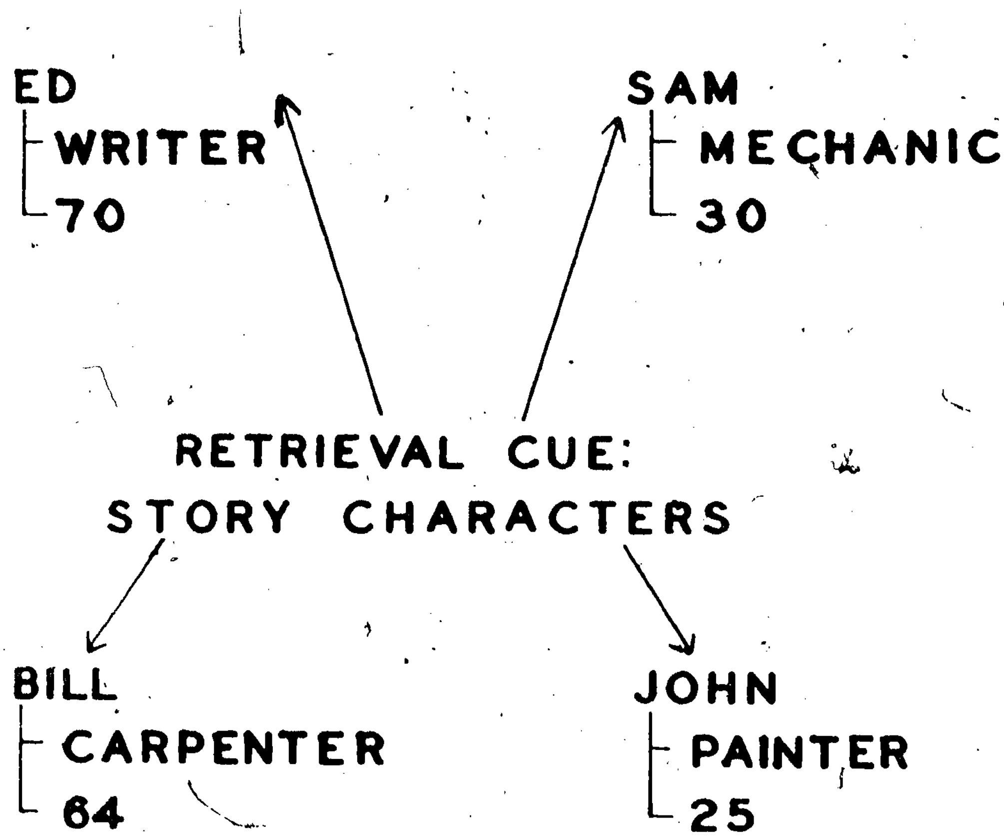
Zadeh, L. A., Fu, K., Tanaka, K., & Shimura, M. (Eds.) Fuzzy sets and their application to cognitive and decision processes. New York: Academic Press, 1975.

Figures

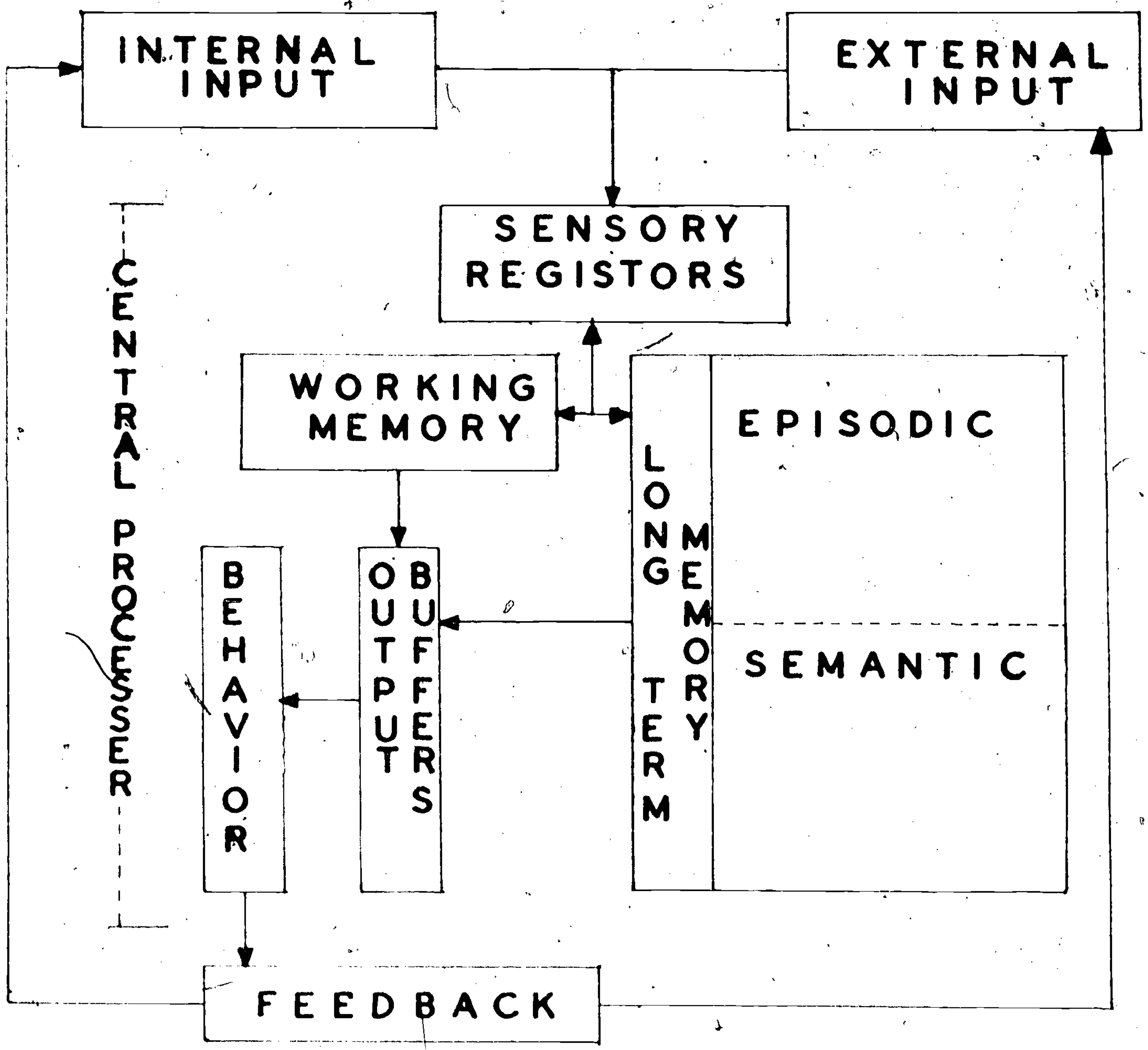

READ QUESTION
STORE RELEVANCE TEST: OCCUPATIONS
READ SENTENCE
OCCUPATION WORD PRESENT?
YES
NO
STORE OCC. WORD IN LIST
MORE SENTENCES
YES
NO
DUMP OCC. LIST


READ QUESTION
STORE RELEVANCE TEST: AGE-OCCUPATION MATCH
READ SENTENCE
A-O NAMES PRESENT?
YES
NO
STORE A-O-NAME TRIPLET
MORE SENTENCES ?
NO
COMPARE A-O APPROPRIATENESS OF LIST ITEMS
PRODUCE RESPONSE

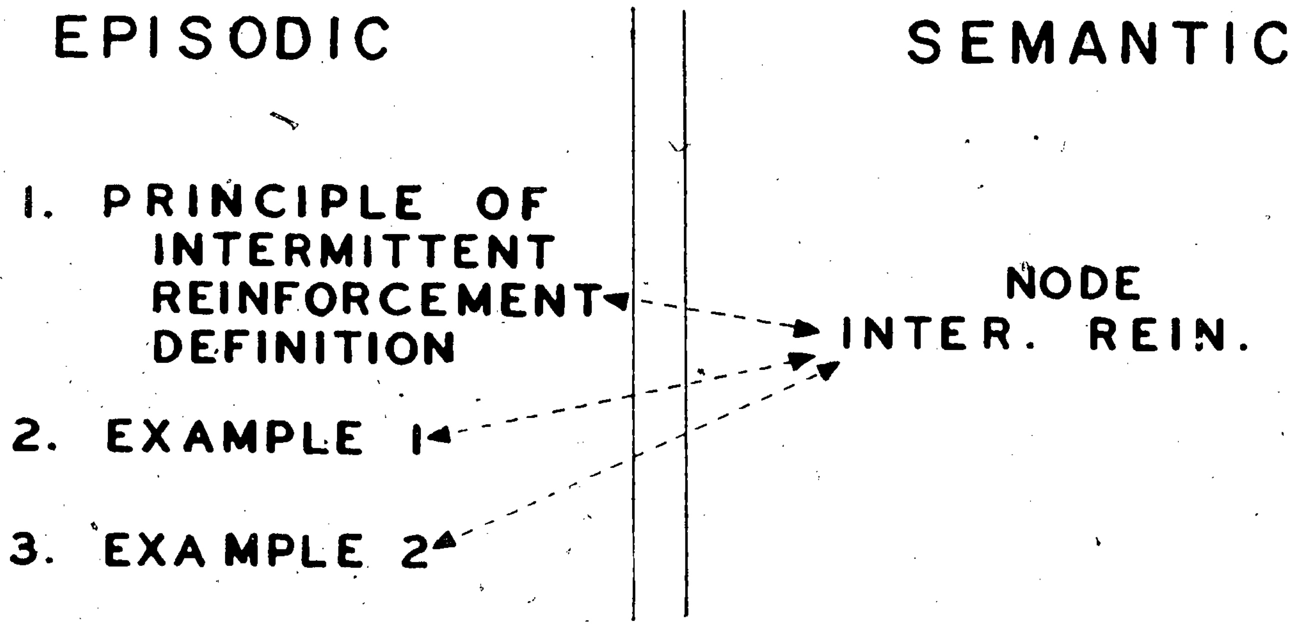

RETRIEVAL CUE:
OCCUPATIONS
MECHANIC
CARPENTER
WRITER
PAINTER

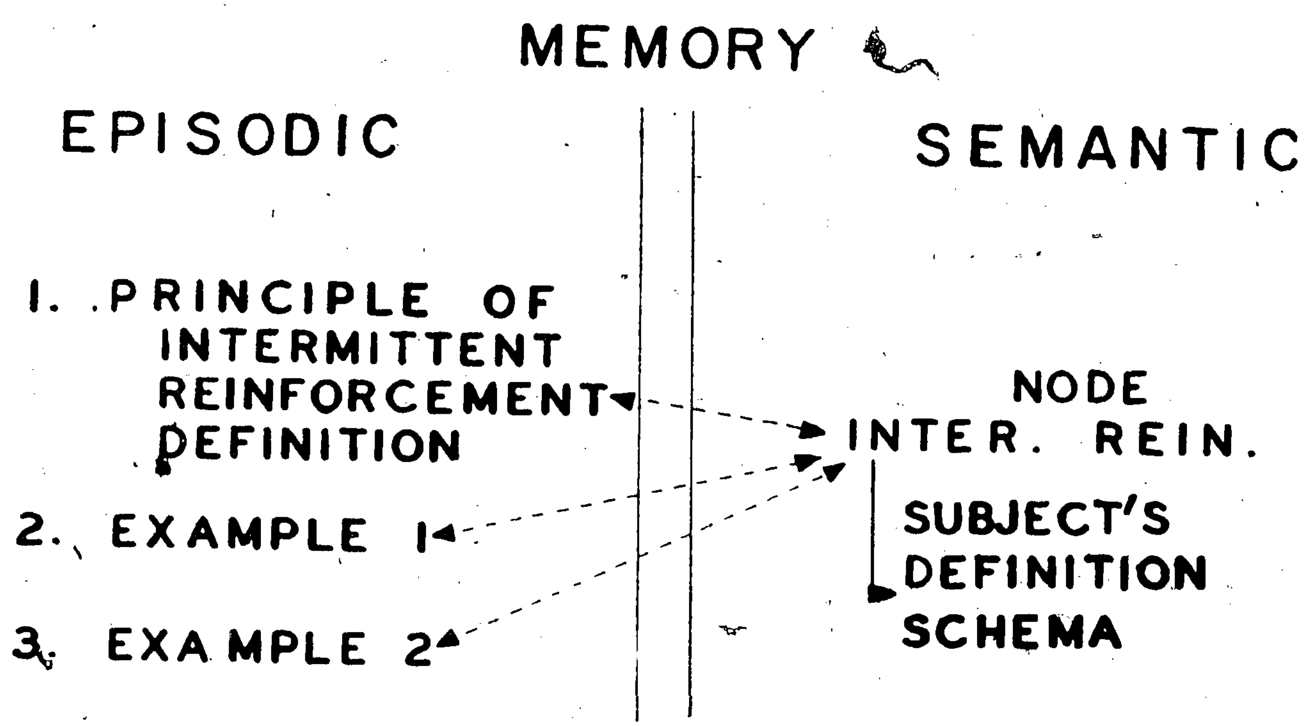


ED
WRITER
70
SAM
MECHANIC
30
RETRIEVAL CUE:
STORY CHARACTERS
BILL
CARPENTER
64
JOHN
PAINTER
25


INTERNAL INPUT
EXTERNAL INPUT
SENSORY REGISTORS
WORKING MEMORY
EPISODIC
SEMANTIC
LONG TERM MEMORY
CENTRAL PROCESSER
BEHAVIOR
OUTPUT BUFFERS
FEEDBACK

MEMORY


EPISODIC
SEMANTIC
1. PRINCIPLE OF INTERMITTENT REINFORCEMENT DEFINITION
2. EXAMPLE 1
3. EXAMPLE 2
NODE
INTER. REIN.


MEMORY
EPISODIC
SEMANTIC
1. PRINCIPLE OF INTERMITTENT REINFORCEMENT DEFINITION
2. EXAMPLE 1
3. EXAMPLE 2
NODE
INTER. REIN.
SUBJECT'S DEFINITION SCHEMA