INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.

University Microfilms International

300 N. ZEEB ROAD, ANN ARBOR, MI 48106
18 BEDFORD ROW, LONDON WC1R 4EJ, ENGLAND

8103922

COSTA, NORMAN DENNIS

AN INVESTIGATION OF SEMANTIC INTEGRATION: THE EFFECTS OF MATERIALS, INSTRUCTIONS, AND AGE ON THE PRODUCTION OF INFERENCES

City University of New York PH.D. 1981

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

Copyright 1980

by

Costa, Norman Dennis

All Rights Reserved

AN INVESTIGATION OF SEMANTIC INTEGRATION: THE EFFECTS OF MATERIALS, INSTRUCTIONS, AND AGE ON THE PRODUCTION OF INFERENCES

by

NORMAN DENNIS COSTA

A dissertation submitted in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

The Graduate School and University Center
The City University of New York

1980

This manuscript has been read and accepted for the Graduate Faculty in Educational Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

9-4-80
date

Barry J. Zimmerman
Chair of Examining Committee

9/4/80
date

Shirley C. Feldmann
Executive Officer

9/4/80
date

Alan L. Gross
A. Gross

The Graduate School and University Center
The City University of New York

ACKNOWLEDGEMENTS

The author acknowledges with great appreciation the assistance of his committee: Shirley Feldmann, Alan Gross, and Barry Zimmerman. The advice and constructive criticism of Barry Zimmerman, Chair, is most noteworthy. The author is especially indebted to the support of Personnel management and other Personnel Research professionals at the International Business Machines Corporation.

TABLE OF CONTENTS

Page

Chapter

I. INTRODUCTION 1

Purpose 1

A General Historical Perspective 2

Recent Research on Semantic Integration 3

An Historical Perspective to the Bransford, et al. Tradition of Research 6

The 'Retain Everything' Position 7

The 'Retain Nothing' Position 8

Between 'Everything' and 'Nothing' 12

The Problem 15

II. REVIEW OF THE LITERATURE 17

The 'Bransford Paradigm' for Linguistic and Semantic Integration 17

Summary of the 'Bransford Paradigm' Studies 30

Studies Critical of the 'Bransford Paradigm' for Linguistic and Semantic Integration 32

A General Critique of the 'Bransford Paradigm' for Linguistic and Semantic Integration 60

Page

Studies Investigating Developmental Effects on Semantic Integration . . 64

A Critique of the Studies Investigating Developmental Effects on Semantic Integration 75

Rationale for Studying Developmental Trends 78

The Inferential Thinking Paradigm for Studying Semantic Integration 78

Summary of the Inferential Thinking Paradigm Studies 98

Rationale 101

The Objectives of this Investigation 102
Rationale for Using Inferential Material 102
Rationale for Determining Time of Inference Production 103
Rationale for Studying Instructional Variation 103

Hypotheses 104

Materials 105
Age Within Materials 105
Instructions 106
Age Within Instructions 107

Page

III. METHODOLOGY 108

Experiment I 108

Purpose 108
Materials 108
Procedure 111
Subjects 113
Design 113
Hypotheses to be Tested 115

Experiment II 117

Purpose 117
Materials 117
Procedure 118
Subjects 122
Design 122
Hypotheses to be Tested 122

IV. RESULTS 125

Experiment I 125

The Data 125
Hypothesis 1 126
Hypothesis 2 126
Hypothesis 3 129
Hypothesis 4 131
Hypothesis 5 132
Hypothesis 6 132
Hypothesis 7 134
Hypothesis 8 135

Page

Experiment II 136

The Data 136
Hypothesis 9 136
Hypothesis 10 138
Hypothesis 11 138
Hypothesis 12 141
Hypothesis 13 142
Hypothesis 14 142
Hypothesis 15 144

V. DISCUSSION 147

Time of Inference Production 147

Developmental Effects 150

Correct Responses to Inference Queries 150
Encoding Time 153
Response Time 153
Developmental Differences in Inference Production 155

Instructional Effects 157

General Comments on Semantic Integration 159

Spontaneity 159
Personal Control 161
Lack of Developmental Trend . . . 161

Educational Implications 161

Page

REFERENCES 164

APPENDICES

A. Materials for Experiment I 170

B. Hypotheses to be Tested in Experiment I . . 180

C. Materials for Experiment II 182

D. Hypotheses to be Tested in Experiment II . 188

E. Raw Data for Experiment I 189

F. Raw Data for Experiment I -- by Logic Type 191

G. Raw Data for Experiment II 197

H. Raw Data for Experiment II -- by Logic Type 199

LIST OF TABLES

Table | Page

1. Mean Scores and Standard Deviations for Three Dependent Variables for Inference Supplied (IS) and No Inference Supplied (NIS) Materials . . . 127

2. Mean Scores and Standard Deviations for Three Dependent Variables for IS and NIS Materials and Logic Type . . 128

3. Mean Scores and Standard Deviations for Three Dependent Variables for Integrative and Verbatim Instructions 137

4. Mean Scores and Standard Deviations for Three Dependent Variables for Integrative and Verbatim Instructions and Logic Type 139

LIST OF FIGURES

Figure | Page

1. Experimental trials and data recording procedures for Experiment I 114

2. Experimental trials and data recording procedures for Experiment II 121

3. Mean correct responses for IS and NIS materials as a function of age in Experiment I 130

4. Mean encoding time for IS, NIS and the average of IS and NIS materials as a function of age in Experiment I . . . 133

5. Mean correct responses to inference queries for integrative, verbatim and total groups as a function of age in Experiment II 140

6. Mean encoding time as a function of age in Experiment II 143

7. Mean response time to an inference query as a function of age in Experiment II 145

CHAPTER I

INTRODUCTION

Purpose

The purpose of this study is to investigate the ability of children to integrate the semantic content of connected discourse. This study shall examine the effects of materials, instructions, and age (developmental effects) on different aspects of human information processing involved in semantic integration.

The concept of semantic integration suggests several properties: (a) that humans spontaneously construct wholistic representations of ideas from partial or fragmentary elements of information; (b) that resulting abstraction is stored in memory independent of the linguistic expression of that information; and (c) that the resulting abstraction may involve more information than was explicitly available during acquisition.

Why is semantic integration a topic worthy of discussion and research in educational psychology? The most obvious reason is that it has a logical and intuitive appeal (indeed, a compelling appeal) to the layman and educator. It appears on its face to apply

to school learning issues. In the detailed review in the next chapter, integrative ability is viewed by some investigators as essential in the process of language comprehension and reading comprehension. Also, the nature of the integrative process and the precise form of its memorial representation are important in the understanding of the human as an information processor.

A General Historical Perspective

The concept of semantic integration has its roots in the philosophy of Plato. The Gestalt writings of Katona (1940), Koffka (1935), and Wertheimer (1959) presuppose the concept of cognitive integration as central to their theoretical frameworks.

Semantic integration (or at least the integration of behaviors in a pure S-R theory) was anticipated in the work of Hull (1930, 1935, 1952) in his notions on foresight. Hull extended his concept of the pure stimulus act to apply to terms like reasoning, insight, inference, and intelligence. In the language of his own system he described such phenomena as the joining together of two previously learned and previously unconnected behavioral chains, for the purpose of solving a problem by the organism (cited in Kendler and Kendler, 1967, p. 159).

The theoretical construct of assimilation in Ausubel's (1963, 1968) theory of meaningful verbal learning, and the assimilation-accommodation construct in the work of Piaget and Inhelder (1969) can be considered variants on the theme of semantic integration. In the Ausubelian scheme, new information is integrated into a framework of prior knowledge from which it is given meaning. The Piagetian concepts of assimilation and accommodation assume that established cognitive schemata are modified and extended as new learning is integrated into its framework. For both Ausubel and Piaget, memory is a constantly changing accumulation of knowledge.

Recent Research on Semantic Integration

Recent research in the educational and verbal learning fields on the subject of semantic integration has been promoted greatly by the work of J. D. Bransford and his colleagues (to be reviewed in detail in Chapter II). That research has sparked continued elaborations and refinements in experimental procedures, presented compelling results, and explored theoretical positions on the abstraction of ideas (Bransford & Franks, 1970a, 1972) and the nature of human verbal comprehension (Bransford & McCarrell, 1974). The importance of the Bransford et al. research is evidenced by: (a) the proliferation of

critical studies challenging the theoretical bases and interpretations of their research; (b) new studies that have accepted some of the Bransford et al. presuppositions, but have varied the experimental procedures and materials (to be reviewed in detail in Chapter II).

The early Bransford, et al. research operationally defined semantic integration as the failure to discriminate on a recognition task between a complete compound sentence and the partial elements that make up the whole sentence. Students acquired a list of simple sentences that could be combined into longer complex sentences. On a criterion task they gave positive recognition responses to complex sentences at least as frequently as they gave them to simple sentences. The students believed they had actually seen the complete complex sentences during acquisition. Actually they only saw the fragments in isolation.

The early Bransford, et al. research was very successful in finding and replicating these results. These studies were interpreted as support for semantic integration and evidence for the psychological reality of semantic content in connected discourse.

However, a significant amount of evidence was later compiled which refuted the interpretation of the Bransford, et al. research and questioned the operational

definition of semantic integration as the failure to discriminate between a complex sentence and its parts. Later studies have shown that the Bransford, et al. findings (a) could be replicated with nonsense material, (b) were mitigated with a changing instructional set, (c) were at most evidence of syntactic integration, and (d) could be explained by a 'guessing strategy' hypothesis (to be discussed in Chapter II). Further, attempts to find developmental differences proved negative.

There was another class of experiments dealing with semantic integration that differed from the early Bransford, et al. research by adopting a different operational definition for semantic integration. Evidence for semantic integration was obtained by observing the production of inferences. Students were presented with sentences or brief paragraphs from which conclusions or inferences could be drawn. However, those conclusions or inferences were not directly forthcoming from a simple concatenation of the simple elements or sentence fragments. The students had to make inferential extensions that were logically or plausibly compatible with the passages as a whole. This was taken to be evidence of the integration of the propositions of the parts of the passage.

This class of research was more successful at demonstrating evidence for semantic integration, and survived, in advance, the criticisms leveled at the early Bransford, et al. research. However, there have been relatively few published studies, and the effects of many task and subject variables have not been investigated.

An Historical Perspective to the Bransford, et al. Tradition of Research

It is not the purpose of this study to review the entire range of historical antecedents to the issue of semantic integration in the field of general psychology. Rather, the historical perspective offered here is that which has most directly formed the basis and motivation of the modern research conducted by Bransford and his colleagues and others.

Reitman and Bower (1973) have characterized the historical antecedents to modern research on semantic integration this way: The research focus has been the age-old controversy between abstraction and the loss of detail on the one hand, and the retention of specific information on the other. The extreme positions in this controversy are that either everything is retained in memory about a concept (all exemplars plus the resulting

abstraction) or only the abstraction is retained and nothing of the detailed exemplars is left intact. Reitman and Bower refer to this controversy as the 'retain everything' versus the 'retain nothing' positions. However, both positions accept that abstractions such as concepts are learned.

The 'retain everything' position

The 'retain everything' position suggests that memory is highly specific and detailed, and that while abstraction in the form of concept development may take place, concept exemplars remain specific and intact once they are learned. Shepard (1967) gives support to this position. After presentation of a very large number of stimuli, students were able to discriminate between OLD (previously seen) stimuli and NEW stimuli. Both OLD and NEW were randomly chosen from the same very large pool of stimuli. Students were able to correctly identify the OLD stimuli in a paired OLD-NEW criterion task to an extremely high degree: The correct percentages were 90, 88, and 98 for words, sentences, and pictures, respectively.

Atkinson and Shiffrin (1968) offer a theory of memory and supporting experiments that argue for a 'retain everything' position. They developed a computer analogy

to human processes with neurological hardware and hard-wired logic functions, control processes under the direction of the human, and different types of memory registers (sensory, short-term, and long-term). The assumption was that these memory structures were specific and nondegenerating.

The 'retain everything' position was also found in research on artificial intelligence and inductive reasoning programs developed in computer applications for pattern recognition (Hunt, Marin, & Stone 1966; Nilsson, 1965; Sebestyen, 1965). The accumulation of complete information on many different exemplars in simulating pattern recognition, deductive, and inductive models of human memory were also found in Collins and Quillian (1969, 1972) and Quillian (1968, 1969).

The 'retain nothing' position

In the 'retain nothing' position it is assumed only the abstractions from a person's experiences are retained, and that the detailed exemplars that were assimilated into the framework of a person's prior knowledge are lost. This position was formalized by Bartlett as early as 1932. Bartlett proposed that abstract schemata were learned in a predominantly integrative and reconstructive way. The central meaning of a passage was

stored in schematic form and recall was achieved by a process of reconstruction. In the words of Bartlett (1932):

> If there is one thing upon which I have insisted more than another..., it is that the description of memories as 'fixed and lifeless' is merely an unpleasant fiction.
>
> ...
>
> ...[R]emembering is 'schematically' determined. The circumstances that allow memory orientation, ...always set up...towards a particular 'schematic' organization. The construction, or reconstruction, ...is...always within the range of this special organization, and...material which is central purports to be drawn from this [pp. 311-312].

Oldfield and Zangwill (1942) examined three theorists who were prominent in British psychology at that time: Head, Bartlett, and Wolters. In a critique and comparison among the three, Oldfield and Zangwill noted numerous differences and inconsistencies. However, all three theorists concluded: (a) the concept of the schema is the most salient feature to account for certain aspects of conscious behavior and experience, such as reconstructive processes in recall, and the nature of conceptual thinking; and (b) the concept of the schema offers the most valuable alternative to trace theories of memorial representation.

In what appears to be one of the earliest (perhaps prophetic) critiques and analyses of the influence of

computer analogies to, and programming simulations of, human memory processes, Oldfield (1954) suggested a general format for the computer modeling of an abstractive memory process.

> An attempt is made to decide whether the memory-model suggested by circuital storage devices in modern computing machines is better adapted to meet empirical data on recall brought to light by Bartlett and others, than is the type of model which postulates storage in the form of simulacra, or traces impressed upon the medium....[It is suggested] if the circuital storage-elements are functionally organized in such a way as to allow of re-coding of the stored message...such a system would [then] have some of the properties of schemata as postulated by Bartlett [p. 14].

Pompi and Lachman (1967) concluded from their data that meanings conveyed in connected discourse were stored in higher order 'surrogate structures' (themes, images, schemata, and words), and that the meanings conveyed by these surrogates were not completely dependent upon the simple concatenation of the semantics of the individual lexical entities.

Begg and Paivio (1969) were willing to suggest that concrete sentences, at least, were stored as unified 'semantic' images. However, they could not extend the same conclusion to abstract sentences which appeared to be retained as literal strings. This harks back to the

'retain everything' position, but other studies attributed this latter finding to greater difficulty levels for the abstract materials. For a critique of the Begg and Pavio paradigm and new experimentation related to it, see Johnson, Bransford, Nyberg and Cleary (1972).

Yuille and Paivio (1969) confirmed the results of the Begg and Pavio (1969) study for concrete sentences. The basic idea of a paragraph may be stored in a unified thematic structure in the form of a visual image.

Dooling and Lachman (1971) lent further support to the conception that semantic information from a sentence is stored in unified representation. They found a thematically related adjunct word facilitated both the integration and retention of words and phrases. The abstract memorial representation of the meaning of a passage serves as a mnemonic device in remembering prose.

In a series of studies dealing with the nature of memorial representations of visual patterns, Franks and Bransford (1971) interpreted their findings as the acquisition of abstract schemata represented by prototypes and their transformations. This report by Franks and Bransford (1971) is an extension of a series of visual pattern studies to be discussed briefly below. Although

the experimental materials were visual and not verbal it will be important to briefly review them because they greatly influenced the theoretical and experimental work of Bransford and his colleagues as they moved into the realm of connected discourse and other verbal materials.

When Bransford and Franks (1971) [to be reviewed in Chapter II] moved into the arena of verbal materials and examined their results, they set the stage for their future theoretical and experimental endeavors.

> ...Ss acquired something more general or abstract than simply a list of those sentences experienced during acquisition. Ss integrated the information communicated by sets of individual sentences to construct wholistic semantic ideas.
>
> ...
>
> In general, Ss do not store representations of particular sentences. Individual sentences lost their unique status in memory in favor of a more wholistic representation of semantic events [p. 348].

In the words of Reitman and Bower (1973): "...[T]he 'store nothing' position has been argued quite forcefully by Bransford and Franks (1971) and Franks and Bransford (1971) on the basis of their data [p. 195]".

<u>Between 'everything' and 'nothing'</u>

The following experiments used visual materials and dot patterns for the learning, recognition, and discrimination of prototypes and prototype transformations.

A study by Posner, Goldsmith, and Welton (1967) laid down a basic set of experimental procedures that established the foundation of the studies reviewed below and helped Franks and Bransford (1971) set up their visual pattern study.

Posner and Keele (1968) taught students to classify different types of distortions of a prototype visual pattern of dots. They later tested the ability of the students to correctly characterize the original prototypes, old distortions, and new distortions. Since old distortions and the prototypes themselves were characterized equally well it was presumed that the students abstracted the correct concepts. However, the better performance on old distortions compared to new distortions suggested that the students learned and retained something specific about original exemplars. Along with the abstraction of concepts, some exemplar information was also stored. These findings were supported in further studies by Posner (1969), Posner and Keele (1970), and Strange, Keeney, Kessel, and Jenkins (1970).

In summary, the Posner, et al. and the Strange, et al. studies suggest

> ...that there are three aspects to 'what is learned', that is the nature of the schema acquired in these studies. (a) The Ss acquire specific information about the individual configurations experienced

during acquisition. (b) The Ss acquire some general information about the common properties or general structure of the set of configurations in a pattern (a prototype). (c) The Ss acquire some information about the possible variation in instances of a pattern (transformations specifying deviance from the prototype) [Cited in Franks and Bransford (1971), p. 65].

Peterson and McIntyre (1973) [to be reviewed in Chapter II] support these concepts by suggesting that their data show both a specific and integrated component process in human memory and that a general theory of memory must assimilate the two phenomena. Or, as Reitman and Bower (1973) suggested: The truth, in all probability, lies between the two positions.

The research on semantic integration with verbal materials that was promoted by J. D. Bransford and his colleagues will be reviewed in depth in Chapter II. The historical perspective, given above, to the Bransford research will afford the opportunity to see it in the context of the 'retain everything' versus the 'retain nothing' controversy. However, it is <u>not</u> the purpose of this investigation to develop an experimental model that will forever solve the controversy. The intent was to give the background from which the Bransford research developed its ideas and experimental procedures.

The Problem

The purpose of this investigation was briefly outlined at the opening of this paper. A general description of the concept of semantic integration was given. Two operational definitions were presented for semantic integration, and a more precise analysis of those definitions will be developed in the next chapter. Brief mention was made of the incorporation of semantic integration into several psychological theories, and that it was intuitively applicable to school learning. Finally, a modest body of research emerged primarily from the work of J. D. Bransford and his colleagues. An historical perspective to this work was outlined.

The problem can now be stated simply.

(a) Is the concept of semantic integration applicable to verbal materials that are characteristic of connected discourse?

(b) Assuming that the integration concept is applicable to connected discourse, what are some of its characteristics?

(c) What would be the nature of integration and what would be the psychological processes underlying this phenomenon?

(d) What are some relevant task variables that influence semantic integration and what are the educational implications of these results?

CHAPTER II

REVIEW OF THE LITERATURE

The 'Bransford Paradigm' for Linguistic and Semantic Integration

In the first section of this chapter a series of studies representing the 'Bransford Paradigm' of semantic integration research will be reviewed. Studies critical of the theoretical assumptions of the 'Bransford Paradigm' and which offer research evidence to support those criticisms will next be reviewed. At this point it should be made clear just what is meant by the 'Bransford Paradigm'.

From an operational perspective, the 'Bransford Paradigm' is simply the procedure for materials construction, experimental design, and data analyses detailed in the Bransford and Franks (1971) study. The 'Bransford Paradigm' refers to combinatorial manipulations of linguistic inputs into sequences of varying lengths. These combinations of discrete elements into strings of varying lengths are called abstractions or wholistic ideas. Several simple sentences are combined into a single complex sentence. Katz, Atkeson, and Lee (1974) refer to what this author calls the 'Bransford Paradigm' as the "...standard Bransford-Franks format [p. 711]."

In reviewing the first study by Bransford and Franks (1971), much detail of the rationale, experimental materials, procedures, and scoring will be presented. This is important because many other research studies have modeled themselves on this study (its materials, methods, and scoring) and/or critically analyzed the same.

Bransford and Franks (1971) noted that other studies of the relationship between language and memory have dealt with memory for single sentences, that syntactic variables have been studied more than semantic variables, and that psycholinguistic theories have been more influential in promoting research on language learning than psychological theories of memory.

In contrast to this, they proposed to study memory for whole ideas that come from the integration of information in several simple sentences. They developed a methodology to study the acquisition of ideas and to seek support to the psychological reality of 'intersententially defined' ideas [p. 332]. The authors offered an awkward definition of an idea as something that "...could be exhaustively characterized as those semantic relations contained in a single sentence ...[p. 332]."

It was hypothesized that students would not be able to discriminate, on a recognition task, between simple sentences actually heard during acquisition and complex sentences which were not presented during acquisition, but which were combinations of actual simple sentences, and which were semantically consonant with the acquisition sentences. It was also hypothesized that the rating of confidence in the recognition of novel complex sentences would be a positive function of the number of simple assertions contained in the novel sentences.

Four complex sentences were invented and each was subdivided into four simple declarative sentences. For example, the complex sentence, _The ants in the kitchen ate the sweet jelly which was on the table_, was subdivided: _The ants were in the kitchen_, _The jelly was on the table_, _The jelly was sweet_, _The ants ate the jelly_. These simple sentences were called ONEs, and the complete complex sentences FOURs. Combinations of two and three simple declaratives were called TWOs and THREEs. Four TWOs and three THREEs were constructed.

Three experiments were conducted. In Experiment I the participants were 15 undergraduate students. They divided into two groups counterbalanced for presentation

order. There was an acquisition phase and a recognition phase.

In the acquisition phase the students were told that they had to answer questions about sentences which were read by the experimenter. After a sentence was read, the students named four colors on a card held by the experimenter, and then answered one elliptical question about the sentence: Did what? What did? Where? The acquisition list consisted of a total of 24 sentences, four sets of six sentences.

Within each set of six sentences there were two ONEs, two TWOs, and two THREEs. These six sentences covered all four simple declarative elements for that sentence set or frame.

The students were naive to the recognition task. In the recognition phase, the students were presented a list of 28 sentences with instructions to indicate which sentences they actually had heard before and which ones they had not. Further, they rated the confidence of their judgements on a five-point scale from 'very low' to 'very high' confidence. Twenty-four (24) sentences were semantically consonant with the acquisition sentences, but were not presented during acquisition. These were called NEWs (two ONEs, two TWOs, one THREE,

and one FOUR for each of the four sentence frames). Four of the acquisition sentences were also included. These were called OLDs.

Experiment II paralleled the first. The materials were constructed the same way and the acquisition procedure was the same. However, on the recognition task, instead of including four OLDs, six NONCASE sentences were employed. A NONCASE sentence used the same 'units' as the other sentences but they were combined <u>across</u> sentence frames rather than <u>within</u> a frame.

The scoring procedure for both experiments was the same. A "Yes" response on the recognition task was scored a plus one (+1); a "No" response a minus one (-1). This was multiplied by the confidence rating which ranged from one, very low, to five, very high. Thus, the scores ranged from plus five (+5) to minus (-5), excluding zero [p. 339].

The results for both experiments were as follows:

1. Many NEW sentences received positive recognition ratings. The students actually thought they had heard these sentences during acquisition.
2. Many NEW sentences received higher scores than the OLD sentences actually heard before.

3. There was a relationship between the number of simple units in the NEW sentences and the recognition-confidence ratings. Recognition ratings were ordered as follows: FOURs - THREEs - TWOs - ONEs. NONCASE sentences received the lowest ratings of all.
4. The students were confident they had not heard NONCASE sentences before.

In Experiment III, materials were constructed as in Experiments I and II with four complex sentences (idea sets). For each of these idea sets there were four ONEs, four TWOs, three THREEs, and one FOUR. The procedure paralleled the first two experiments. Two acquisition lists were constructed, one for each group, containing 24 sentences, six related to each of the idea sets. The two acquisition lists were complementary, exhausted all sentences, and did not overlap.

The students were naive to the recognition task. The recognition task contained only one presentation of all 48 sentences. Thus, half the sentences were OLD and half were NEW. The OLD sentences for group I were the NEW sentences for group II and vice versa.

The results indicated that the students could not discriminate between NEW and OLD sentences for FOURs, THREEs, and TWOs. At the level of ONEs, OLD sentences

received higher recognition-confidence ratings than NEW sentences. Also, the rank order effect noted in Experiments I and II was replicated. This was the correlation of recognition-rating scores to sentence complexity.

> Data [from these experiments] indicate that Ss acquired something more general or abstract than simply a list of those sentences experienced during acquisition. Ss integrated the information communicated by sets of individual sentences to construct wholistic semantic ideas.
>
> ...
>
> In general, Ss did not store representations of particular sentences. Individual sentences lost their unique status in memory in favor of a more wholistic representation of semantic events [p. 348].

The authors noted that there were other issues to be clarified in future experiments:

1. What is learned? What is the nature of the semantic ideas that are acquired?
2. What will be the effects of different types of acquisition instructions? The authors cited the findings in a Ph. D. thesis by Curnow (1969) which suggest that instructions to process similar materials either rotely or semantically yield results consonant with the immediate study.
3. The authors further suggested that there might be two independent memory representations:

(a) a memory for semantic structures that stores wholistic semantic representations; and

(b) a representational storage for input style, that is, information about the style of the acquisition list.

Franks and Bransford (1972) felt that previous research had demonstrated the semantic integration of partial information into stored complete ideas for concrete verbal input. They now wanted to test the generalization of this phenomenon to abstract ideas. Other research indicated that concrete ideas were stored as images, while the process was different for abstract information.

The authors noted that in their previous study (Bransford and Franks, 1971) they used concrete material that could be easily imaged. They cited a study by Begg and Paivio (1969) which suggested that concrete and abstract material are differentially stored in memory and that the 'integration' of units into wholistic ideas might not hold for abstract materials which could not be easily imaged. Their new study would determine if partial meanings of abstract verbal content could be integrated by students into a complete idea and stored in memory, similar to what had been found for concrete

ideas. A complete idea was rather loosely described -- not defined -- as the imbedding of four propositions into a single complex sentence.

Two experiments were conducted. In Experiment I four sets of sentence frames were constructed to parallel the materials in the Bransford and Franks (1971) study. The sentences were screened by a previous group who rated them on a five-point scale, from very concrete, easily imaged, to very abstract, not easily imaged. An example of a complex idea or FOUR was: _The arrogant attitude expressed in the speech led to immediate criticism_. A ONE, TWO and THREE were respectively: _The attitude was arrogant_; _The arrogant attitude was expressed in the speech_; and _The arrogant attitude expressed in the speech led to criticism_. In addition, four NONCASE sentences were constructed.

The acquisition procedure was the same as in the Bransford and Franks (1971) study. A 40-sentence recognition list contained 24 sentences, including the FOURs that related to each complete idea, which were not used during acquisition (NEWs). In addition 12 OLDs and four NONCASEs were used.

The naivete to the recognition task and the scoring procedure were also similar to that in the earlier study.

However, the recognition phase was presented orally, one sentence at a time, with instructions to indicate whether it was presented before and to rate the confidence of the reply.

The results showed an ordering effect -- a linear relationship between sentence complexity and recognition-rating scores similar to the prior study. FOURs elicited the highest mean recognition-rating responses and ONEs the lowest. NONCASEs had lower scores than ONEs.

Experiment II was conducted to precisely assess the differences between NEWs and OLDs. Two complementary acquisition lists were formed containing 24 sentences each. No NONCASEs were used. Acquisition, naivete, recognition and scoring procedures were identical to Experiment I.

Half the students received acquisition list 1 and the other half received list 2. Both received the same 48 sentence recognition list so that half the sentences were OLD and half NEW. Further, the OLD and NEW sentences for one group were the NEW and OLD sentences, respectively, for the other group.

The results confirmed the ordering effect whereby the sentences with the greater number of propositions elicited the greater mean recognition-rating responses.

This effect was manifest within OLDs and NEWs. This was interpreted as supporting the hypothesis that students integrated partial meanings and stored the complete ideas. Further, the OLDs were not discriminated from the NEWs except for the ONEs, where OLDs received higher mean scores. This also confirmed the findings of the earlier study. With these latter results in mind the authors retreated slightly from their earlier 'retain nothing' position. At least for the ONEs and to a lesser extent TWOs, students retained some information specific to the simple sentences that were presented in acquisition. Finally, they concluded that there were similarities for both concrete and abstract materials in the memorial representations of ideas.

Peterson and McIntyre (1973), in the introduction to their experiment, referred to two demonstrations of memory effects which appeared to contradict one another. The first was the specific recognition memory demonstrated by Shepard (1967). After viewing semantically unrelated sentences, students were able to discriminate with a high degree of accuracy between NEW (previously unviewed) and OLD (previously viewed) sentences. The theoretical framework for this memory phenomenon in a

specific or reproductive framework was offered by Atkinson and Shiffrin (1968).

The second was the <u>Integrated recognition memory</u> demonstrated in the Bransford and Franks (1970a, 1970b, 1971) studies. After hearing semantically <u>related</u> sentences, students were unable to discriminate between OLD (previously heard) sentences and NEW (previously unheard but consonant with prior meanings) sentences. The theoretical roots of this integrative or reconstructive framework go back to Bartlett (1932).

Peterson and McIntyre attempted to clarify the conditions under which integrated recognition memory and specific recognition memory were found. The materials, acquisition procedure, naivete to the recognition task, the recognition-rating tasks, and scoring procedures were essentially the same as in the Bransford and Franks (1971) study.

In the acquisition phase two groups received 24 related sentences plus three unrelated sentences. Students in two other groups heard 24 unrelated sentences plus three related sentences.

During recognition each student heard 72 sentences: 16 for each of four complex sentence frames and eight

NONCASE items. Four were within-frame NONCASEs and four were between-frame NONCASEs.

The findings of this study showed that students in the unrelated condition generally performed better on sentence recognition than students in the related condition. In fact, they performed as well as the students in the Shepard (1967) study. However, while subjects in the related and unrelated conditions demonstrated equal performance on recognizing OLD sentences, those in the related condition recognized just as many NEW sentences.

For the related acquisition conditions, confidence ratings varied with the complexity level of the related recognition sentences. For the unrelated acquisition conditions, confidence ratings did not vary with complexity level. These findings confirmed the earlier findings of Bransford and Franks. Similarly, NONCASE sentences were rejected.

The results supported both a specific and integrative recognition memory for sentences.

> However, it is not clear why...relatedness is so important [for integration recognition memory]. It may be...that the semantic relatedness contained within the acquisition sentences themselves is the important consideration -- that subjects use it as an organizational structure within which individual

> sentences are integrated into a general representation. On the other hand, it may be that the degree of semantic relatedness between the acquisition and noise sentences used during recognition is the important consideration -- that subjects simply are unable to discriminate between the sentences. Most likely, it is some combination of both these considerations that is important [Peterson and McIntyre (1973), p. 705].

This statement was noteworthy in that Peterson and McIntyre (1973) suggested, for the first time, that the results from the Bransford, et al. studies could be accounted for, not by semantic processes, but by non-semantic operations of poor discrimination and inappropriate generalization.

Summary of the 'Bransford Paradigm' Studies

The Bransford, et al. studies, in general, not just studies using the 'Bransford Paradigm', have been a reaction to the prior dominance of psycholinguistic theory in investigating sentence processing: memory and comprehension. Fillenbaum (1971) cogently expressed the concern:

> One general problem sort of arises with regard to much of this [previous] work -- it is often not clear whether it is primarily directed toward the assessment of the psychological reality of some syntatice structure or, say, toward an analysis of how a sentence might be learned and remembered [p. 277].

Psycholinguistic theory was not criticized as incorrect, but rather limited in a complete understanding of sentence comprehension and memory.

The results of the 'Bransford Paradigm' studies which provide support for an integration memory model are:

1. Students could not discriminate on a recognition task between OLD sentences and NEW, but semantically consonant, sentences never before seen.
2. There was a linear relationship between recognition-confidence ratings and sentence complexity for OLD and NEW sentences. The author's interpretation was that increased sentential complexity tended to exhaust all the semantic relations contained within the semantic whole.
3. Students were confident they had never heard NONCASE sentences, indicating they 'could not be fooled.'
4. The effect of integration was accomplished despite random interspersion of the acquisition sentences.

However, it was never made clear why these findings were taxing to a psycholinguistic model of sentence processing.

Studies Critical of the 'Bransford Paradigm' for Linguistic and Semantic Integration

Singer and Rosenberg (1973) cite Bransford and Franks' (1970b, 1971) finding that integration of linguistic inputs into semantic units was demonstrated. Singer and Rosenberg do not dispute the implication that people abstract and store the meanings of sentences apart from the linguistic surface features within which they are represented. They question Bransford and Franks' conclusions regarding a correlation between confidence ratings and sentence complexity. Singer and Rosenberg reference the findings of two other studies which suggested that prompted recall of a sentence depends, not on the complexity of a sentence, but on the grammatical function of the prompting word. Blumenthal (1967) found that "[n]ouns involved in the more inclusive relation (subject-of-sentence) were more effective prompts than modifier nouns [p. 203]." This finding was supported in a subsequent study by Blumenthal and Boakes (1967).

Singer and Rosenberg replicated Experiment II of Bransford and Franks (1971) to show that the original

grammatical function of a word imbedded in a sentence determines the likelihood that a student will report recognizing that sentence.

The verbal materials were those used in the Bransford and Franks (1971) report. From four complex ideas: two THREEs, two TWOs, and two ONEs were selected, totaling 24 sentences for the acquisition phase. The sentences were individually presented on a tape recorder, followed by an elliptical question, and a filler task of color naming.

The students were naive to the recognition task. During recognition the students heard 36 randomized sentences not presented in acquisition, including 12 filler sentences and two ONEs, two TWOs, one THREE, and one FOUR from each of four complex frames. The 12 fillers were divided into six within-frame (wf) NONCASEs and six between-frame (bf) NONCASEs. It was felt that students would be more likely to give positive recognition rating responses to wf than bf NONCASEs.

As in the Bransford and Franks studies, the students indicated "Yes" or "No" whether the exact sentence had appeared during acquisition and rated the confidence of their choice.

The basic findings of the Bransford and Franks work were replicated with mean recognition-rating responses increasing with sentence complexity and wf NONCASEs being intermediate between ONEs and bf NON-CASEs. This supports the notion that linguistic units were integrated. However, when the absolute confidence ratings were examined among the ONEs, TWOs, THREEs, and FOURs, there were no significant differences. Thus, ignoring whether the students responded "Yes" or "No" to the recognition sentences, they were no more or less confident of their choices.

While the recognition rate of THREEs and FOURs was uniformly high with means of 80.0 and 81.7 and standard deviations of 16.3 and 6.4, the same mean recognition rates of ONEs and TWOs were lower but with much higher standard deviations: the means were 44.2 and 67.5, and the standard deviations were 29.6 and 19.7. This meant that some ONEs and TWOs were recognized with very high frequencies. The question is: Why?

A fairly complex regression analysis was applied to the Singer and Rosenberg data and the Bransford and Franks data. The results indicated that the grammatical relations of the ONEs and TWOs resulted in average

lower recognition rates. If the MAIN relation was expressed in the ONE or TWO, the recognition rate was high; if not, the rate was low.

For example, a relevant FOUR would be _The scared cat running from the barking dog jumped on the table_. The MAIN relation would be _The cat jumped on the table_. The MAIN relation is the relation expressed by the principal clause of the relevant FOUR.

Consequently, Singer and Rosenberg felt it was incorrect to assume that the confidence a person had in his rating was related to the complexity of the recognition sentence. The presence of the MAIN relation in a sentence of any complexity contributes most to the tendency to recognize it. This supported the idea that grammatical form and function of linguistic units influenced the probability that it would be integrated with other information. High confidence scores for FOURs were greatly due to the fact that they all have the MAIN relation, not greater sentence complexity.

Franks and Bransford (1974) responded to Singer's and Rosenberg's (1973) contention that the order of frequency of recognition of FOURs > THREEs > TWOs > ONEs was due to increased likelihood of the presence of the MAIN relation in the complex sentences. Franks and

Bransford retabulated their earlier data using only sentences containing the MAIN clause. The significant ordering of FOURs > THREEs > TWOs > ONEs still held, using a Monte Carlo technique described in an earlier Bransford and Franks (1971) paper.

Franks and Bransford (1974) restated their belief that the data lend strong support for the notion of semantic integration. Franks and Bransford have re-responded to the first serious questioning of their data and conclusions.

Reitman and Bower (1973) presented a study that was both important and dramatic for three reasons: first, they offered an historical and theoretical perspective to semantic integration research; second, they were the first to label the 'Bransford Paradigm' as a concept attainment task and to refer to the sentence components as exemplars of a concept; and third, their data and interpretations offer the first substantive critique of the semantic integration hypothesis.

In rehearsing the major findings of the semantic integration research, Reitman and Bower noted the following: (a) Students were unable to discriminate between old (previously acquired) and new (never before presented) exemplars of a concept. The complex embedded

sentence in the 'Bransford Paradigm' was the <u>concept</u> and the individual sentence parts were the <u>exemplars</u>. (b) The confidence ratings of the students increased as did the number of exemplars in the recognition task. In fact, the complete concept containing all exemplars, while never before seen during acquisition, received the highest confidence rating. The explanation proposed by Bransford, et al. was that recognition-confidence ratings increase with the exhaustion of all possible semantic relations that are characteristic of a complete idea. (c) The conclusion from the 'Bransford Paradigm' studies was that general concepts were more likely to be retained relative to individual exemplars of concepts.

Reitman and Bower were not satisfied with the conclusions and explanations from the 'Bransford Paradigm' studies. They offered several explanations to the apparent retention of schemata over exemplars:

1. Since students are exposed to seven or eight exemplars of many concepts and were asked to remember which combinations of exemplars occurred, it was quite possible that task memory was overloaded. In addition, each exemplar

was a full sentence with peripheral syntactic information.

2. A massive amount of forgetting due to interference might have occurred since there was great overlap within and among concept frames (Osgood, 1949).

3. There was a lack of a control group to examine forgetting (and consequent lack of discrimination) and interference among old acquired and newly presented exemplars.

As such, the authors decided (1) to repeat the Bransford and Franks (1971) procedure with simpler learning materials; (2) include the appropriate control group; and (3) to resolve the important issue contrasting retention of specific exemplars versus the retention of specific combinations of exemplars called concepts.

The material elements consisted of simple letters and digits, although the generation of concept strings followed the same rules as the Bransford and Franks (1971) procedure.

There were three conditions: (1) the obvious concept group (Ob) with three concepts, ABCD, 1234, and wxyz; (2) the difficult concept group (Df), 41ab, 3yzD, and w2Cx; and (3) the random group (R) where neither concepts

nor rules existed. The elements were randomly taken from the *Ob* and *Df* groups. Finally, two other control groups used elemental materials but differed in the existence or obviousness of a concept.

There was an acquisition phase and a recognition phase. Each string was presented only once for recognition. Acquisition was carried out in the guise of a short-term memory test. Presentation was on a memory drum for each exemplar for two seconds, followed by a 10-second filler task. Twenty-four (24) strings were presented, eight from each of three concepts.

During the recognition task, the students rated each of 45 or 65 items (depending on the conditions) as to whether they were clearly not presented during acquisition, minus five (-5) or clearly were, plus five (+5). On recognition the *Ob* and *Df* groups received the eight OLD strings from each of three concepts, seven NEW:LEGAL strings which were the complement of the eight OLD strings, and 20 NEW:ILLEGAL strings or non-exemplars. Half of the NEW:ILLEGALs were misordered within a concept set, and half misordered across concept sets. The *R* group received no ILLEGALs. The OLD, NEW:LEGAL, NEW:

ILLEGAL (ORDER), and NEW:ILLEGAL(ORDER and SET) correspond to Bransford's and Frank's OLD, NEW, wf NONCASE, and bf NONCASE, respectively.

Students discriminated well between OLD items seen previously and NEW:LEGAL items that fit the rule but were not seen previously. However, they did abstract a rule since the NEW:LEGAL items were recognized more than the NEW:ILLEGAL items for both concept conditions (Ob and Df). As for confidence ratings and complexity, when the rule was obvious the score went up with increasing test item length (condition Ob); when there was no rule the score went down with increasing test item length (condition R); the intermediate group (Df) showed intermediate results. Thus, when a rule or concept was obvious, the tendency was to identify a more complex exemplar of a rule as familiar; and when there was no rule, the tendency was to identify the more complex exemplar as never seen.

Reitman and Bower suggested several questions to be answered in light of the above results. Is particular information relevant to an abstraction retained? How does recognition differ for rule-governed judgments and similarity judgments? Why do complete instances

elicit greater frequencies of recognition? The following were tentatively hypothesized to answer these questions:

1. There was memory for the specifics of the abstraction.
2. Students store the full unit or concept presented, and the recognition for any exemplar (a presented-before, OLD, or a never-presented, NEW) follows a probablistic decision theory approach: Store the whole concept with all its parts, show any part thereof on a recognition task, and with a given probability (relative to its length) it will be recognized.
3. The conceptual rule _per se_ was not learned. Instead a set of training strings generated a certain frequency distribution for sub-units into memory. Recognition ratings for test strings were then computed from these frequency distributions.

Katz (1973) reviewed the relevant literature (cited earlier) that suggested that unified semantic representations were stored in memory in the process of comprehending connected discourse. However, Katz questioned the

way in which the results were obtained. Katz offered the most incisive evaluation and critique, to date, on the Bransford and Franks (1971) results.

> Since Bransford and Franks (1971) have been concerned to explain _semantic_ processes, their interpretation of these has focused on memory for the underlying representation of linguistic inputs and not on memory for other aspects of sentences [such as meaning]. Therefore, it seems somewhat inappropriate that during the recognition phase of their experiments they instructed Ss to evaluate sentences only according to whether such sentences had _actually_ been heard during acquisition. There is little doubt that under such circumstances Ss are attentive not only to the meaning of the sentences, but to various aspects of their form and the context in which they were presented. In particular, it seems likely that Ss have trouble remembering whether they encountered a given simple idea alone or in one of many possible combinations, even though they are certain that it _means_ the same thing as something they have heard earlier.
>
> The purpose of the present study was to determine whether the Bransford and Franks (1971) results can be obtained when nonsemantic criteria are disregarded and recognition is based only on meaning [p. 79].

In order to accomplish this purpose Katz reproduced the Bransford and Franks (1971) study with two sets of instructions: actual instructions (AI), those used in the Bransford and Franks study, which asked the students to give a recognition-rating response to a set of sentences if they thought they heard them during acquisition; and meaning instructions (MI), which required a

response based on congruence of meaning with the acquisition sentences.

Katz's hypotheses flow from the Bransford and Franks results and the concept of semantic integration. First, he reasoned that students should not be able to discriminate between OLD and NEW sentences in either the AI or MI conditions. Second, the 'linear effect' of recognition-rating responses and sentence complexity should provide the same slope in both the AI and MI conditions.

The materials were derived from four complex embedded English sentences each analyzed into four simple declaratives, similar to the Bransford and Franks materials. For example, The rock which rolled down the mountain crushed the tiny hut at the edge of the sea, was analyzed into, The rock rolled down the mountain, The rock crushed the hut, The hut was tiny, and The hut was at the edge of the sea.

The results confirmed the Bransford and Franks data that students could not discriminate between OLD and NEW sentences. More important were the results on the 'linear effect'. In the AI condition the 'linear effect' of recognition-rating responses to sentence complexity confirmed the Bransford and Franks results,

that students integrate related ideas into a unified representation. This effect intensified from one trial block to a second trial block. However, for the MI condition there was a slight effect in trial block 1 but virtually a flat curve in block 2.

Katz was still at a loss to describe the differential processes operating under the AI and MI conditions, since the data were not revealing on the matter. However, he defers to the same tentative suggestions offered by Reitman and Bower (1973).

> If an individual has fully integrated and memorized a complex idea, he can 'recognize' any part of it. In light of this, the explanation for the linear effect offered by Bransford and Franks (1971), namely, that recognition is a function of the number of basic ideas in the integrated idea [that is] exhausted by the semantic representation of the input is unrealistic. If true, it would mean that individuals are literally unable to recognize small components (basic ideas) of a holistic idea in isolation from the whole [p. 83].

Citing other empirical evidence to support this tentative, exploratory hypothesis, Katz referred to his own doctoral thesis (1971) where, although there were minor variations, Katz used the 'Bransford Paradigm' to investigate various aspects of the semantic integration process. It was concluded that meaning was used by the students in the processing of and recognition of NON-CASEs. However, the processing of high imagery and low

imagery materials, and the ensuing results, suggested that the 'linear effect' of sentence complexity on recognition-confidence responses was the result of integrating perceptual information. After all, the effect was obtained with low imagery material that was virtually meaningless (e.g., <u>The original event created an unusual state in the structure of the system</u>).

Katz also cited the recent findings of Reitman and Bower (reviewed above) where n-tuples of letters and numbers in rote, meaningless concepts were used to obtain the 'linear effect'. Finally, conceding that the exact cause of the 'linear effect' was not known, Katz suggested that it appeared to be influenced by instructional artifacts.

An ancillary finding was that the confidence ratings were superfluous scores, and numbers of recognition responses, per se, were adequate to both analyze the data and find the same results.

Katz and Gruenewald (1974) took issue with the interpretations of Bransford and Franks (1971, 1972) that their results supported the interpretation (1) that there was constructed a unified semantic representation in the comprehension of connected discourse; and (2) that the recognition of new linguistic strings was a

function of the extent to which those input strings exhausted all of the information in that unified semantic whole.

They suggested that the results might have nothing to do with semantic processes, since Reitman and Bower (1973) were able to replicate the 'linear effect' with letter and number n-tuples, since Katz (1971) showed the same results with virtually incomprehensible material, and since Katz (1973) was able to show its disappearance by changing the instructional set.

Katz and Gruenewald (1974) wanted to test directly whether or not the 'linear effect' was based on sentence meaning. They compared the sentence complexity curves of sentences that were meaningful with sentences that were virtually incomprehensible. Meaningful (M) sentences were similar to the original Bransford and Franks (1971) study while meaningless ($\overline{M}$) sentences retained syntactic structures but eliminated lexical meaning.

The materials were four M and four <u>M</u> complex sentences. The M sentences were similar to those referenced in earlier studies and each was analyzed into four simple declaratives. For example, <u>The rock which rolled down the mountain crushed the tiny hut at the edge of the sea</u>, was analyzed into <u>The rock rolled down the</u>

mountain, *The rock crushed the hut*, *The hut was tiny*, *The hut was at the edge of the sea*. The $\bar{M}$ sentences were derived from four M sentences by a rather detailed procedure employing rules outlined in Forster (1966) and Underwood and Schulz (1960). The corresponding $\bar{M}$ sentence and its simple declaratives for the M sentence given above were, *The soto which tehoed inow the feexteva voneed the tioc len ic the froo ab the adex*, *The soto tehoed inow the feexteva*, *The soto voneed the len*, *The len sim tioc*, and *The len sim ic the froo ab the adex*.

As in the earlier studies, four ONEs, four TWOs, three THREEs, and one FOUR (a set of 12 sentences) were generated from each frame. The 12-sentence idea group was used to generate the acquisition and recognition lists. In addition, two M between-frame NONCASEs and two $\bar{M}$ between-frame NONCASEs were generated. The acquisition list contained two each of ONEs, TWOs, THREEs, and two FOURs, totalling 26 sentences. The recognition list had 32 sentences: three OLD sentences from each frame, a ONE, TWO, and THREE; three NEW sentences from each frame, a ONE, TWO, and THREE; the four FOURs (two OLD and two NEW); and the NONCASEs were added.

The acquisition phase was posed as a short-term memory task with interpolated activity between exposure to and writing down each sentence. The students were naive to the recognition task. After a two-minute rest the students went through a recognition deck and gave "Yes" or "No" recognition responses and confidence ratings.

The results indicated: (a) Using the recognition-confidence scores as the dependent measure, a significant 'linear effect' was found for both the M and $\overline{M}$ sentences. (b) There was an interaction between sentence complexity and meaningfulness, whereby the slope of the $\overline{M}$ condition, while significant, was not as steep as the M condition. However, this difference in slopes was demonstrated to be an artifact of greater difficulty in the $\overline{M}$ condition and does not mitigate the subsequent interpretations of the data.

The authors concluded that the 'linear effect' found in the $\overline{M}$ condition could not be attributed to semantic integration. The question that remained was: What was the cause of the 'linear effect'?

> We offer the relatively simple hypothesis that... Ss...adopt a guessing strategy in which they assign ratings, not according to what they have experienced (or integrated), but on the basis of the probability that a sentence having a particular number of

> "components" (i.e., ideas, phrases, or anything Ss interpret as being manipulated combinatorially) could have occurred in acquisition. They do this presumably, by estimating the total possible number (or "set size") of sentences of varying sentence complexity and then formulating a probability based on the inverse of that set size. The linear effect itself can thus be easily explained: Since the maximum set size is greatest for ONEs and least for FOURs, the probability estimates and, hence, recognition confidence ratings would yield a reverse ordering, i.e. ratings would be greatest for FOURs and least for ONEs...[p. 740].

Recall that the guessing strategy was suggested by Reitman and Bower (1973) and Katz (1973).

Katz, Atkeson, and Lee (1974) attempted to test the guessing strategy hypothesis further. They hypothesized that what students were learning was not necessarily integrated semantic content, but the idea that the acquisition sentences appear in various combinations. If this was so then the increased misrecognition of sentences of greater complexity may be an outcome of a guessing strategy; that is, that a particular combination of ideas, probably did occur in acquisition.

The procedure closely paralleled that of Bransford and Franks (1971). The Control condition received one each of the ONEs, TWOs, THREEs, and FOURs for each idea group, totalling 16 sentences. The recognition list was identical to the acquisition list, though in a different random order.

The Experimental condition received only the 16 ONEs in acquisition and recognition, with different random orders of the same sentences for acquisition and recognition. For the Experimental group there was no reason to expect that the sentences would be combined.

> According to Bransford and Franks...either [condition]...should...result in the formation of wholistic semantic representations...Thus, recognition-confidence ratings for old ONEs ought to be identical...According to the procedural artifact hypothesis...ONEs in the experimental condition should be significantly higher because the Ss are no longer required to remember what the precise combinations of ideas were in acquisition. They need only remember whether they saw the ideas at all [p. 712].

Acquisition was an incidental learning task, although the students were told that the sentences were related. The students were naive to the recognition task. Each sentence was presented orally only once, followed by a brief interpolated task, and then the students wrote down the sentences. After all 16 sentences, they rested for five minutes. The mean recognition-confidence scores were used as the dependent measure.

The results indicated that the control condition replicated the findings of Bransford and Franks (1971). The crucial comparison was between the mean responses of the ONEs for the Experimental and Control conditions.

The Experimental condition elicited significantly higher scores. There were no differences in the recognition-ratings of NONCASEs. The results supported the artifact hypothesis.

To explain the results of the 'linear effect' as found by Bransford and Franks (1971), Katz, Atkeson, and Lee (1974) invoked their guessing strategy hypothesis which was elaborated upon earlier. Although students tried to make exact identification, statistical probability estimates are made in an attempt to precisely identify a sentence during recognition.

Moeser (1976) presents a lengthy series of four experiments which bear directly upon the topic under discussion. Moeser discusses Tulving's (1972) description of a semantic memory system and an episodic memory system. Though it will not be elaborated upon here, the semantic memory system encodes new information so that it will be organized with prior knowledge or rule structures, for the generation of logical consequences, interrelations, and new information that was not actually presented to the learner (i.e., inferences). This was in contrast with the episodic memory system which was a representation of unique events and which, theoretically, could not generate new inferential information.

Moeser (1976) was concerned with reasoning in the episodic memory system. It was assumed that the 'Bransford Paradigm' results were operating off an episodic memory system.

The first experiment by Moeser was designed to investigate whether various encoding conditions would affect a student's ability to integrate a set of basic premises into complex sentences. There were three experimental encoding conditions. In the first (holistic sentences group), 'complex' sentences were presented as a syntactic unit: The ants ate the sweet jelly on the table in the kitchen. In the second (ordered propositions group), the same information was presented sequentially in a set of simple sentences: The ants ate the jelly. The jelly was on the table. The table was in the kitchen. In the third (ramdom proposition group), the simple sentences were presented non-sequentially so that the semantically related sentences were temporally separated.

It was hypothesized that memory performance of the actual materials presented for encoding should not differ for the three conditions. However, memory performance for the 'complex' sentences should decrease from the first to the last condition. Thus, the first group was

supplied with the actual 'complex' sentences while the second and third groups had to attain the 'complex' sentences which were not explicitly stated.

Four complex sentences comprised of three basic premises were constructed. The holistic and ordered groups were presented the ideas, in acquisition, in the same order. Each idea was presented twice. The random group received the same number of exposures of the basic simple sentences as the ordered group, but with appropriate blocking and randomization.

The test was a forced-choice (see Anderson and Bower, 1973) situation which required the discrimination between true and false complex sentences and true and false basic premises.

Children were presented the sentences one at a time, were asked to repeat the sentences, and then were asked questions about the sentences (e.g., Who ate the jelly?). They were initially told that they were going to be tested on how well they could remember. However, the children were naive to the test situation at the end of the total presentation sequence.

The results were as predicted: Memory performance for actual encodings were identical across groups, while identification of complex sentences declined from the

ordered proposition group to the random proposition group.

The purpose of Experiment II was to determine if the logical connection between discrete semantically related sentences was encoded into the memory system. This would be determined by examining memory performance for complex sentence chains under two conditions: with and without the logical connective as a cue.

For example, consider the acquisition sentences: The doll was in the small crib; and The small crib was under the tree. The connective is the word crib. On a forced-choice test, one group (the TRANSITION group) would choose between: The doll is in the crib under the tree; and The doll is in the crib on the sidewalk. Another group (the NO TRANSITION group) would choose between: The doll is under the tree; and The doll is on the sidewalk.

> If the memory representation for the information is such that a logical connection exists between two propositions, the subjects given the coordinating cue should find it easier to retrieve the [complex sentences]...than those not given the cue. If no logical connections exist among the memory representations, the coordinating cue should not facilitate the making of the [complex sentence]...[p. 199].

Eight problems were constructed as sets of four sentences. An example of a complete problem set would

be: The doll is in the small crib; The toy truck is on the sidewalk; The small crib is under the tree; The clown is riding the donkey. Each problem set contained parallel alternatives -- for example, sidewalk and tree in this set. A line drawing illustrating the relationships in the sentences was presented with each sentence.

Following the presentation of each problem set the student was asked one of two question sets. One question set tested memory for two explicitly-stated propositions; the second question set contained one complex chain question and one memory question. Each question was of the force-choice type.

The results indicated that memory questions elicited better performance than complex chain questions. However, there were no differences between the TRANSITION and NO TRANSITION conditions. The students did not appear to be able to make use of the coordinating cue in the premises. It appears that the common element for two premises was differentially represented in memory for the students in the two conditions.

In Experiment III, the purpose was to see if students enhance their complex chaining as a function of presentations of individual premises to force the motivation to make semantic relations.

Materials similar to Experiment I were used and presented in either an Ordered proposition condition or a Random proposition condition as in Experiment I. The students either received two or six presentations of the materials. The criterion task consisted of 16 forced-choice items: four tested complex chaining and 12 tested memory for explicitly-presented propositions. The procedure was similar to Experiment I and the subjects were naive to the criterion task.

The results indicated that the Random group was not able to handle complex chaining at all. The Ordered group was able to produce chains. However, repetitions had no effect on specific memory or complex chaining.

The purpose of Experiment IV was to see if integrative encoding is facilitated by providing a miniature cognitive structure which might be used to encode new information. Students were provided with a short, two-sentence story before exposure to the acquisition materials.

The results of the preceding three experiments showed that it was not always possible for students to encode information into an organized structure. In contrast to the previous experiments, Experiment IV was an intentional rather than incidental learning task.

There were two conditions. In the STORY condition, the students were required to memorize four stories. In the NO STORY condition there was no prior activity. There were two additional control groups to ascertain the influence of the mini-stories alone on criterion performance.

Four complex ideas were constructed similar to those used in the prior experiments. For example: <u>The ants ate the jelly on the table under the tree</u>. The short story used in the STORY condition as a miniature cognitive structure for the aforementioned complex sentence was: <u>It was very hot inside the house, so the family planned to have dinner in their shady back yard. But after carrying out the food, they decided to go for a swim in the pool before they ate</u>.

After memorizing the four stories in the STORY condition the students were instructed:

> You will now hear a number of sentences, each of which is related to one of the stories you have just learned. There will be 12 sentences in all, and each story has three sentences which are related to it. You are to combine each sentence with its related story, and later you will be tested on how well you have been able to do this. The sentences will be presented a number of times in order to insure that you will be able to relate them all [p. 207].

In the NO STORY condition, the instruction differed as follows:

> You will hear a number of sentences, some of which are related to one another, and you are to combine together those sentences which have the same words in them so that they form one complex idea. Later you will be tested on how well you have been able to do this. There will be 12 sentences in all, arranged so that there are four sets of complex ideas and three sentences in each complex idea. The sentences will be presented a number of times in order to ensure that you are able to relate them all [p. 207].

For both conditions, the acquisition sentences were presented six times in the same random sequence. These were followed by 4 forced-choice chaining questions, and 12 forced-choice memory items, similar to those in Experiment III.

The control groups indicated no undue influence of the stories, per se, on criterion performance. However, this critical observation must be made. There should have been a control group to test whether memorizing prior stories, per se, had any effect on criterion performance. The experimenters must distinguish between the effects of a miniature cognitive structure, plus instructions, on the one hand, and developing a learning set, on the other.

The results indicated no difference between the STORY and NO STORY conditions on memory items. However,

the STORY condition showed superior performance on the chaining items. Although students were encouraged to form complex chains, they were unable to do so when the sentences were given without additional context.

Moeser seems to be making a strong case for the fact that the 'Bransford Paradigm' materials were stored as discrete episodes, and only with great difficulty was there evidence of integration. The students needed a lot of help. Either give them the entire integrated chain in one complex sentence, order the simple sentences logically, or give them an integrative context. Apparently, people do spontaneously integrate information in some way (as evidenced by the 'Bransford Paradigm' studies). However, if they are given a forced-choice test, they show lower levels of integration. This speaks for memory decrement, poor discrimination, and inappropriate generalizations, all of which were evident in the 'Bransford Paradigm' studies because a non-forced-choice recognition test rather than a forced-choice test was used. This supports the Katz, et al. criticism that the 'Bransford Paradigm' results were artifact and not integration. After all, a transitional cue in Experiment II did not help at all. In Experiment IV, merely instructing students to combine related

sentences did not produce integrative chains beyond chance performance.

A General Critique of the 'Bransford Paradigm' for Semantic and Linguistic Integration

In the 'Bransford Paradigm', semantic integration was merely the combinatorial manipulation of linguistic units into sequences of varying lengths. These apparent juxtapositions of short discrete elements into long strings were called abstractions and wholistic ideas.

The 'Bransford Paradigm' research could demonstrate that unified wholistic strings were acquired (though not necessarily connected with prior knowledge), but the products of integration were not necessarily at a higher level of abstraction than any of their elements. Transfer was never demonstrated except under the guise of stimulus generalization and poor discrimination. Further, the underlying processes appear not to be semantic.

Further, and this is most important, the 'Bransford Paradigm' was an incidental learning task. The students were naive to the conditions of the recognition task. They might have been instructed that the sentences relate to one another; they were never (with the exception of Moeser, 1976) instructed to integrate them, to make the inductive leap to extrasentential inferences, or to

draw conclusions as a deliberate information processing strategy. It will be apparent in further studies, yet to be reviewed, that the effects of such an instructional set have implications for determining personal control over semantic integration. However, no study systematically investigated instructional variation on semantic integration. Instructions within different paradigms were constant. The experimental question remains as to how strategies can change with instructions.

The Bransford and Franks (1971) material was concrete and easily imaged. The FOURs were just as concrete and susceptible to imagery as the ONEs. Therefore, on what basis does one say that the FOURs were linguistically abstract relative to the ONEs? In my opinion, what they called semantic integration or linguistic abstraction (which implies some order of conceptualization) was merely a sequencing of distinct units that maintained their logical interrelationships. The logic of the relationship came either from the lexical interpretation (deep structure) of the unit elements and/or the syntactic constraints they contain. The juxtapositioning and logical sequencing could be carried out in a non-semantic manner. The end result was no more or less abstract or concrete than the elements.

Begg and Paivio (1969) suggested that concrete material was imaged and abstract material was processed as strings of sequential dependencies. Since Bransford and his colleagues showed that concrete and abstract material yield similar results in their experiments, the conclusion was that part meanings were integrated into whole meanings for concrete and abstract input according to some semantic process. However, if concrete and abstract materials were indistinguishable in terms of the experimental effects, why assume that the effects were the outcome of a semantic process? Why not conclude that both sets of data were processed according to Begg and Paivio's description of abstract verbal processing and storing: string processing by sequential dependencies (see Miller and Selfridge, 1950)?

Singer and Rosenberg (1973) suggested that the correlation of recognition to sentential complexity was an artifact of grammatical form and function of the recognition cue, and that the correlation of confidence ratings to sentential complexity does not exist when the recognition response is ignored. Franks and Bransford (1974) responded to those criticisms but the question remains far from resolved.

Reitman and Bower (1973) using n-tuples of alpha and numeric characters found a specific recognition

memory for exemplars and a storage for the full concept unit. They tentatively hypothesized that the recognition of any exemplar follows a probabilistic decision process which is non-semantic.

Katz (1971, 1973) suggested that the 'linear effect' was due to perceptual integration. If a concept was learned, then any part could be recognized with equal frequency. If a sequential string was learned, then recognition-ratings would follow a probabilistic model. Katz and Gruenewald (1974) provided even stronger evidence that the results of the 'Bransford Paradigm' may have nothing to do with semantic processes at all, since the 'linear effect' was obtained for both meaningful and meaningless sentences. They formalized the decision theoretic approach into the 'guessing strategy' hypothesis.

Katz, Atkeson, and Lee (1974) were able to show that semantic integration, as manifested by the 'linear effect' of Bransford and Franks was an artifact of the method of stimulus presentation. Finally, Moeser (1976) was able to support the artifact over the integration explanation for the results of Bransford and Franks by using a far more sensitive criterion task (forced-choice) in a series of experiments.

The concept of semantic integration still remains alluring and intuitively appealing in applied educational psychology. But, some serious criticisms have been raised. Perhaps the 'Bransford Paradigm' data supports some other non-semantic process as suggested by Katz and Gruenewald. The question remains open.

Studies Investigating Developmental Effects on Semantic Integration

Donaldson (1963) suggested that children will commonly encode the elements of a potentially integrative task as separate units. However, we might expect older children to change their encoding strategy to an integrative orientation.

Smedslund (1963) designed a study to compare the assertions from informal observations by Piaget, Inhelder, and Szeminska (1960) and an experiment by Braine (1959), where Braine reported that transitivity (a kind of integration) occurs at least two years earlier than Piaget, et al. suggested. In a study on concrete transitivity of length inequalities (A is longer than B, B is longer than C, therefore A is longer than C), though, Smedslund concluded that the average age of acquisition of concrete transitivity of length was probably around eight years, and that Piaget was supported over Braine.

However, Smedslund was careful to make the following observation that reflects on the subject matter of this proposal: "Concrete transitivity, involving inferences from actual observations, should not be confused with formal transitivity which permits the subject to make inferences from verbally stated, hypothetical premises [p. 389]."

Kendler and Kendler (1967) presented another study of children's ability to integrate learned behaviors.

> When [problems that require the integration of separate segments are]...applied to a cross section of human beings, we find that if we use either grade level or MA as a developmental index, solutions are very infrequent at the lower developmental levels [kindergarten]. However, as developmental level increases, solutions become increasingly frequent until at the highest levels, they are the overwhelmingly dominant mode of response [p. 186].

The tasks, briefly stated, were to press a button on a red panel to obtain a glass marble, to engage in an interpolated behavior, and to place the glass marble in a slot to obtain a toy. The students on the integration task were asked to obtain a toy. The integrated sequence of events was obvious to the reader, but not so obvious for the less mature students.

Paris and Carter (1973) felt that the integrative memory hypothesis had received support in the work of

Bransford and Franks (1972) on adults. They extended the research to determine if children also demonstrated integrative memory ability and to explore age-related differences.

For acquisition, seven stories were constructed. Each was comprised of three sentences, two premises and a filler. For example, The bird is inside the cage; The cage is under the table; and The bird is yellow. The entire acquisition list was read aloud to the individual students. They were instructed to remember the sentences since they would be asked questions about them. After the acquisition phase, the students engaged in a five-minute interpolated task. The students were naive to the task of drawing an inference.

The recognition materials consisted of four sentences for each of the seven stories: a true premise, The bird is inside the cage; a false premise, The cage is over the table; a true inference, The bird is under the table; and a false inference, The bird is on top of the table. The students were required to give recognition-rating responses as in prior studies.

The results indicated that the students consistently misrecognized true inferences far more frequently than they erred on other test sentence types. The pattern of

results was the same for both second and fifth grade children, except that the older children made fewer overall errors. There were no main or interaction effects with sex. The certainty judgments indicate that the students made their errors in the misrecognition of valid inferences with strong conviction.

It will be briefly mentioned here, that: (a) instructions were not varied regarding the nature of the criterion task; (b) it was not apparent from the overt materials whether they suggested extra-linguistic inferences; and (c) no developmental differences were found. This study suffers from many of the same criticisms of the 'Bransford Paradigm' studies.

Barclay and Reid (1974) investigated developmental trends in semantic integration from kindergarten through the fifth grade. As such they concentrated on two aspects of language comprehension and memory: (a) memory for a particular syntactic structure, the truncated passive sentence (e.g., The ball was hit); and (b) semantic integration. Semantic integration was described, in a very general way, as the distribution of syntactically isolated information throughout the content of a passage.

Barclay and Reid (1974) felt it was important to know whether and to what extent semantic integration improves in childhood. However, in a review of the

literature, Barclay and Reid could find no hard evidence of an increase in semantic integration with age in early childhood.

In a review of the literature, Barclay and Reid noted that full passives were often recalled or recognized as full actives (e.g., Mehler, 1963; Sachs, 1967). However, Slobin (1968) found that while full passives might be recalled or converted to full actives, truncated passives were usually immune from this transformation and recalled verbatim.

The means of assessing semantic integration was the observation of the syntactic conversion of truncated passives (e.g., The ball was hit) to full passives or full actives when the deleted actor was mentioned in a later sentence. The interpretation was that the information in the truncated passive was semantically encoded and distributed to each sentence in a passage of discourse so that the later-mentioned actor was integrated with the earlier-mentioned truncated passive. This integration (by virtue of the availability of an actor) was the necessary condition for the syntactic conversion to take place. By observing the conversion it was inferred that integration took place.

The following are examples of the materials. TRUNCATED PASSIVE/ACTOR: On the first day of school Bob was introduced to his new teacher. The principal who introduced him was very nice. TRUNCATED PASSIVE/NO ACTOR: On the first day of school Bob was introduced to his new teacher. All his friends were glad to see him. FULL PASSIVE: On the first day of school Bob was introduced to his new teacher by the principal. The principal was very nice.

Students were read three passage types in the following order: target sentences that were TRUNCATED/ACTOR, TRUNCATED/NO ACTOR, FULL PASSIVE. In the first passage, the actor was given in the immediately following sentence. The students, participating in a memory game, had to repeat the story from memory.

The results indicated that verbatim truncated passives were recalled most frequently when no actor was supplied, and recalled with a frequency equal to that for actives when an actor was supplied. Barclay and Reid felt that the results supported a semantic integration hypothesis, and supported the psychological reality of semantic content. If syntax, rather than semantics, were primarily the basis for recall, the ACTOR - NO ACTOR conditions should have displayed equal predominance of truncated passives in recall. The opportunity for semantic

integration allowed the students to form a condensed, wholistic version of the story in memory to ease the burden of retention. However, four age groups did not differ significantly in the pattern of results for ACTOR, NO ACTOR, and FULL PASSIVE conditions. Thus, no developmental trend was evident.

The authors recommended further investigations. Integration in their study spanned only adjacent sentences. Other studies might investigate integration over a wider span of sentences for different age groups. It would be necessary to partial out age-related memory span effects from developmentally different linguistic processes. Finally, a legitimate applied school learning issue is whether a link can be made between deficient semantic integration and problems in learning and reading comprehension. The students, participating in a memory game, had to repeat the story from memory.

The results indicated that verbatim truncated passives were recalled most frequently when no actor was supplied, and recalled with a frequency equal to that for actives when an actor was supplied. Barclay and Reid felt that the results supported a semantic integration hypothesis, and supported the psychological reality of semantic content. If syntax, rather than

semantics, were primarily the basis for recall, the ACTOR - NO ACTOR conditions should have displayed equal predominance of truncated passives in recall. The opportunity for semantic integration allowed the students to form a condensed, wholistic version of the story in memory to ease the burden of retention. However, four age groups did not differ significantly in the pattern of results for ACTOR, NO ACTOR, and FULL PASSIVE conditions. Thus, no developmental trend was evident.

The authors recommended further investigations. Integration in their study spanned only adjacent sentences. Other studies might investigate integration over a wider span of sentences for different age groups.

Hale and Piper (1974) presented two interesting experiments on the "effect of pictorial integration on children's incidental learning." At first blush, this study would seem out of place in this critique. However, it was deliberately included for three reasons:

1. The entire study -- its rationale, procedures, and results -- offer an analogue to the previously reviewed work in this section.
2. It addresses a developmental factor which was investigated in the prior study using connected discourse.

3. A critique of the results sows the seeds of a non-semantic process as the basis for an apparent semantic integrative process.

Hale and Piper (1974) cite research in the literature whereby children's learning of incidental stimuli from pictorial events remains constant over age, while memory for central events increases developmentally. The interpretation of the decline in the ratio of incidental to central events was that children would attend more to task-relevant cues with increasing age (Hagen, 1972).

> It is proposed that if the stimulus elements were to be integrated into a scene, a child would more naturally view a compound stimulus as a single unit and would find it a more demanding task to analyze the compound into its separate elements. Materials of this sort should thus permit a more appropriate test of children's ability to analyze stimuli into their relevant and extraneous features and to attend selectively to the former [p. 847].

Pictorial stimuli in four conditions, varying in integration between the components of the pictures were examined by 8, 11, and 14 year olds. There were two action conditions, WEAK and STRONG action, wherein an animal was depicted performing an action with an object; one STANDARD condition, wherein the animal and object were independent entities; and one STATIC condition, with the animal holding the object.

The authors' review of paired-associate studies revealed that children's incidental learning of pictorial associations improved when components were presented in an action relation. The degree of facilitation was found to increase with age.

Thus, incidental learning of animal-object relations was expected to increase with age in the ACTION conditions, but not in the STANDARD and STATIC conditions. The mediational factor underlying this difference was hypothesized to be stimulus integration.

The experiments were divided into a learning phase and an incidental learning phase. In the learning phase, six animal frames in a horizontal array (deer, pig, dog, sheep, horse, and rabbit) were exposed simultaneously for six seconds. The student was then cued with a picture of one animal and asked to point to the correct array position of that animal. Immediate feedback was provided. The <u>central learning</u> score was computed from the number of correct responses.

All students were naive to the incidental learning phase. Immediately following the learning phase the six incidental objects were simultaneously displayed (tennis racket, guitar, bucket, kite, broom, ball), the animals were shown one by one, and the children indicated the correct object with each animal. The number of correct

animal-object associations constituted the _incidental learning_ score.

Consistent with previous findings, the central learning scores increased with age for all conditions. The incidental learning scores, contrary to expectations, did not increase with age and in some cases declined. The incidental learning was greatest in the STRONG action condition, but no age differences were notable.

Hale and Piper concluded that pictorial integration automatically increases the likelihood that a child will recall the components of a stimulus together. But this effect does not increase developmentally; it is of roughly the same magnitude for children throughout the range from childhood to adolescence. By observing the recalling together of stimulus components we might incorrectly jump to the conclusion that what we observe is integration and abstraction of semantic content. The process may simply be the juxtaposition of discrete elements on a perceptual level. If semantic content was not involved, and therefore, semantic integration could not be observed, then why should a developmental trend be expected?

As an aside, it is interesting to note that the integration in the Hale and Piper study was imposed on the students by the construction of the stimulus

pictures themselves. By comparison, the integration in the 'Bransford Paradigm' studies, reviewed so far, was purportedly a product of the mental processes of the students.

In Experiment I of Moeser's (1976) study (reviewed earlier) differences among three experimental encoding conditions on a developmental basis were investigated. It was expected that younger children would have more difficulty integrating information. However, the results were the same at all age levels. There were no developmental effects. In other words, there were no developmental differences in ability to construct complex sentences.

In Experiment II of Moeser's (1976) study, there was a slight but non-significant trend in integrative ability. It is interesting to note that Grade 2 and 6 showed identical performance while, theoretically, great changes in logical reasoning are supposed to take place (Smedslund, 1963).

A Critique of the Studies Investigating Developmental Effects on Semantic Integration

Early studies on integration behaviors in children have been successful in demonstrating developmental effects (Donaldson, 1963; Kendler & Kendler, 1967;

Piaget, Inhelder, & Szeminska, 1960; Smedslund, 1963). However, these studies were limited to the observation of concrete operational tasks while the thrust of this investigation is on the integration of semantic content on a verbal level using connected discourse.

Four studies reviewed in this chapter were unsuccessful in finding developmental trends in semantic integration. In the Barclay and Reid (1974) study, the pattern of results for the TRUNCATED PASSIVE/ACTOR, TRUNCATED PASSIVE/NO ACTOR, and FULL PASSIVE conditions was the same for all age groups. Thus, there were no observable developmental trends. In the Hale and Piper (1974) study central learning increased with age but not incidental learning. Incidental learning was greatest in the STRONG action condition, which suggested some sort of integration of central with incidental stimuli. The effect was the same for all age groups. Thus, there were no observable developmental trends.

Paris and Carter (1973) and Moeser (1976) were also interested in the ability of children to integrate semantic information and, therefore studied age-related effects. While they did find that children could integrate materials in a 'Bransford Paradigm' type of study, and that older children manifested greater memorial behavior,

they found that the pattern of results was the same for all age groups.

At this point it may be profitable to take a second look at the Hale and Piper (1974) study of pictorial integration to determine why a developmental trend might not have been found and to suggest how a developmental effect on semantic integration could be studied more effectively. It was noted that pictorial integration in the Hale and Piper (1974) experiment served only to increase the likelihood that stimulus events would be recalled together. This integration, as elaborated in the extensive critique of the 'Bransford Paradigm' studies, could be attributed to the juxtaposition of adjacent events by virtue of perceptual processes, syntactic constraints, sequential dependencies or other non-semantic operations. Therefore, one would not expect a trend in semantic integration since semantic processes were not involved. With increasing age, discrimination between relevant (central) and irrelevant (incidental) information improves. The discrimination of the central stimulus event from the total stimulus field presupposes an ability to abstract at a higher level than the overt level of the actual stimulus materials. That is why recall of central events increases with age. But incidental stimulus events required no abstraction operation.

Therefore, no recall advantage was observed for higher age groups.

Lastly, the studies which failed to find developmental effects on semantic integration have used the 'Bransford Paradigm' and suffer from the same criticisms, mentioned earlier.

Rationale for studying developmental trends

The failure to note a developmental trend in the aforementioned studies does not obviate the possibility of a developmental effect for semantic integration. A close scrutiny of the above studies suggests that semantic processing was not, in fact, tested. The recalling together of stimulus components, as observed in these studies, was incorrectly labeled integration and abstraction on a semantic level.

However, a better test of semantic integration would occur if inferences beyond the information initially input were tested. This would involve inferences that were neither perceptual nor syntactic, but semantic.

The Inferential Thinking Paradigm for Studying Semantic Integration

The studies to be reviewed below are similar to the 'Bransford Paradigm' of research in that semantic integration was investigated. However, the crucial

distinction between the two classes of research rests upon the description of the integrative process. The next series of studies presuppose that input materials are integrated with existing knowledge or rule structures, and with other input materials, to yield a new product that: (a) represents a wholistic unit; (b) involves some abstractive process; (c) permits the observation of transfer phenomena by way of inferences, extra-linguistic products, and logical conclusions; and (d) assumes that the underlying verbal content to account for the above are propositional (Kintsch, 1970). These studies were more directly aimed at demonstrating the psychological reality of semantic integration and will survive the criticisms of the 'Bransford Paradigm'.

Bransford, Barclay, and Franks (1972) reported that investigations of sentence memory and comprehension have relied on concepts developed within transformational linguistic theory, sometimes called deep structural theory (see Blumenthal, 1967; Blumenthal & Boakes, 1967; Rohrman, 1968; Sachs, 1967). It was felt that a more complete psychological description of the nature of semantic information processing was necessary. Consider the following sentence, <u>Three turtles rested on a floating log, and a fish swam beneath them</u>. Bransford, Barclay,

and Franks suggested that students would make the inference that the fish swam beneath the log and the turtles. However, if the preposition beside was substituted for on that same inference could not be made. Thus we have a potential inference (PI) sentence and a non-inference (NI) sentence.

Three experiments were conducted. In Experiment I 14 sentence frames were constructed such as: Three turtles rested (on/beside) a floating log, and a fish swam beneath (it/them). The 14 frames could yield a possible four sentences each. They differed as being PI or NI and in changing the final pronoun. One sentence was chosen from each frame.

In random order, sentences were read to each student with instruction to remember them because questions would be asked at a later time. The students were not naive to the recognition task. After a three-minute rest, 35 recognition sentences were read by them: 14 originally presented PI and NI sentences, an additional 14 sentences which were the final pronoun counterparts to the originals, and seven filler (F) sentences. The students were asked to give recognition-rating responses as in the previous studies by Bransford and Franks.

The mean recognition-ratings were compared for the following six categories of items, NEW and OLD versus

PI, NI, and F. The results indicated that for PI sentences alone, students could not distinguish those sentences they had heard before from those they had not. These results were interpreted as being consistent with an integrative component in memory.

Experiment II dealt with the problem that the results obtained for PI sentences could have been due only to poor memory or weak discrimination among PI sentences. Thus, a memory decrement, specific to PI sentences could account for the results in Experiment I. The object was to demonstrate that memory decrement for the PI form was specific to the final pronoun and not to the PI form itself.

Two groups received seven PI, seven NI, and seven F sentences. The acquisition procedure was the same as in Experiment I. During the recall phase the students were read sentence prompts, consisting of the noun phrases of each sentence, and asked to recall the rest of the sentence.

The results showed that the PI and NI sentences were paraphrased equally well, when the accuracy of the final pronoun was disregarded. When the accuracy of the final pronoun was required, PI sentences showed a lower rate of recall. These results were consonant with previous findings in Experiment I.

In Experiment III the authors wanted to investigate extra-linguistic inferences that could be made from the construction of a wholistic semantic product out of elementary linguistic units. In other words, the semantic construction from a passage might yield, The chair is to the right of the tree. Then an extra-linguistic product would be, The tree is to the left of the chair.

The materials consisted of six descriptive passages and four filler passages. An example of a descriptive passage is: There is a tree with a box beside it, and a chair is on top of the box. The box is to the right of the tree. The tree is green and extremely tall. Two recognition sheets were constructed. The first (A) had four sentences related to each passage: OLD - original sentence; I - legitimate inference; and two non-legitimate sentences. The second (B) recognition sheet changed the linguistic format and had two legitimate inference sentences and two non-legitimate sentences.

In the acquisition phase the students were told to listen carefully and comprehend them since they were to be asked questions about them. In the recognition phase, to which the students were not naive, task A was given to group 1 and task B to group 2.

The results for set A showed that OLDs were recognized more than Is indicating a tendency to remember

the linguistic form. But, given that the subject could not remember the specific form, the tendency was to recognize an I sentence. For set B the two legitimate inference sentences were recognized equally well and better than the other two types. The results were congruent to those in Experiments I and II.

The authors acknowledged a very important limitation to their study, that nothing much was said about the nature of the semantic descriptions constructed by the students. However, they concluded that the results supported a synthesis of present input with existing knowledge to produce a final product.

Johnson, Bransford, and Solomon (1973) began their study with the assumption that much comprehension and learning of verbal materials involved inferential thinking. The authors referenced the Bransford, et al. (1972) article whereby knowledge of spatial relations was presumed necessary to integrate the semantic content of a passage into a cognitive whole. However, there are many kinds of prior knowledges that influence the inferences that can be made from novel input. The study by Johnson, et al. investigated the production of inferences about relations among objects and about consequences suggested by different operations.

The materials consisted of 20 short descriptive stories. Eight were filler passages. The _experimental_ group had 12 inference stories: six inferred an object relation and six inferred a consequence. An inferred object relation was: _The man was shot_. The inference was: _The man was shot by an instrument presumably a gun_. An inferred consequence relation was: _The boy hit the baseball and watched as it flew into the picture window in the house_. The inference was: _The window was shattered_. For the _control_ group, a prepositional or verb phrase was changed so that an inference was not forthcoming, relative to the recognition sentences.

In the acquisition phase the students were told to listen carefully to the stories as they would be asked questions about them later. The acquisition list was read once by the experimenter. The recognition list had two fillers, 12 OLD sentences (identical to the acquisition sentences), 12 UNRELATED sentences (inconsistent with the acquisition sentences), and 12 INFERENCE sentences used for both control and experimental groups.

The results indicated that the experimental and control students did not differ in the mean number of positive recognition responses given to UNRELATED sentences. The experimental and control students did not

differ on the positive recognition of OLD items. However, the EXPERIMENTAL group responded more frequently to INFERENCE items.

> The results are consistent with the notion that an S's understanding depends not only on what he hears, but on the implications of that information in light of relevant knowledge he already posesses. The Ss were likely to think that information available only by implication was actually given during the acquisition task. [However, on the basis]...of the present procedures one cannot determine whether such implications were actually realized in acquisition or whether they were determined in recognition [p. 204].

It shall be one purpose of this proposed study to determine whether the inferences were generated from and encoded with primary input during the acquisition phase, or whether inferences were made later during the retrieval phase.

Kintsch and Monk (1972) felt that conclusions could be made about the representation of information in memory by observing reading time, time needed to make an inference, and the probability of making a correct inference for two kinds of materials, simple and complex.

The making of an inference is a manifestation of semantic integration in that new information in its propositional form is integrated with rules of inference to yield a new proposition.

> Two versions of a paragraph were prepared. In one, the simplest possible syntax was used, generating a string of simple sentences which expressed in as straightforward a manner as possible, the underlying proposition of the paragraph. In the second version, all conceivable syntactic and semantic transformations were used to generate one long, complex sentence, which however, still expressed the same propositions [p. 26].

It should take longer to read the complex than the simple paragraphs. If the information is stored in a propositional form, the speed of making an inference should be the same for two versions. If information is stored in verbatim form, the complex version should require more processing time to make an inference. The above analysis presumes that simple and complex versions share the same meaning.

Three experiments were conducted. In Experiment I three types of paragraphs were used. Each type had four propositions A, B, C, and D. Type 1 was constructed as follows: A causes B, B and C cause D (or -D); assert A and C (or A and -C); question D. The assertion and response operationally defined the inference. Type 2: A causes B, B causes C, C causes D; assert A (or -A); question D. Type 3: A causes B, C is an instance of A, D is an instance of B; assert C (or -C); question D.

For each problem type four paragraphs were constructed, for a total of 12. Making them simple and complex

netted 24 paragraphs. A sample of a Type 1 Simple paragraph:

> The council of elders in the land of Syndra meets whenever a stranger arrives. If the council meets and if the stranger presents the proper gifts to the council, he is not molested by the natives. The explorer Portmanteau came to Syndra without any valuable gifts [p. 27].

A sample of a Type 1 Complex paragraph:

> The arrival of strangers in the land of Syndra, like the explorer Portmanteau, who did not bring valuable gifts, always resulted in a meeting of the council of elders, which insured that the stranger was not molested by the natives upon receipt of the proper gifts [p. 27].

The question: Was Portmanteau molested by the natives? Other paragraphs covered topics of smog, the price of soap, and crime. Eight filler paragraphs were included for a total of 32.

Two experimental lists were constructed: Set 1 and Set 2. Four warmup problems (filler items), 12 experimental problems, and the remaining four filler paragraphs were organized according to a latin square in terms of Simple-Complex paragraph versus "Yes"-"No" response to the assertion. Set 2 was the complement of Set 1.

The students were told that all questions required an inference rather than a simple answer from memory. They were asked to work fast, but that the important

thing was to be correct. They were further instructed to answer on the basis of overt information, not on logically extraneous factors.

The paragraphs were presented with a slide projector and the advance to the question frame was controlled by the student (unlimited time to read), otherwise the advance to the question frame was experimenter-controlled after a 10-second reading time (restricted reading time).

The results were as follows: Reading time. Given the occurrence of a correct response to an assertion, the mean reading time for Complex paragraphs was significantly longer than for Simple paragraphs. Inference time. For the self-paced condition, inference times did not differ for Simple and Complex paragraphs, whether or not the inference required a "Yes" or "No" response. Proportion of correct responses. The likelihood of a correct response was less under conditions of limited reading time for both paragraphs.

In Experiment II 20 paragraphs were constructed, differing in subject matter, length, logical structure, and type of inference required (classification, comparison, arithmetic problem, and logical reasoning). There were two versions of each: Simple and Complex.

The self-paced condition was the same as in Experiment I except the student's response was "True" or

"False". In the restricted-time group, the experimenter-controlled time was proportional to average time needed in the self-paced group and varied from five to 16 seconds.

The results paralleled Experiment I. Paragraph complexity increased reading time. Inference times, given the occurrence of a correct response, were not influenced by paragraph complexity. The likelihood of a correct response was greater under self-paced than unlimited reading time.

In Experiment III the materials consisted of 12 classical syllogistic arguments. Arguments consisted of two premises plus logically irrelevant material. The conclusion was the question. Half the arguments were valid and half were invalid. The procedure paralleled Experiment II.

The results also paralleled Experiments I and II. Complexity increased reading time, and time to inference (valid or invalid) was the same for Simple and Complex paragraphs. Kintsch and Monk concluded that memorial representation of information was independent of the linguistic complexity of the input. This was presumed to be the propositional base of the verbal material.

Kinstch and Monk acknowledge, however, that it was uncertain when the inference was made and encoded. Was

the inference made during acquisition and encoded along with its generative base? Or was the generative base only processed and stored during acquisition and the inference made later during the retrieval phase? This shall be studied in this investigation.

In a study by Anderson and Ortony (1975), the authors deferred to a description of language comprehension as typified by the writings of Bransford, Barclay, and Franks (1972) and Bransford and McCarrell (1974). They suggested that representational and symbolic experience were more elaborate and particularized than the surface or deep structure features of sentences allowed us to infer. In their words:

> ...[R]epresentation is generally more detailed than the words in the utterance might appear to entail; that the words only loosely constrain the representation; and that one's store of knowledge about the world and analysis of context are heavily implicated in sentence comprehension and memory [pp. 167-168].

The purpose of the study by Anderson and Ortony (1975) was to demonstrate that elaborating and particularizing were routine and predictable in language comprehension. Integration of prior knowledge, context and input were presumed to be the basis of the elaboration and particularization.

First, the authors spend a good deal of time analyzing, critiquing and contrasting their own views with those of Human Associative Memory -- HAM (Anderson and Bower, 1972; Anderson and Bower, 1973). As an illustration they considered the following situation. A cue was used to retrieve a previously-encoded sentence. A cue which was characteristic of the integration of context, prior knowledge, and input would retrieve the sentence with a probability greater than two partial cues: one related only to the subject phrase and one related only to the object phrase. HAM would predict a cue that was integrated to the whole sentence would elicit recall no more frequently than the sum of partially related cues.

The materials and method were elaborate and complicated. The following was an example of a sentence set:

A. Nurses are often beautiful.
B. Nurses have to be licensed.
C. Landscapes are often beautiful.
D. Taverns have to be licensed.

There were two cues: _Actress_ and _Doctor_. For _Actress_, sentence A was the target sentence because the cue was contextually related to the whole sentence. B and C are subject and predicate controls, that is, related to only part of those sentences. D is a double control, related to neither subject nor predicate. For the

cue Doctor B is the target, A the subject control, D the predicate control, and C the double control. Doctor and Actress were either remote or close cues relative to the sentences used. A target and control sentence from each of 14 sets were assigned to list A or list B depending on the cue used. Each list was divided into two sublists.

Students were instructed to learn the sentences. Each student received two presentations on a tape recorder of sublist 1 and then two tests on sublist 1. The first test examined learning by cueing recall with the subject noun. The second test examined recall by cueing with the remote and close cue. The two presentations and two testings were repeated with sublist 2.

There were four levels of scoring from almost strict verbatim to loose derivation. The interrater reliability was .97 to .99. The dependent variable was the proportion of target sentences recalled on test 2, conditional upon demonstrated learning on test 1.

The results showed that at every level of scoring, the probability of recalling a target sentence with the context-relevant cue was greater than the sum of the probabilities of recall for the subject and predicate cues.

Thorndyke (1976) states that theories of discourse comprehension have rarely addressed the issue of

inferential thinking and its role in providing an integrating context for the understanding of incoming information. Thorndyke suggests two basic processes which account for the generation of inferences.

One process is where a sentence, taken in isolation, is given a frame of reference or context by the learner from which plausible inferences are generated and held in immediate memory. The frame or context is either given directly to the learner or the learner supplies one from his general storehouse of knowledge (Bransford & Johnson, 1973; Dooling & Lachman, 1971; Thorndyke, 1975). The generated inferences are used to evaluate new information yet to be encountered in a paragraph.

The second process is where later information must be reconciled with earlier information in the absence of direct and obvious connections. Inferences are generated to fill in the gap between later and earlier assertions to make them logically compatible or at least plausible (Clark, 1975). The inferences from the second process might reinforce, invalidate, or be unrelated to inferences from the first process within the same paragraph of discourse.

The first process described might be called early inferencing; the second process might be called later inferencing.

The predictions tested in Thorndyke's experiments were: (a) the plausibility of early inferences are increased by later inferences that are reinforcing; (b) the plausibility of early inferences are reduced by later inferences that are invalidating; and (c) the plausibility of early inferences are unchanged by neutral later inferences. The reinforcing, invalidating, or neutral influences of the later inferences are a function of two things: the sentence which generated the early inferences (called the target sentence); and the subsequent sentence which generated the later inferences (called the continuation sentence).

In Experiment I four unrelated narrative passages of 20 sentences were constructed. Each passage contained two target-continuation sentence sets. For each target-continuation set, three early inferences were constructed that were equally plausible with respect to a control continuation sentence. That is, the later inferences from the continuation sentence were neutral in their influence on the early inferences. The same early inferences were differentially influenced (reinforced, invalidated, neutral influence) by an experimental continuation sentence.

An example of a target sentence was: (A) The hamburger chain owner was afraid his love for his french fries would ruin his marriage. The postulated early inferences were: (B) The hamburger chain owner got his french fries for free; (C) The hamburger chain owner's wife didn't like french fries; (D) The hamburger chain owner was very fat. The experimental continuation sentence later in the passage was: (E) The hamburger chain owner decided to join Weight-Watchers in order to save his marriage. Relative to the experimental continuation sentence, the plausibility of early inference B was not affected (neutral); the plausibility of C was reduced; and the plausibility of D was reinforced.

The control continuation sentence was: (F) The hamburger chain owner decided to see a marriage counselor in order to save his marriage. The postulated later inference chain would be: (A) The hamburger chain owner was afraid his love for his french fries would ruin his marriage. (G) He wanted to save his marriage. (H) He thought a marriage counselor would help save the marriage. (F) He decided to see a marriage counselor in order to save his marriage. The perceived plausibility of B, C, and D remain unchanged relative to the control continuation sentence F. The authors collected normative

productions and ratings of plausibility to verify the appropriateness of the preselected inferences.

There were two conditions: Experimental and Control. The Experimental group received the experimental continuation sentence E, the other the control continuation sentence F. As indicated above, relative to the continuation sentence E, the early inferences were eithe appropriate, inappropriate, or neutral. Relative to the F continuation sentence, they were all neutral.

The subjects read each of four passages, paused after each target sentence and continuation sentence and wrote down three possible inferences with unlimited time. After this exercise they rated the plausibility of the postulated set of inferences on a seven-point scale.

As expected, the probability of producing the appropriate, inappropriate, or neutral inferences given after each target sentence was the same across inferences and across Experimental and Control groups.

Also as expected, the probability of producing the appropriate, inappropriate, and neutral inferences after the continuation sentence was the same across inference types for the Control group. However, for the Experimental group the probability was lower for the inappropriate inference and higher for the appropriate inference.

Further, the plausibility ratings were the same across inference types for the Control group, while the Experimental group showed lower ratings for the inappropriate inferences and higher ratings for the appropriate inferences.

In Experiment II it was hypothesized that inferences generated during comprehension were encoded in memory along with the text information from which they were generated, and subsequently were undifferentiated from that information.

On a recognition test of sentences that included sentences presented during acquisition and the preselected target sentence inferences, the misrecognition of inferences would indicate the encoding of inferences and lack of differentiation from acquisition sentences. Also, it would be expected that appropriate, inappropriate, and neutral inferences would be misrecognized (false alarm) with equal probability for the Control group. The Experimental group should produce greatest false alarm rates for appropriate inferences and the least for inappropriate inferences.

The students read each of the four passages and rated comprehensibility, imagery, and meaningfulness after each one. Afterward, four recognition tests were

administered. Each test consisted of six true sentences and two sets of three inferences for each target sentence. The students indicated whether a sentence actually appeared in acquisition or was an inference about the story.

> As predicted, inference type had a marked effect on false alarm rates. For the Experimental group, the probability of a false alarm was highest for Appropriate inferences at .58, .40 for Neutral inferences, and lowest for Inappropriate inferences at .06. For the Control group, false alarm rates were nearly equal for the three inference types... [pp. 443-444].

The above experiments strongly suggested that much inferential thinking took place during acquisition, and the inferences were encoded into memory at that time. This proposed study will examine whether or not inferences were indeed encoded in acquisition as opposed to retrieval.

Summary of the inferential thinking paradigm studies

In contrast to the definition of semantic integration implicit in the 'Bransford Paradigm', semantic integration can be characterized in at least two other ways that are closely related. Integration may involve assimilation of input into a previously acquired schema to produce a new schema (a modified schema). Also,

integration may be observed where a previously acquired propositional system influences new input to allow deduction, induction, conclusions, or inferences to be made. However, the propositional base itself is not modified.

The inferences investigated in the Bransford, Barclay, and Franks (1972) Experiment III relied on 'universal' knowledge of spatial relations. Both integrative and specific recognition memory were supported. Students could not discriminate between any two legitimate inferences; but there was a discrimination between OLDs and inferences. Further, investigation showed that the superior recognition for potential inference sentences over non-inference sentences was not attributable to poor memory or weak discrimination.

The Johnson, et al. (1973) study assumed the prior existence in cognitive structure of concepts that were primitive and well established in our culture, in general, like spatial relations. They found that an experimental group could recognize an inference if it was forthcoming from the original acquisition list while a control group could not. This was not really an experiment; it was more a demonstration. But, it was an

indication of an integration of new input with what the student already knows.

Kintsch and Monk (1972) concluded that syntactically SIMPLE and COMPLEX sentence structures yield the same meaning, since inference time did not differ in a self-paced reading condition. They suggested that the results were better explained by a propositional base of sentence meaning.

In the Anderson and Ortony (1975) study integration involved the melding of prior knowledge, context, and input; this integration was key to sentence comprehension. The knowledge of the world that was important in the interpretation of the data for Anderson and Ortony was comparable to the knowledge of spatial relations in the Johnson, et al' study, permitting semantic integration to take place. This type of semantic integration was contrasted to non-semantic integration which may be a function of perceptual integration, grammatical form and function, syntactic integration, and sequential dependencies.

In the most far-reaching study to date, Thorndyke (1976) showed how the generation of inferences played an important role in the comprehension of connected discourse. In a rather ingenious set of experiments

Thorndyke demonstrated a two-fold role of inference generation. Inferences were generated from new input to access a general framework or context. If no framework was found then a backward inferential chain was generated to find some anchoring point so that the total paragraph would be integrated. Sentences were integrated with an early generated inferential chain to find a contextual framework, or were integrated with a later inferential chain to an antecedent.

Rationale

Is semantic integration a real phenomenon that is a cognitive process and is important to school learning or is it a perceptual phenomenon and artifact of experimental materials, instructions, procedures, and criterion tasks? The objective of this study is to obtain support for semantic integration and thereby indirectly support the psychological reality of semantic content. Semantic content is what Kintsch (1972) refers to as propositional representations which are distinct, though not mutually independent, from surface and deep structure.

The objectives of this investigation

The objectives of this research are to determine the following:

1. Are inferences encoded during acquisition or are they processed from the primary input during the retrieval phase?
2. Is there a developmental trend in the semantic integration process, as indicated by time to produce an inference on query, time to encode input material, and production of a correct inference?

Rationale for using inferential materials

The Kintsch and Monk (1972) materials were adapted for this investigation. These materials were composed of elements that, if integrated, could enable students to make and recognize valid inferences or conclusions that go beyond the information given (Bruner, 1957). The rationale was that the Kintsch and Monk (1972) materials could provide evidence of semantic integration if transfer to implications, inferences, and conclusions could be observed. There was also a presumption of cognitive processing at a semantic level. These materials obviate the previously-mentioned criticisms of the 'Bransford Paradigm.'

Rationale for determining time of inference production

Three studies using inferential material and investigating conditions under which inferences were generated (Johnson, Bransford, & Solomon, 1973; Kintsch & Monk, 1972; & Thorndyke, 1976) acknowledged that it could not be determined from their data whether inferences were made during acquisition and then encoded in memory, or if the semantic base only was encoded during acquisition and then the inference made during the retrieval phase. However, there is consensus, if only implicit, that inferential thinking occurred in acquisition as opposed to retrieval.

Rationale for studying instructional variation

In the studies reviewed under the 'Bransford Paradigm' and every study, except one, which criticized the 'Bransford Paradigm' findings, the students were naive to the recognition task or the requirement to integrate the acquisition materials. In the later studies which do support semantic integration the students expected the criterion task and/or understood that they had to make an inference or reach a conclusion based on the acquisition materials.

Why should naivete versus non-naivete to the recognition task requirements be important to this investigation of semantic integration? The condition wherein a student has an intent to learn, remember, and integrate constitutes a task-specific learning set for the inferential materials that will be used in this study. Past research supports the importance of intention to remember and the prior establishing of a meaningful learning set (Ausubel, 1968; Ausubel, Schpoont, & Cukier, 1957; Harlow, 1949). The establishment of a set to integrate and go beyond the information given can be controlled by instructions to the learner.

Hypotheses

The focus of this proposal is the concept of semantic integration. Semantic integration will be studied by observing the production of inferences. By examining various aspects of inferential thinking it is expected that some properties of semantic integration will yield to scientific inquiry. Inferential thinking will be examined in two major contexts: one in which stimulus materials are experimentally manipulated, and the other in which instructions are experimentally manipulated. Within these two major contexts, developmental effects will

be studied. The developmental effects will be main effects for age, and interactions of age with materials and instructions.

Materials

Stimulus materials will be varied by either explicitly supplying the logical inference during acquisition or not doing so. In the latter case, the inference is available to the learner only by implication and plausibility. Given that inferential thinking takes place during acquisition, it is hypothesized that:

Encoding time during acquisition will be greater when inferences are not explicitly supplied;

Time to produce a correct response, during retrieval, to a query about an inference will not differ across stimulus materials.

If inferential thinking spontaneously takes place, it is hypothesized that:

The number of correct responses to a query about an inference will not differ across stimulus materials.

Age within materials

If reading speed, memory, and comprehension increase with age, it is hypothesized that:

Older children will encode stimulus materials faster than younger children;

The number of correct responses to a query about an inference will increase as age increases;

The time to produce a correct response to a query about an inference will decrease as age increases.

If spontaneous integrative ability increases with age, it is hypothesized that:

More inferential thinking, during acquisition, should be evident for older children than for younger children. Therefore, the difference between stimulus materials on encoding time should be greater for older children.

The explicit availability of an inference during acquisition should differentially benefit younger children in terms of the number of correct responses to a query about an inference.

Instructions

There will be two instruction sets: instructions to process materials in a rote-verbatim fashion, and instructions to integrate stimulus materials so as to produce inferences or conclusions. If the variation in

instructions influence semantic integration as observed by inference production, it is hypothesized that:

Encoding time during acquisition will be greater for the Integrative Instruction group;

Time to produce a correct response, during retrieval, to a query about an inference will be less for the Integrative Instruction group;

The number of correct responses to a query about an inference should be greater for the Integrative Instructional group.

<u>Age within instructions</u>

Older children will encode stimulus materials faster than younger children;

The number of correct responses to a query about an inference will increase as age increases;

The time to produce a correct response to a query about an inference will decrease as age increases;

Integrative Instructions should differentially benefit younger children as manifested by the number of correct responses to a query about an inference.

CHAPTER III

METHODOLOGY

Experiment I

Purpose

The first experiment was designed to investigate the relationships of Materials and Age to inferential thinking. Making an inference from stimulus materials was presumed to reflect semantic integration. Thus, Experiment I investigated the effects of Materials and Age on semantic integration.

Materials

Three types of paragraphs were used, modeled on the Kintsch and Monk (1972) study. Each type had four propositions: A, B, C, and D. Type 1 was constructed as follows: A causes B, B and C cause D (or -D); assert A and C (or A and -C); question D. The assertion and response operationally defined the inference. Type 2: A causes B, B causes C, C causes D; assert A (or -A); question D. Type 3: A causes B, C is an instance of A, D is an instance of B; assert C (or -C); question D.

For each problem type, four paragraphs were constructed. Each paragraph was presented in two ways: once without explicitly stating the inference, and once with the inference explicitly stated. A sample of Type 1 without the explicit inference is:

> Every Christmas there is a winter party. The ice-boats have a race on the lake as long as the ice is one foot thick. All winter the ice on the lake was half a foot thick. The winter party is held in the mountains.

A sample of Type 1 with the inference explicitly stated is:

> Every Christmas there is a winter party. The ice-boats have a race on the lake as long as the ice is one foot thick. All winter the ice on the lake was half a foot thick. The iceboat races were not held this year.

The query:

> Were the iceboat races held this year?

Two experimental lists, Set 1 and Set 2 were constructed. Set 1 contained 12 paragraphs, six without the inference explicitly stated and six with the inference stated. Set 2 was the complement of Set 1. Each list began with two warmup paragraphs (fillers), one with and one without the inference. Thereafter,

paragraphs were arranged in four blocks of three. Each block contained one problem of Type 1, 2, and 3.

Whether a paragraph appeared with or without an inference, and whether a YES or NO query was used was determined by a counterbalancing scheme following a latin square. Two different orders were prepared for each set. Each paragraph was represented once in each set, either with or without an inference. In each set the correct answer was YES for half the inference paragraphs, and NO the other half. However, only one response form, YES or NO was used for a given paragraph presented with or without an inference. (The paragraphs used are described in Appendix A.)

The paragraphs were selected from a larger set written by a paid, experienced, fifth-grade teacher and modified by this writer. The materials were at the third grade reading level, as determined by the Lorge Readability Index (Lorge, 1959). A panel of 41 adults examined the paragraphs to determine if the conclusions to be drawn from the materials could be inferred; and that the inferences should only be attainable if all the pre-inferential propositions were available to the reader. Thus, paragraphs were eliminated which did not lead

to an appropriate inference, or where the correct answer was obvious merely by asking the question.

Procedure

Index cards (4" X 6") were made for all paragraphs and queries. The students were exposed to one paragraph at a time and a stop watch (accurate to 1/5 of a second) was started. The experimenter stopped the watch when reading was completed. A color naming task was interpolated while the experimenter recorded the encoding time. Another 4" X 6" card was shown to the student as the experimenter started the timer a second time. This card contained the question on the paragraph inference. The experimenter stopped the timer as the student gave an answer. After the response and time were recorded, the experimenter gave the student a "ready" signal and proceeded to the next paragraph-color naming-YES/NO item.

The instructions were read to the students as follows:

> This is a reading game. You and I will play together. You will read 12 little stories, one at a time. Read them carefully. There is no time limit, but I will time how long you take. Do not waste too much time on any one story, but you can take all the time you need.

After each story you will be asked to name some colors. This is a rest period.

After the rest period I will show you a question about what you have just read. Read the question and answer "Yes" or "No". There is no time limit, I will time how long you take to answer.

Now let's do two practice stories before we play the game.

1. When I turn the card, read the story. When you have finished, look up at me and say "Stop". Ready, begin.

...."Stop".

2. Look at these colors. Tell me the names of these colors from left to right.

....

3. When I turn the card, read the question. When you know the answer, say "Yes", or "No".

....

Now let's do one more practice story.

1. Read the story. When you finish, look at me and say "Stop". Ready, begin.

...."Stop".

2. Name these colors.

....

3. Answer the question "Yes" or "No".

....

Very good. Now we are ready to begin. Are there any questions?

1. Read the story. When you finish, look at me and say "Stop". Begin.

...."Stop".

2. Name these colors.

....

3. Answer "Yes" or "No".

....

[Repeat numbers 1, 2, and 3.]

A flow diagram of the experimental trials and data recording is presented in Figure 1.

Subjects

Fifty-eight (58) students from five grades were enlisted: grades four, five, six, seven, and eight. They were randomly assigned to two groups, 29 in each. The school was a coeducational learning center affiliated with a university campus, but whose population was drawn from the community-at-large. As such, a broad range of backgrounds was represented. The students were accustomed to frequent testing, both group and individual, and had cooperated with a number of graduate student research projects over the years.

Design

The first experiment had three dependent variables: time to encode a stimulus passage, the number of correct inferences, and the time to produce a response to a query about an inference from a passage.

START TIMER	ACQUISITION ENCODING
	TIMER STOPPED
	COLOR NAMING (E RECORDS TIME & RESETS TIMER)
START TIMER	QUERY INFERENCE
	YES/NO RESPONSE STOPS TIMER
	E RECORDS TIME & RESPONSE & RESETS TIMER

Figure 1. Experimental trials and data recording procedure for Experiment I.

There were two independent variables: Age (110 - 180 months), and Stimulus Materials -- the logical inference was supplied (IS) or was nor supplied (NIS).

Hypotheses to be tested

The hypotheses stated at the end of Chapter II can now be restated. (A more technical statement of the hypotheses is displayed in Appendix B.)

H_1: If inferential thinking spontaneously takes place, the mean number of correct responses to a query about an inference should not vary across stimulus materials.

H_2: The number of correct responses to a query about an inference will increase as age increases. There is a positive correlation between age and correct responses.

H_3: If younger children possess less spontaneous (natural) semantic integrative ability than older children, then an inference supplied during acquisition should differentially benefit younger children as manifested by the number of correct responses to a query about an inference. That is to say, there is an interaction between age and materials on the

number of correct responses to an inference query.

H_4: If inferential thinking takes place during acquisition, then mean encoding time will be greater when inferences are not explicitly supplied.

H_5: Older children will encode the stimulus materials faster than younger children. There is a negative correlation between age and encoding time.

H_6: If older children possess greater semantic integrative ability, they would spend more time during acquisition on inferential thinking than younger children, in the absence of any supplied inference. As a result, the difference between older and younger children on encoding time should be greater for NIS than IS materials. That is to say, there would be an interaction between age and materials on encoding time.

H_7: If the encoding of inferences, when they are not explicitly supplied, takes place during acquisition rather than retrieval, the mean time to produce a correct response to a query

about an inference will not vary across stimulus materials.

H_8: The time to produce a correct response to a query about an inference will decrease as age increases. There is a negative correlation between age and response time.

Experiment II

Purpose

The second experiment was designed to investigate the relationships of Instructions and Age to inferential thinking. Making an inference from stimulus materials was presumed to reflect semantic integration. Thus, Experiment II investigated the effects of Instructions and Age on semantic integration.

Materials

Three types of paragraphs were used, modeled on the Kintsch and Monk (1972) study. These were the same as those described in the preceding section on Experiment I materials. For each type of problem, four paragraphs were constructed. The inferences were not explicitly stated. A sample of Type 1 paragraph without explicit inference was shown previously on page 98. A single

list of 12 paragrpahs was constructed. (The paragraphs are described in Appendix C.)

For each paragraph there were two questions. One query related to a specific sentence that was actually presented. The other was a query about an inference from the story. The order of the queries and whether they require a "Yes" or "No" response was determined by a counterbalancing scheme following a latin square.

A set of random orders for the 12 stories was established prior to the experiment. Each student received a different random order. However, the same random order sets were used across instruction conditions.

Procedure

As in Experiment I, 4" X 6" index cards were made for all paragraphs and queries. The students were exposed to one paragraph at a time with unlimited reading time. Each student received only one set of instructions: Verbatim or Integrative. The choice of instruction set was random, chosen by the experimenter prior to the experiment and read to the student by the experimenter. The Verbatim Instructions were as follows:

> This is a reading game. You and I will play together. You will read 12 little stories, one

at a time. Read them carefully. There is no time limit, but I will time how long you take. Do not waste too much time on any one story, but you can take all the time you need.

After each story you will be asked to name some colors. This is a rest period.

After the rest period I will show you two questions about what you have just read. You must answer "Yes" or "No". There is no time limit, but I will time how long you take to answer.

One question is important and the other is not. One question is about a sentence you actually read in the story. This is the important question. So pay close attention when you read the stories. Remember the words you actually see.

Another question makes you think about the story. This question is not important, but answer it any- way. Just remember only the words you actually see.

Now let's do two practice stories before we play the game.

1. When I turn the card read the story. When you have finished, look up at me and say "Stop". Ready, begin.

...."Stop".

2. Look at these colors. Tell me the names of these colors from left to right.

....

3. When I turn the card read the question. When you know the answer, say "Yes" or "No".

....

4. When I turn the card read the question. When you know the answer, say "Yes" or "No".

....

Now let's do one more practice story.

1. Read the story. When you finish, look at me and say "Stop". Begin.

...."Stop".

2. Name these colors.

....

3. Answer "Yes" or "No".

....

4. Answer "Yes" or "No".

....

Very good. Now we are ready to begin. Are there any questions?

[Repeat the last set of instructions 1, 2, 3, and 4.]

The Integrative Instructions varied in the fourth and fifth paragraphs as follows:

One question is important and the other is not. One question is about a conclusion from the story. This is the important question. The conclusion is not really a sentence in the story, but you could figure it out from the rest of the story. When you read the story try to figure things out. So pay close attention to what the story means.

The unimportant question is about the words you actually read in the story. Only remember the meaning, not the words.

A flow diagram of the experimental trials and data recording is presented in Figure 2.

INSTRUCTIONS
VERBATIM OR
INTEGRATIVE

START TIMER — ACQUISITION ENCODING

TIMER STOPPED

COLOR NAMING
(E RECORDS TIME
& RESETS TIMER)

START TIMER — QUERY 1

YES/NO RESPONSE
STOPS TIMER

E RECORDS TIME
& RESPONSE
& RESETS TIMER

START TIMER — QUERY 2

YES/NO RESPONSE
STOPS TIMER

E RECORDS TIME
& RESPONSE
& RESETS TIMER

Figure 2. Experimental trials and data recording procedures for Experiment II.

Subjects

Fifty-nine (59) students from five grades were selected: grades four through eight. Twenty-nine (29) were randomly assigned to the Verbatim group and 30 to the Integrative group. They were drawn from the same school and population described in Experiment I; however, the students across the two studies were different.

Design

The second experiment had three dependent variables: time to encode a stimulus passage, number of correct responses to a query about an inference, and time to produce a correct response to a query about an inference from a passage.

There were two independent variables: Age (110 - 180 months), and Instructions (Verbatim and Integrative).

Hypotheses to be tested

The hypotheses stated at the end of Chapter II can now be restated. (A more technical presentation of the hypotheses is displayed in Appendix F.)

H_9: Inferential thinking will take place during acquisition for the Integrative group and during retrieval for the Verbatim group. However, there should be no differences in

the attainment of the correct inference between the instruction groups.

H_{10}: The number of correct responses to a query about an inference will increase as age increases. There will be a positive correlation between age and correct response.

H_{11}: If younger children possess lesser spontaneous (natural) semantic integrative ability than older children, then an integrative instruction set should differentially benefit younger children, as manifested by the number of correct responses to a query about an inference. That is to say, there will be an interaction between age and instructions on the number of correct responses to an inference query.

H_{12}: The integrative instructions will elicit more inferential thinking during acquisition than verbatim instructions. Thus, the mean encoding time for the Integrative group will be greater than the Verbatim group.

H_{13}: Older children will encode the stimulus materials faster than younger children. There will be a negative correlation between age and encoding time.

H_{14}: The integrative instructions will compel more encoding of inferences during acquisition along with the actual sentences, while the verbatim instructions will foster the encoding of only the sentences. Thus, inferences for the Verbatim group must be processed from the sentences after they are recalled, while the Integrative group merely will recall the inference. Thus, response time will take longer for the Verbatim group than the Integrative group.

H_{15}: The time to produce a correct response to a query about an inference will decrease as age increases. There will be a negative correlation between age and response time.

CHAPTER IV

RESULTS

Experiment I

The data

The summary descriptive statistics are displayed in Table 1. The raw data are presented in Appendix E. Paired means were compared by the use of *t* tests. Zero order correlations were tested in a linear regression format, because the statement of R^2 and the sign of the regression weight convey information about variance explained and direction of a relationship. Comparisons of independent means were done with dummy coding in a regression model.

The previous chapter described three types of stimulus materials used in this study: three distinct logic types. The main analyses collapsed the data across all three logic types. However, post hoc analyses were computed separately for each logic type. The purpose of these additional analyses was to determine whether the results were unique to a particular logical construction in the stimulus materials or generalizable across a variety of logical constructions. The raw data for each logic type are presented in Appendix F.

Hypothesis 1

The first hypothesis stated that the number of correct responses to a query about an inference would not differ across stimulus materials. This assumed that inferential thinking was a natural, spontaneous operation in acquisition.

The hypothesis was not supported. The mean correct responses for No Inference Supplied (NIS) and Inference Supplied (IS) materials from Table 1 were, respectively, 3.29 and 5.10. The difference was significant, t (57) = 11.16, $p < .001$.

A post hoc analysis revealed that each logic type displayed the same significant difference. As seen in the summary descriptive statistics in Table 2, the smallest difference between means was for logic type 2. The mean correct responses for NIS and IS materials are, respectively, 1.34 and 1.72, t (57) = 3.16, $p < .005$.

Hypothesis 2

The second hypothesis stated that older children would give more correct responses to a query about an inference than younger children. It was reasonable to assume that reading comprehension and integrative ability would increase with age.

Table 1

Mean Scores and Standard Deviations for Three Dependent Variables for Inference Supplied (IS) and No Inference Supplied (NIS) Materials

Variable	M	SD	N
Correct Responses			
IS	5.10	0.98	58
NIS	3.29	1.21	58
Encoding Time[a]			
IS	15.53	4.74	58
NIS	16.03	5.16	58
Response Time[a]			
IS	2.33	0.84	56
NIS	2.75	1.02	56

[a]Measured in seconds

Table 2

Mean Scores and Standard Deviations for Three Dependent Variables for IS and NIS Materials and Logic Type

	Logic Type 1			Logic Type 2			Logic Type 3		
Variable	M	SD	N	M	SD	N	M	SD	N
Correct Response									
IS	1.81	0.40	58	1.72	0.59	58	1.56	0.53	58
NIS	1.01	0.71	58	1.34	0.69	58	0.93	0.49	58
Encoding Time									
IS	14.97	5.24	58	15.16	4.61	58	16.48	5.06	58
NIS	16.06	5.23	58	15.09	5.17	58	16.93	5.79	58
Response Time									
IS	2.32	1.11	58	2.22	0.83	58	2.37	1.12	58
NIS	2.94	1.38	58	2.95	1.24	58	2.43	1.15	58

The hypothesis was not confirmed. However, while the value of R^2 was .098 and significant, F (1, 57) = 6.20, $p < .025$, the sign of b_1 was negative (the reverse of the hypothesis). This unexpected finding indicated that older children made more errors. The regression line is plotted in Figure 3.

A post hoc analysis showed that each logic type displayed the same relationship between age and the number of correct responses. There was no significant interaction of age with logic type in the relationship of age to correct responses, F (1, 53) = 1.00, $p > .05$.

Hypothesis 3

The third hypothesis stated that younger children would be differentially benefitted by the IS materials in terms of the number of correct responses to an inference query. Older students, it was reasoned, would not need the benefit of an inference explicitly available as much as younger students.

The hypothesis was not confirmed. Younger children were not differntially benefitted by the IS materials, F (1, 57) = 2.48, $p > .05$. This non-interaction effect is plotted in Figure 3 along with the results from hypothesis 2.

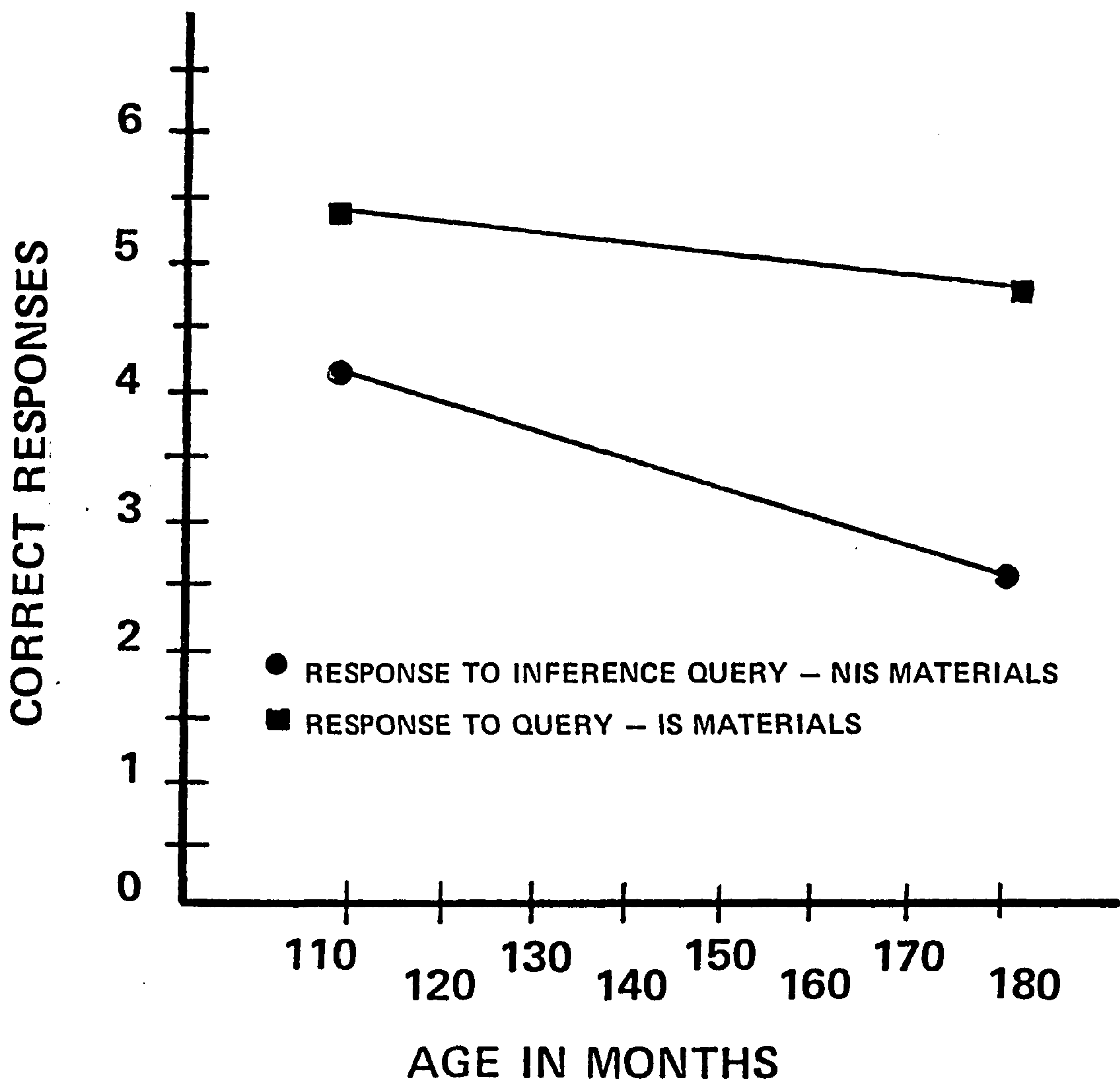

Figure 3. Mean correct responses for IS and NIS materials as a function of age in Experiment I.

A post hoc analysis failed to find any significant interaction effect for any of the three logic types. The increment in variance explained by adding an interaction term -- age with stimulus materials -- was highest for logic type 3, .0071, and clearly not significant, F (1, 54) = 1.00, $p > .05$.

Hypothesis 4

The fourth hypothesis stated that encoding time during acquisition would be greater when inferences were not explicitly supplied. This assumed that inferential thinking took place during acquisition. Encoding time for each student was computed by averaging the time to acquisition of six IS paragraphs and six NIS paragraphs. Since each student in Experiment I was exposed to both types of materials (IS and NIS) a t test comparing means for paired data was executed.

The hypothesis was supported. Encoding time for NIS and IS materials from Table 1 were, respectively, 16.03 seconds and 15.53 seconds. The difference was significant, t (57) = 2.44, $p < .05$.

However, a post hoc analysis showed this significant difference between IS and NIS materials on encoding time to be true for only logic type 1, t (57) = 3.11, $p < .005$. As seen in Table 2, the mean encoding time for IS and NIS

materials are, respectively, 14.97 and 16.06 seconds. Neither logic type 2 or 3 approached significance, with the larger *t* value as follows: $t\ (57) = 0.93$, $p > .05$. This interaction of stimulus materials with logic type will be discussed at length in the next chapter.

Hypothesis 5

The fifth hypothesis stated that older children would encode stimulus materials faster than younger children. It was assumed that reading speed would increase with age.

The hypothesis was not confirmed. Encoding time was not related to age, $F\ (1, 57) = 1.38$, $p > .05$. The regression line is plotted in Figure 4.

A post hoc analysis showed no significant relationship of age to encoding time for any logic type, with the largest non-significant value of *F* for logic type 1: $F\ (1, 57) = 2.82$, $p > .05$.

Hypothesis 6

The sixth hypothesis stated that the difference between stimulus materials on encoding time should be greater for older children. The reasoning was that more inferential thinking, during acquisition, should be evident for older than younger children.

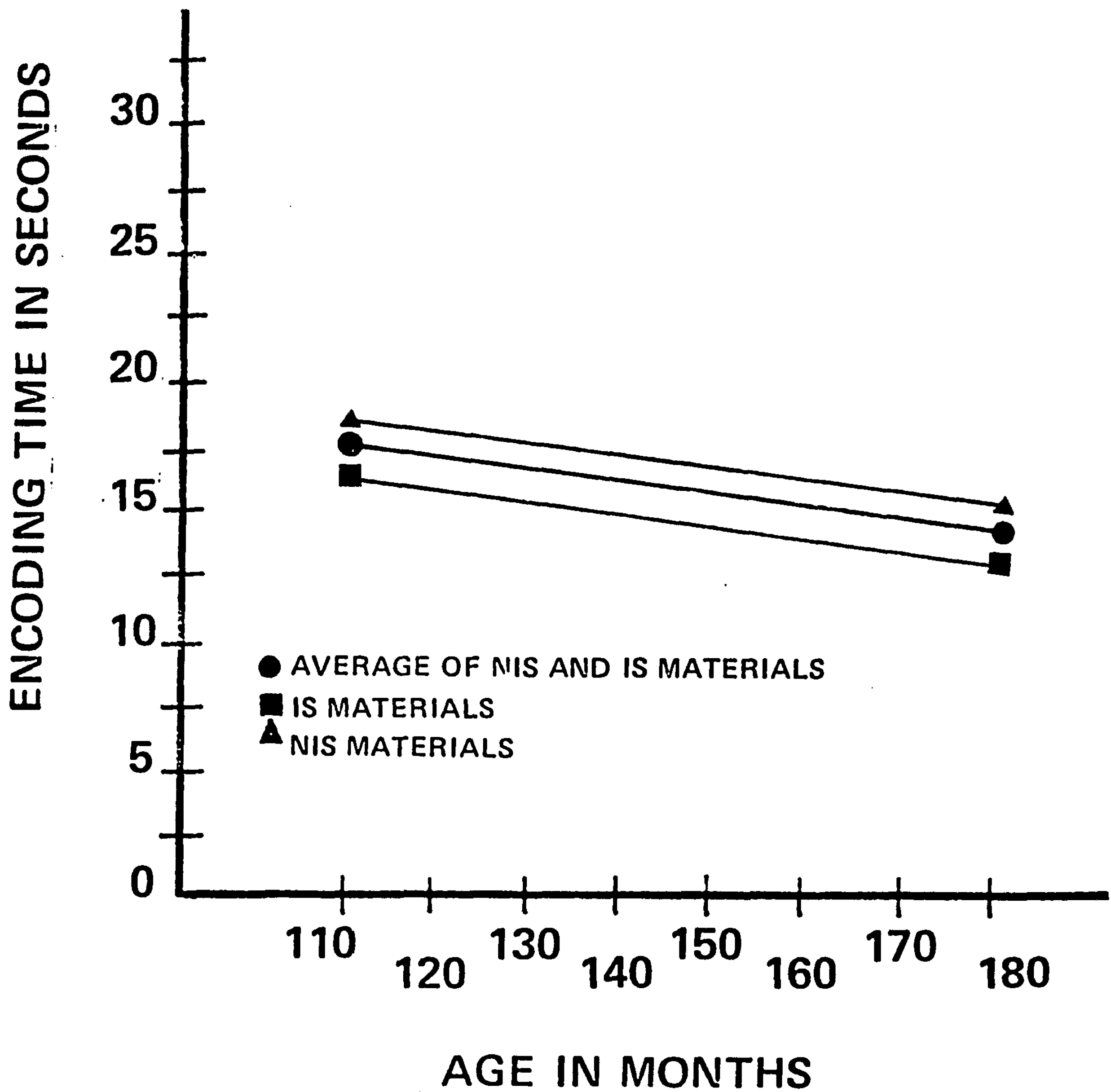

Figure 4. Mean encoding time for IS, NIS and the average of NIS and IS materials as a function of age in Experiment I.

The hypothesis was not confirmed. Any difference between stimulus materials on encoding time, was constant across all ages, F (1, 57) = 1.00, $p > .05$. These means are plotted in Figure 4 along with the means for hypothesis 5.

A post hoc analysis dod not show any significant interaction effect of age with stimulus materials on encoding time for any logic type. The largest F observed was for the interaction term for logic type 1, F (1, 54) < 1.00, $p > .05$.

Hypothesis 7

The seventh hypothesis stated that the time to produce a correct response, during retrieval, to a query about an inference would not differ across stimulus materials. This presumed that inferential thinking took place during acquisition. Response time for each student was computed by averaging the response times of each correct response, and doing so separately for the two types of paragraphs. Two students did not respond correctly to any of the queries for the NIS materials. As a result, they were dropped from the analysis for this hypothesis.

The hypothesis was not supported. Mean response times for NIS and IS materials from Table 1 are, respectively, 2.75 seconds and 2.33 seconds. The difference was significant, t (55) = 3.13, $p < .01$, but contrary to the predicted direction.

A post hoc analysis indicated a significant difference in response times between stimulus materials for logic type 1, t (57) = 3.17, $p < .005$, and logic type 2, t (57) = 3.91, $p < .001$. As seen in Table 2, the mean response times for IS and NIS materials for logic type 1 are, respectively, 2.32 and 2.94 seconds; and for logic type 2 they are 2.22 and 2.95 seconds. While the difference for logic type 3 was not statistically significant, t (57) = 0.39, $p > .05$, the observed difference was in the same direction as the other two logic types (IS = 2.37 seconds and NIS = 2.43 seconds).

Hypothesis 8

The eighth hypothesis stated that the time to produce a correct response to a query about an inference would decrease as age increases. It was assumed that processing time in memory would be faster for older children.

The hypothesis was not confirmed. Response time was not related to age, F (1, 55) = 2.27, $p > .05$.

A post hoc analysis did not reveal any significant relationship of age to response time for any logic type. The largest, but not significant, relationship was observed for logic type 1, F (1, 57) = 2.58, $p > .05$.

EXPERIMENT II

The data

The summary descriptive statistics are displayed in Table 3. The raw data are presented in Appendix G. As in Experiment I, post hoc analyses were computed separately for each logic type. The raw data for each logic type are presented in Appendix H.

Hypothesis 9

The ninth hypothesis stated that the number of correct responses to a query about an inference would be greater for the Integrative instruction group. It was reasoned such instructions should produce a higher frequency of correct inferences.

The hypothesis was not supported. Correct responses did not differ between instructional groups, F (1, 57) = 1.00, $p > .05$.

A post hoc analysis did not find any significant differences in correct responses between instructional groups for any of the three logic types. The largest

Table 3

Mean Scores and Standard Deviations for Three Dependent Variables for Integrative and Verbatim Instructions

Variable	M	SD	N
Correct Responses			
Integrative	6.46	1.96	30
Verbatim	6.65	2.37	29
Encoding Time[a]			
Integrative	17.05	6.51	30
Verbatim	17.35	5.61	29
Response Time[a]			
Integrative	2.91	1.14	30
Verbatim	2.93	1.01	29

[a]Measured in seconds

observed difference was for logic type 3 with an F (1, 57) = 2.33, $p > .05$. As seen in Table 4, the mean correct responses for Verbatim and Integrative groups for logic type 3 are, respectively, 2.27 and 1.87.

Hypothesis 10

The tenth hypothesis stated that the number of correct responses to a query about an inference would increase as age increases. It was assumed that reading comprehension and inferential ability would increase with age.

The hypothesis was not confirmed. Correct responses did not increase with age, F (1, 57) = 1.00, $p > .05$. The regression line is plotted in Figure 5.

A post hoc analysis did not find any relationship between age and correct responses for the three logic types. The onserved F for the largest, but not significant, relationship was for logic type 3, F (1, 57) = 1.00, $p > .05$.

Hypothesis 11

The eleventh hypothesis stated that younger children would differentially benefit from the integrative instructions in giving a correct response to a query about an inference. It was reasoned that older children would

Table 4

Mean Scores and Standard Deviations for Three Dependent Variables for Integrative and Verbatim Instructions and Logic Type

	Logic Type 1			Logic Type 2			Logic Type 3		
Variable	M	SD	N	M	SD	N	M	SD	N
Correct Response									
Verbatim	2.03	1.11	29	2.31	1.34	29	2.27	0.95	29
Integ've	2.17	0.94	30	2.39	1.03	30	1.87	1.04	30
Encoding Time									
Verbatim	17.44	5.69	29	16.36	5.60	29	18.34	5.78	29
Integ've	17.42	6.79	30	15.96	5.81	30	17.84	7.03	30
Response Time									
Verbatim	2.87	1.07	26	3.27	2.28	28	3.05	1.65	28
Integ've	2.81	1.35	30	2.93	1.31	30	2.57	1.22	29

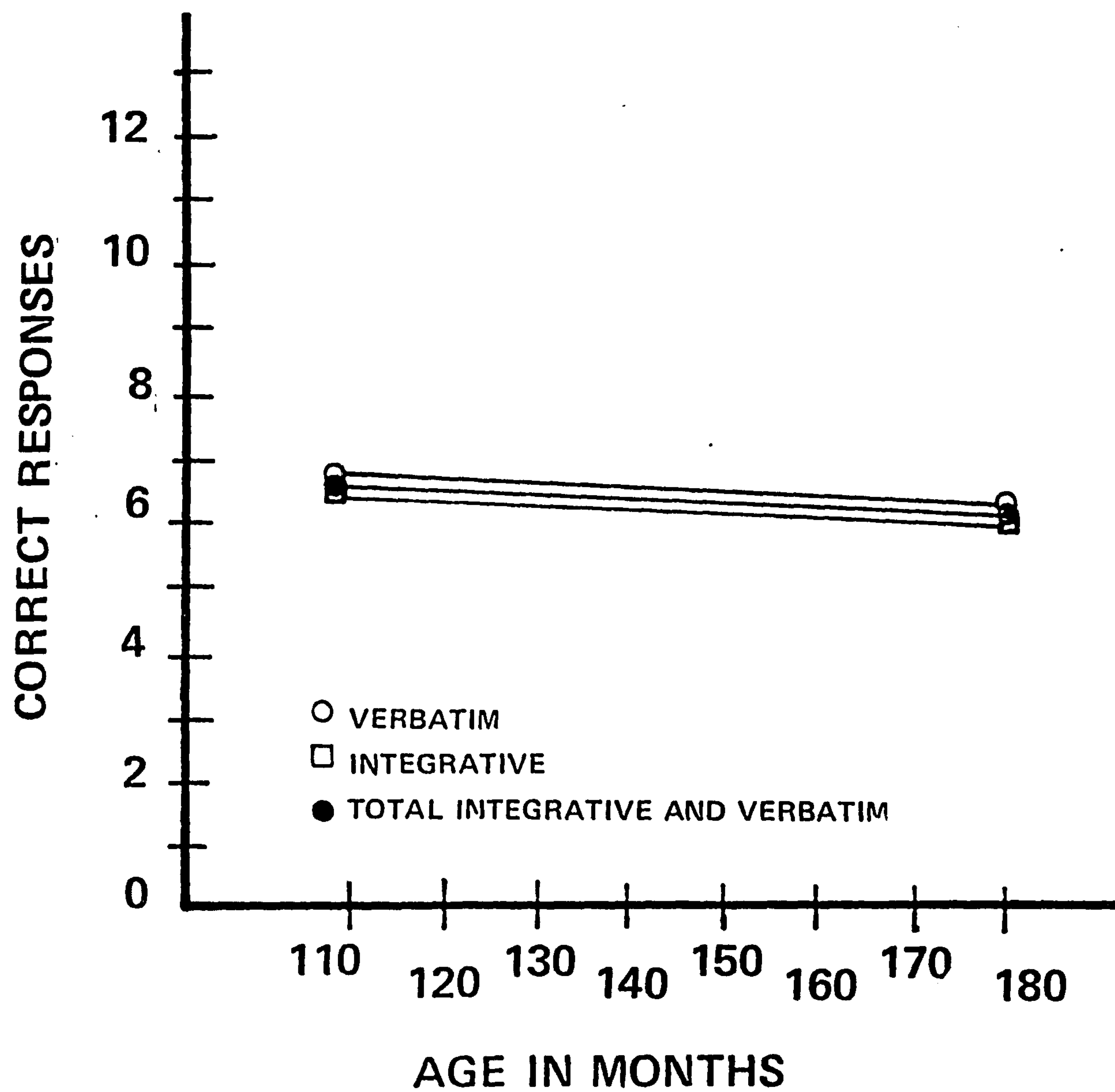

Figure 5. Mean correct responses to inference queries for integrative, verbatim and total groups as a function of age in Experiment II.

find the process more natural and spontaneous. The younger children would benefit more from an artificial support and explicit prompting.

The hypothesis was not confirmed. Younger children did not differentially benefit from Integrative instructions, F (1, 57) = 1.00, $p > .05$. This non-interaction is plotted in Figure 5 along with the results for hypothesis 10.

A post hoc analysis did not find any significant interaction effect of age with instructions on correct response for the three logic types. The largest, but not significant, interaction effect was for logic type 3, F (1, 57) = 2.79, $p > .05$.

Hypothesis 12

The twelfth hypothesis stated that encoding time during acquisition would be greater for students receiving integrative instructions than those receiving verbatim instructions. It was reasoned that instructions to produce an inference would instigate more cognitive processing, that required more time, than instructions to simply remember the overt content of a passage.

The hypothesis was not supported. Encoding time was the same across instructions, F (1, 57) = 1.00, $p > .05$.

A post hoc analysis indicated there was no significant difference in encoding time between instruction groups for any of the logic types. As seen in Table 4, the largest difference was for logic type 3 (Verbatim = 18.34 seconds; Integrative = 17.84 seconds) but not significant, F (1, 57) = 1.00, $p > .05$.

Hypothesis 13

The thirteenth hypothesis stated that older children would encode stimulus materials faster than younger children. It was assumed that reading speed would increase with age.

The hypothesis was confirmed. There was a significant relationship between encoding time and age, F (1, 57) = 5.83, $p < .05$. The regression line is plotted in Figure 6.

A post hoc analysis demonstrated a significant relationship between age and encoding time for each logic type. The smallest, but yet significant, relationship was for logic type 1, F (1, 57) = 4.84, $p < .05$.

Hypothesis 14

The fourteenth hypothesis stated that the time to produce a correct response, during retrieval, to a query about an inference would be less for the group receiving integrative instructions. It was reasoned that for the

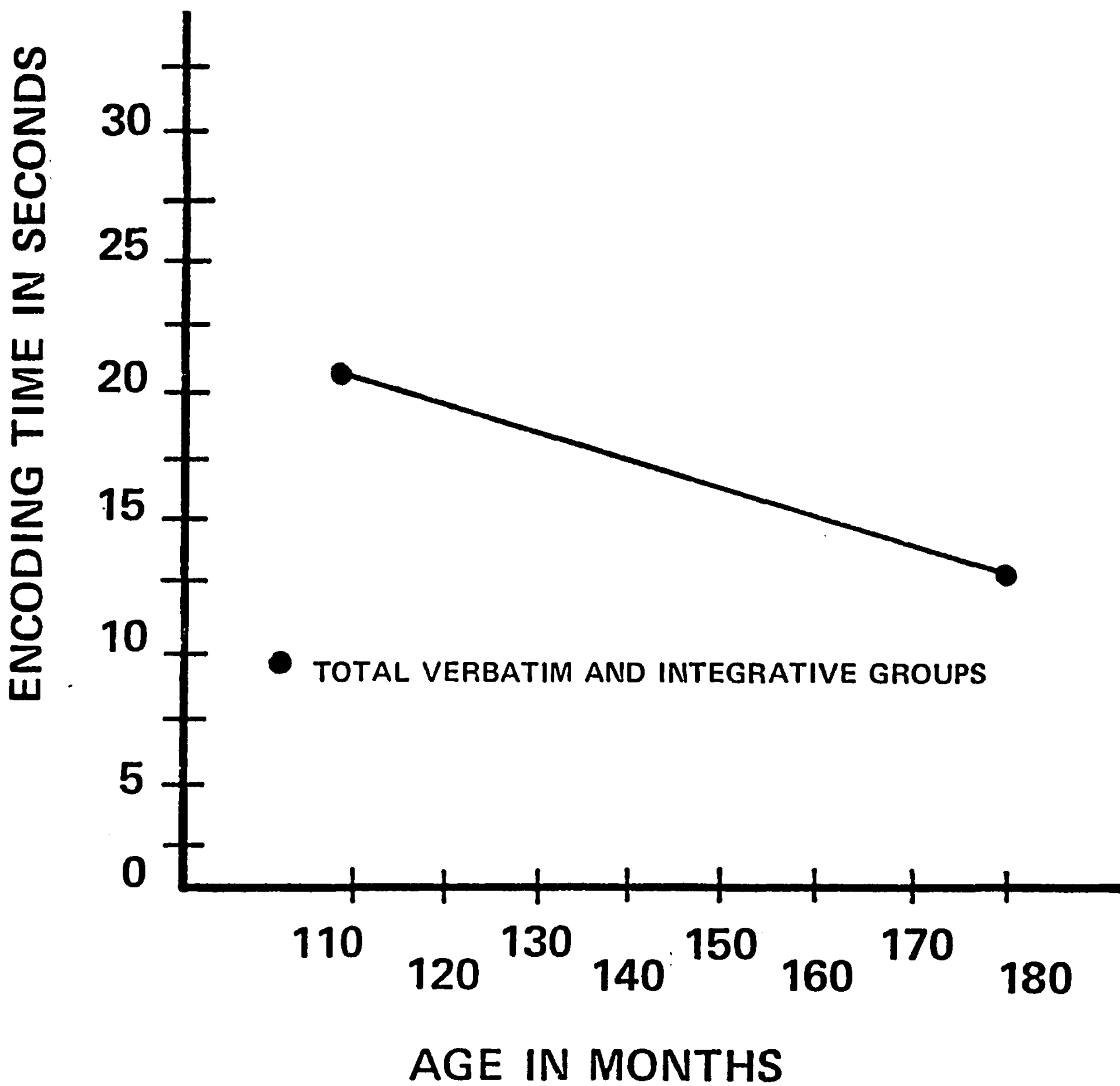

Figure 6. Mean encoding time as a function of age in Experiment II.

Integrative group, more inferential thinking would take place during acquisition and save time during the retrieval phase to produce a response.

The hypothesis was not supported. Response time was the same for both instructional groups, F (1, 57) = 1.00, $p > .05$.

A post hoc analysis found no significant differences between instruction groups on response time for any of the three logic types. As seen in Table 4 the largest difference between instruction groups was for logic type 3 (Verbatim = 3.05 seconds; Integrative = 2.57 seconds), but it was not significant, F (1, 57) = 1.59, $p > .05$.

Hypothesis 15

The fifteenth hypothesis stated that the time to produce a correct response to a query about an inference would decrease as age increases. It was assumed that cognitive processing would be more efficient for older students.

The hypothesis was confirmed, F (1, 57) = 8.19, $p < .01$. The regression line is plotted in Figure 7.

A post hoc analysis indicated a significant relationship between age and response time for logic type 1, F (1, 55) = 8.19, $p < .01$, and logic type 2, F (1, 55) = 3.67, $p < .05$. The realtionship for logic type 3 was

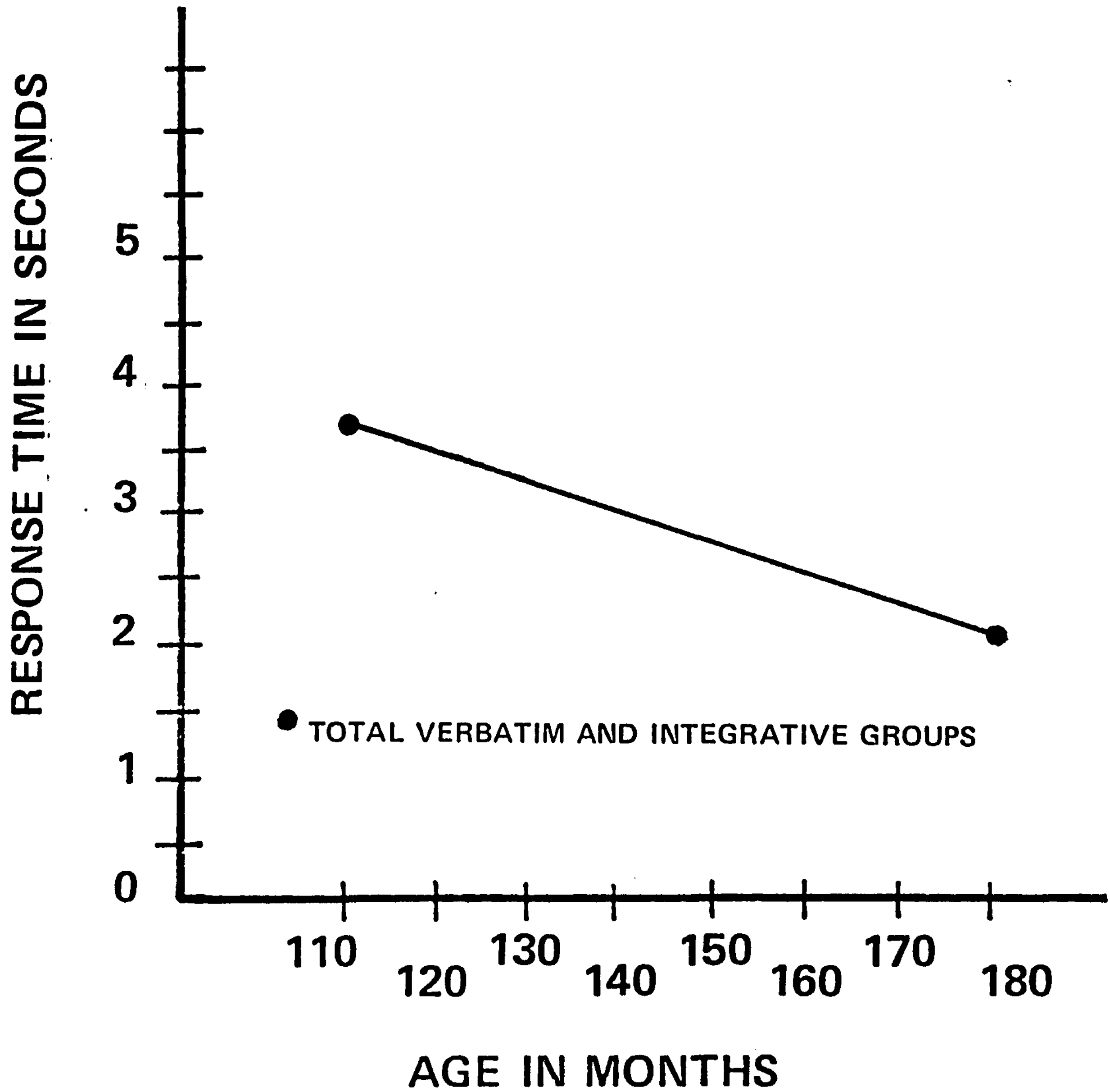

Figure 7. Mean response time to an inference query as a function of age in Experiment II.

not statistically significant. However, the observed direction of the relationship was the same as logic types 1 and 2.

CHAPTER V

DISCUSSION

Time of Inference Production

The issue to be resolved was whether inferences were generated during acquisition and encoded along with primary written material or generated during retrieval after first recalling the primary input.

The results of the first hypothesis indicated that inference production took place in retrieval rather than acquisition. This hypothesis stated that the number of correct responses to a query about an inference would not differ across NIS and IS passages. It was assumed inferential thinking was a natural, spontaneous operation which took place in acquisition.

The results actually showed that for NIS passages the mean correct reponses (3.29 was significantly lower than for the IS materials 5.10). It was obvious that the inferences were not made during acquisition. The inferences were generated during retrieval. If inferences were generated during acquisition, they should be as readily available as other information that was actually given during acquisition. This was not the case. Since the inferences were dependent on the availability

of the primary input during recall, any loss due to forgetting of the generative material would, of necessity, reduce the likelihood of being able to produce a correct response.

The results from the fourth hypothesis showed that encoding time was greater for the NIS materials. The difference in encoding time between the two groups was small, 0.5 seconds, but it was significant and does suggest some differential mental reaction to an incomplete proposition. However, this additional encoding time did not lead to any significant increase in inferential thinking, as noted above from the results on correct inference responses.

The results of the seventh hypothesis support the results of the first. The mean response time to the NIS materials (2.75 seconds) was significantly greater than that for the IS materials (2.33 seconds). The extra time (0.42 seconds) needed for a response indicated additional processing during retrieval for NIS queries. This would suggest that inferences were produced in retrieval after the recall of the primary generative material.

The generalizability of these results based on collapsing data across logic types is mitigated somewhat by the finding that the significant difference for

encoding time was found only for logic type 1. There were no significant differences in encoding time observed for the other two logical constructions. A possible explanation might be based on the relative difficulties of these three types of logical constructions. An examination of the number of correct inferences for each type of logic (Table 2) indicates that logic type 1 was intermediate in difficult between the easier logic type 2 and more difficult logic type 3.

An inspection of the logical models shown in Chapter III suggests the same relative ordering of difficulty of types of logic (transformations, formal analysis, indirect proposition, etc.). When the inferences are supplied student's encoding time on the easiest and hardest were not influenced as much. Since logic type 2 was easier, perhaps supplying an inference was not needed to reduce encoding time. However, at the very least, these differential results are not contradictory since the data do not go in the opposite direction.

It should be noted that no interactions were observed for logic type in hypotheses one and seven. Thus, the results are not unique to a particular logical construction of the stimulus materials.

While Thorndyke (1976) inferred from his results (see pp. 85-88) that inferences were generated and encoded during acquisition, this conclusion must now be reexamined. Inferences appeared to be made during acquisition in Thorndyke's study only because the experimenter required the students to do so after reading a few early passages.

The present data also answer the question raised by the Johnson, Bransford and Solomon (1973), and the Kintsch and Monk (1972) studies as to the time of inference production. It appears to occur during retrieval.

Developmental Effects

Correct responses to inference queries

The results of hypothesis 2 in Experiment I were a reversal of what was expected. The data showed that older children displayed more errors than younger children (see Figure 3).

This may be one of the more interesting findings of the study. Two possible, although not mutually exclusive, explanations are: (a) older children, by virtue of possessing more information and knowledge about the world in general, were able to consider more possibilities and alternatives to the correct answer

established in the pilot testing; and (b) older children were more reactive to the questions than either the younger children in the same experiment or the adults studied during pilot testing.

For example, one student who gave the wrong answer to the inferential query, "Did the fish live with the turtle?" (correct answer was "No"), volunteered the following comment: "Yes, the fish lived with the turtle until the time that the turtle got around to eating the fish." It was possible the older children suspected they were being asked a "trick" question. In their effort to be scrupulously correct they chose the wrong dichotomous option. Younger children, less sophisticated in judging the purpose of the testing, might not have been overly reactive.

This reactivity on the part of older students to the NIS materials may present a problem to the interpretation of the data and call into question whether or not the NIS materials were a valid measure of inferential thinking. The key to resolving this problem was to determine if the reactivity of the older students was unique to the NIS materials or generalizable to all materials. From an inspection of Figure 3 it is apparent that older children made more errors even with the IS materials.

Since the slopes of the regression lines were not significantly different it must be concluded that the reactivity of the older students was constant across all materials.

The observed reactivity (greater error rate) of older students does not alter the interpretation of results for other hypotheses. For example, mean differences in correct responses between IS and NIS materials can be interpreted unambiguously whether the slopes are positive, negative, or zero (as long as the slopes for the IS and NIS materials are not significantly different). The time to produce a correct response to a query about an inference can also be interpreted unambiguously. By definition, the response time was computed only for correct responses. Response times were computed as the materials were successfully and appropriately promoting the production of an inference. Thus, there was no compelling reason to question the appropriateness of the materials as they reflect on the validity of the results as a whole.

Comparing experiments, it is noteworthy that correct responses were correlated with age in Experiment I, but not in Experiment II. In both experiments, the results observed were the same across all logic types. In

neither experiment was there an interaction with logical construction.

Encoding Time

The results of hypothesis 5 from Experiment I indicated that encoding time did not increase with age, as expected. Perhaps this was due to a 'floor effect' since the reading difficulty was held constant at or below the reading level of the youngest student. Reading level was controlled, since this was not an issue of interest.

However, these findings contrast with the results of hypothesis 13 from Experiment II where age was related to encoding time. The only thing that could be inferred from these data was that the two experimental frameworks, themselves, set the conditions for the age-encoding time effect to manifest itself. Future research may clarify the issue. It should be noted though, that post hoc analyses did not demonstrate any interactions with logic type for either hypotheses 5 or 13.

Response Time

The results of hypothesis 8 from Experiment I showed that older children did not have faster response times to an inference query. Given that reading time and reading

difficulty were the same, these data indicated no quantitative time difference in information processing within the age range studied. However, there was a possible alternative explanation. It may be that while older children processed information faster, the observed results were confounded by the fact that they were processing more information. After all, older children possess a greater storehouse of ready-to-use information. Further research is needed to confirm and tease out these separate effects.

However, the results of Experiment II, hypothesis 15, do not replicate the findings of Experiment I. In Experiment II there was a relationship of response time to age. Again, this coupled with the non-replication cited above on the age-encoding time relationship, and the age-correct response relationship indicated that the two experimental procedures produced different results in and of themselves. However, it was noteworthy that mean correct inference responses, mean encoding time, and mean response time were not appreciably different (by inspection) across the two experiments. The results of hypotheses 8 and 15 are not unique to the logical construction of the stimulus materials, since no intereactions were observed with logic type.

Developmental differences in inference production

The non-replication of results, cited above, on number of correct responses, encoding time, and response time deals with age differences that are quantitative and linear in nature -- reading speed, memory span, size of mental data base, and processing speed. These non-replications across Experiments I and II do not jeopardize the answer to the more crucial question: Is there a qualitative difference in the production of inferences as age increases?

The results of hypothesis 3 in Experiment I indicated that different age groups were not differentially benefitted by the IS materials in giving correct responses. It was hypothesized that younger children would possess less spontaneous inferential ability and thus benefit more than older students if inferences were provided during acquisition. The benefit of the IS materials on the number of correct responses to an inference query was constant across the entire age range studied. Thus, no developmental difference in inferential ability could be inferred from these data.

The results of hypothesis 11 in Experiment II indicated that differences on mean correct responses to inference queries, between instructional groups,

was the same across the entire age range in this study. It was originally hypothesized that younger children would possess less spontaneous inferential ability and thus benefit more than older students by the artificial support and explicit prompting of the integrative instructions. Different age groups were equally impacted by the instructional variation. Thus, no developmental difference in inferential ability could be inferred from these data.

The results of hypothesis 6 in Experiment I indicated that the difference in encoding time between NIS and IS passages did not increase with age. It was originally hypothesized that more inferential thinking in acquisition would take place among older students. This was not the case. The significant difference was constant across the entire age range studied. Thus, no developmental differences in inferential ability could be inferred from these data. Further, the post hoc analyses on hypothese 3, 11 and 6 did not show any interaction with logic type.

These results, cited above, were consistent with findings reviewed earlier which found no developmental trend in the semantic integrative process. Paris and Carter (1973) (see pp. 58-59) found no developmental

effects between the second and fifth grades. Barclay and Reid (1974) found evidence of integration but no developmental trends from kindergarten to the fifth grade. Moeser (1976) (see p. 66) found no developmental trends from grades two through six. The data reported in this study further support Donaldson (1963) who suggested that children will commonly encode the elements of a potentially integrative task as separate units. We did not find what was expected: that older children would naturally change their encoding strategy to an integrative orientation.

Instructional Effects

The results of hypothesis 9 from Experiment II indicated that integrative instructions did not modify the strategy for information processing so as to yield more correct responses. In other words, instructions, per se, did not influence the production of inferences.

Data bearing on hypothesis 12 in Experiment II indicated that encoding time during acquisition was not greater for students who received integrative instructions. It was originally hypothesized that integrative instructions would initiate more cognitive processing which would require more time than the verbatim instructions. Either the instructions had no effect at all in

modifying information processing strategy, or the instructions were efficacious but the different strategies required the same amount of time. In line with the first explanation, perhaps the students in this age range do not have sufficient volitional control over cognitive processing to be modified by an instructional set.

It was obvious from the results of hypothesis 14 that integrative instructions did not decrease response time during retrieval. As such, these data, in conjuction with results from hypotheses 9 and 12, would suggest that integrative and verbatim instructions did not produce different strategies that just happened to take the same period of time.

The evidence indicated that integrative and verbatim instructions simply did not modify whatever strategy the students were using. The generalizability of the results of hypotheses 9, 12 and 14 are not mitigated by logical construction of the stimulus materials since no interaction with logic type was observed.

General Comments on Semantic Integration

Spontaneity

The evidence discussed above seriously challenges the assertion by Bransford and Franks (1971) that, "[i]ndividual sentences lost their unique status in memory in favor of a more wholistic representation of semantic events." Only the specifics were stored. Memory was modified only during, and as a result of the problem-solving task in, recall. These data support a highly specific and detailed model of memory, consistent with the "retain everything position described in Chapter I.

The results suggest that students were not constructing 'wholistic' representation of ideas from partial or fragmentary elements of information -- at least not during acquisition and passive storage. The resulting representation from the acquisition phase did not include more information than was explicitly available during acquisition (see p. 1). Only when confronted with a problem-solving task did the students retrieve the original elements, integrate the semantic content, and then produce an inference. What was stored was the content that was acquired, which was no more than what was presented.

The integrative process was not spontaneous. Integration, so as to produce an inference, took place only when specifically cued. The cue was the recall task itself. Bartlett's (1932) memory research supported a reconstructive memory, but only when the subjects were asked to give the stories back in recall. The reconstruction and elaboration was a performance ability manifest only during retrieval and not necessarily a spontaneous and autonomous activity prior to that time.

The process of recall may be considered a problem-solving task. Given a problem-solving task which demands integration and inference production, the retrieval phase appears to be a very active and generative process. Semantic integration, while being psychologically real, is not automatic. Memory is not reconstructive, recall is.

These findings support the observation by Moeser (1976) that integration is not a spontaneous, natural phenomenon, in contrast to the findings of the 'Bransford Paradigm' studies. Discrete episodes were stored. Only with great difficulty was there evidence of integration.

Personal control

The control by the student over the mental processes involved in information processing appeared to be minimal. These data, particularly on instructional variation, suggested little personal, volitional control over semantic integrative processes. Moeser (1976) also found that instructions to integrate did not enhance integrative performance.

Lack of developmental trend

The fact that no qualitative developmental differences in inferential processing were found lends support to the stability of inferential processes within this range of ages. Further study would have to be done to extend the lower and upper limits of the age range to see if qualitative shifts occur.

Educational Implications

Learning appeared to be largely the acquisition of information. The drawing of conclusions and making inferences should be done in either of two ways: deliberately and explicitly providing them to the student during acquisition, or designing the recall task to demand this performance from the student. Discovery of inferences and conclusions -- the inductive leap,

perhaps -- does not appear to be a natural and spontaneous process. Thus efforts should be made to explicitly structure text passages for students.

Further, students in this age range may do not appear able to exercise volitional control over the way in which information is processed or integrated. The instructions to integrate, per se, apart from the method and sequencing of materials presentation had no influence at all.

While instructions might be able to control a student's attention to different stimuli and thus influence perceptual input to cognitive processing, it appears that instructions cannot be used to change the cognitive process itself. If instructions do not control this in a student, it may mean that the student, himself, does not have control over the underlying process that is going on.

One can attend or not attend to different stimuli, or do so in varying degrees, but the extent to which the student in this age range can control the <u>set</u> used for information processing appears to be very limited. The educational implication is that the teacher and learning materials should not expect to alter "the way a student thinks" by merely imploring the student or reader to do so.

There seems to be little drawing of inferences unless there are cues which closely direct inferencing. Students who were not cued to draw inferences or were only told in a general way to form inferences did not differ from one another. Inferencing appeared not to occur until specific questions were asked, such as during retrieval. These data suggest that when teaching children to read for maximum comprehension, the material and teacher must structure the task very carefully if particular inferences are to be drawn. General instructions to draw inferences apparently are not very effective. The question of the best way to teach this inferencing process was not addressed in this study, but is a topic in need of study in light of the present results.

REFERENCES

REFERENCES

Anderson, J. R., & Bower, G. H. Configural properties in sentence memory. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 594-605.

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

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

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. J. 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 and Stratton, 1963.

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

Ausubel, D. P., Schpoont, S. H., & Cukier, L. The influence of intention on the retention of school materials. Journal of Educational Psychology, 1957, 48, 87-92.

Barclay, J. R., & Reid, M. Semantic integration in children's recall of discourse. Developmental Psychology, 1974, 10, 277-281.

Bartlett, F. C. Remembering. Cambridge: Cambridge University Press, 1932.

Begg, I., & Paivio, A. Concreteness and imagery in sentence memory. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 821-827.

Blumenthal A. Prompted recall of sentences. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 203-206.

Blumenthal, A., & Boakes, R. Prompted recall of sentences: A further study. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 674-675.

Bransford, J. D., & Franks, J. J. Temporal integration in the acquisition of complex ideas. In "Understanding Sentences", a symposium chaired by J. J. Jenkins and presented at the Midwestern Psychological Association, Cincinnati, May 1970a.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas. University of Minnesota, 1970b.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas: A review. Cognition: International Journal of Cognitive Psychology, 1972, 1, 211-249.

Bransford, J. D., & McCarrell, N. S. A sketch of a cognitive approach to comprehension: Some thoughts about what it means to comprehend. In W. B. Weimer & D. S. Palermo (Eds.), Cognition and symbolic processes. New York: Winston and Sons, 1974.

Bransford, J. D., Barclay, J. R., & Franks, J. J. Sentence memory: A constructive versus interpretive approach. Cognitive Psychology, 1972, 3, 193-209.

Bruner, J. S. Going beyond the information given. In Contemporary approaches to cognition. Cambridge, Mass.: Harvard University Press, 1957.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 240-247.

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

Curnow, P. F. Integration of linguistic materials. Unpublished Ph. D. Dissertation, University of Minnesota, 1969.

Dooling, D. J., & Lachman, R. Effects of comprehension on retention of prose. Journal of Experimental Psychology, 1971, 88, 216-212.

Fillenbaum, S. Psycholinguistics. In P. H. Mussen and M. R. Rosenzweig (Eds.), Annual review of psychology. Vol. 22. Palo Alto, California: Annual Reviews Inc., 1971.

Forster, K. I. The effect of syntactic structure on non-ordered recall. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 292-297.

Franks, J. J., & Bransford, J. D. Abstraction of visual patterns. Journal of Experimental Psychology, 1971, 90, 65-74.

Franks, J. J., & Bransford, J. D. The acquisition of abstract ideas. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 311-315.

Franks, J. J., & Bransford, J. D. A brief note on linquistic integration. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 217-219.

Hale, G. A., & Piper, R. A. Effect of pictorial integration on children's incidental learning. Developmental Psychology, 1974, 10(6), 847-851.

Harlow, H. F. The formation of learning sets. Psychological Review, 1949, 56, 51-65.

Hunt, E. B., Marin, J., & Stone, P. Experiments in induction. New York: Academic Press, 1966.

Johnson, M. K., Bransford, J. D., & Solomon, S. K. Memory for tacit implications of sentences. Journal of Experimental Psychology, 1973, 98, 203-205.

Johnson, M. K., Bransford, J. D., Nyberg, S. E., & Cleary, J. J. Comprehension factors in interpreting memory for abstract and concrete sentences. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 451-454.

Katona, G. Organizing and memorizing. New York: Columbia University Press, 1940.

Katz, S. The role of representational information and instructional set in the integration of semantic ideas. Unpublished doctoral dissertation, University of Michigan, 1971.

Katz, S. Role of instructions in the abstraction of linguistic ideas. _Journal of Experimental Psychology_, 1973, _98_, 79-84.

Katz, S., & Gruenewald, P. The abstraction of ideas in "meaningless" sentences. _Memory and Cognition_, 1974, _2_, 737-741.

Kintsch, W. Notes on the structure of semantic memory. In E. Tulving and W. Donaldson (Eds.), _Organization of memory_. New York: Academic Press, 1972.

Kintsch, W., & Monk, D. Storage of complex information in memory: Some implication of the speed with which inferences can be made. _Journal of Experimental Psychology_, 1972, _94_, 25-32.

Koffka, K. _Principles of gestalt psychology_. New York: Harcourt, Brace and World, 1935.

Lorge, I. _The Lorge formula for estimating difficulty of reading materials_. New York: Bureau of Publications, Teachers College, Columbia University, 1959.

Mehler, J. Some effects of grammatical transformations on the recall of English sentences. _Journal of Verbal Learning and Verbal Behavior_, 1963, _2_, 346-351.

Miller, G. A., & Selfridge, J. A. Verbal context and the recall of meaningful material. _American Journal of Psychology_, 1950, _63_, 176-185.

Nilsson, N. J. _Learning machines_. New York: McGraw-Hill, 1965.

Oldfield, R. C. Memory mechanisms and the theory of schemata. _British Journal of Psychology_, 1954, _45_, 14-23.

Oldfield, R. C., & Zangwill, O. L. Head's concept of the schema and its application in contemporary British psychology. British Journal of Psychology, 1942, 32, 267-286.

Osgood, C. E. The similarity paradox in human learning: A resolution. Psychological Review, 1949, 56, 132-243.

Peterson, R. G., & McIntyre. C. W. The influence of semantic "relatedness" on linguistic integration and retention. American Journal of Psychology, 1973, 86(4), 697-706.

Piaget, J., & Inhelder, B. The psychology of the child. New York: Basic Books, Inc., 1969.

Pompi, K. F., & Lachman, R. Surrogate processes in the short term retention of connected discourse. Journal of Experimental Psychology, 1967, 75, 143-150.

Posner, M. I. Abstraction and the process of recognition. In G. H. Bower and J. J. Spence (Eds.), Psychology of learning and motivation. Vol. 3. New York: Academic Press, 1969.

Posner, M. I., & Keele, S. W. On the genesis of abstract ideas. Journal of Experimental Psychology, 1968, 77, 353-363.

Posner, M. I., & Keele, S. W. Retention of abstract ideas. Journal of Experimental Psychology, 1970, 83, 304-308.

Posner, M. I., Goldsmith, R., & Welton, K. E. Perceived distance and the classification of distorted patterns. Journal of Experimental Psychology, 1967, 73, 28-38.

Quillian, M. R. Semantic memory. In M. Minsky (Ed.), Semantic information processing. Cambridge, Mass.: M. I. T. Press, 1968.

Quillian, M. R. The teachable language comprehender: A simulation program and theory of language. Communications of the ACM, 1969, 12, 459-476.

Reitman, J. S., & Bower, G. H. Storage and later recognition of exemplars of concepts. Cognitive Psychology, 1973, 4, 194-206.

Rohrman, N. L. The role of syntactic structure in the recall of English nominalization. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 904-912.

Sachs, J. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 1967, 2, 437-442.

Sebestyen, G. Decision making processes in pattern recognition. New York: McGraw-Hill, 1965.

Shepard, R. N. Recognition memory for words, sentences, and pictures. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 156-163.

Singer, M., & Rosenberg, S. T. The role of grammatical relations in the abstraction of linguistic ideas. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 273-284.

Slobin, D. I. Recall of full and truncated passive sentences in connected discourse. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 876-881.

Strange, W., Keeney, T., Kessel, F. S., & Jenkins, J. J. Abstraction over time of prototypes from distortions of random dot patterns. Journal of Experimental Psychology, 1970, 83, 508-510.

Underwood, B. J., & Schulz, R. W. Meaningfulness and verbal learning. New York: Lippincott, 1960.

Wertheimer, M. Productive thinking (Enlarged ed.: Michael Wertheimer, Ed.). New York: Harper, 1959.

Yuille, J. C., & Paivio, A. Abstractness and the recall of connected discourse. Journal of Experimental Psychology, 1969, 82, 467-471.

APPENDICES

APPENDIX A

MATERIALS FOR EXPERIMENT I

Set 1

Filler 1

When a man enters the great city, he is called to see the king. The man is sent to the castle tower if he does not give the king a gold ring. Simon entered the great city without a gold ring. Simon was sent to the castle tower.

Question: Was Simon sent to the castle tower?
Answer: Yes.

Filler 2

A poor food diet can make people sick. A poor food diet is when you have only one or two kinds of food to eat. When a person gets sick the stomach swells up. Meg has a food diet of only bread and water. Meg has a red dress.

Question: Will Meg's stomach be alright?
Answer: No.

Paragraph 1, Logic Type 1, Inference Supplied

Every year some birds lose their feathers. The birds can not fly until their feathers grow back. During this time doctors catch them to see if they are sick. Sick birds are taken away and killed. Canadian geese lose their feathers every year. The sick Canadian geese are killed.

Question: Do the sick Canadian geese get better?
Answer: No.

Paragraph 2, Logic Type 3, No Inference Supplied

Talking is needed so lions can work together. Lions talk to each other by using the same language. By using the same language they know what other lions are saying. Many lions must work together to catch enough food. If they do not work together they will not catch enough food. There are ten lions together. Each lion speaks a different language. The lion is king in his land.

Question: Will the lions catch enough food?
Answer: No.

Paragraph 3, Logic Type 2, Inference Supplied

An animal with brown and green spots can hide from its enemies. If an animal can hide from its enemies then it can walk far to find food. With enough food an animal can live a long life. The fox does not have brown and green spots. The fox lived only a short life.

Question: Did the fox live only a short life?
Answer: Yes.

Paragraph 4, Logic Type 1, No Inference Supplied

The sister of Karnak died and went to the Valley of the Dead. Karnak went there to bring her back to the Land of the Living. Karnak had to lead her back but he could not look at her. If he turned to look at her, she would be lost forever. Karnak was so happy he turned to help her out. His sister was very beautiful.

Question: Was the sister of Karnak lost forever?
Answer: Yes.

Paragraph 5, Logic Type 2, No Inference Supplied

When people do not eat breakfast they get hungry later in the day. When they get hungry later in the day they eat too much food. Too much food makes people fat. Bobby does not like to eat breakfast. Bobby wears black shoes.

Question: Will Bobby get fat?
Answer: Yes.

Paragraph 6, Logic Type 3, Inference Supplied

Children do not like to do hard work. They become very unhappy. Cleaning an attic is very hard work. When people are unhappy with their work they do a poor job. Susan and Mary are ten years old. They had to clean the attic. They did a poor job.

Question: Did Susan and Mary do a good job?
Answer: No.

Paragraph 7, Logic Type 1, No Inference Supplied

Every Christmas there is a winter party. The iceboats have a race on the lake as long as the ice is one foot thick. All winter the ice on the lake was half a foot thick. The winter party is held in the mountains.

Question: Were the iceboat races held this year?
Answer: No.

Paragraph 8, Logic Type 2, No Inference Supplied

When Mrs. Brown needs milk she asks Johnny to go to the store. When Johnny goes to the store he walks by Mike's house. When he walks by Mike's house he stops to talk to Mike's sister. Mrs. Brown does not need any milk today. Johnny has a new, brown and white dog.

Question: Will Johnny talk to Mike's sister today?
Answer: No.

Paragraph 9, Logic Type 3, Inference Supplied

All over the world animals are used to run machines. An ox is an animal used to run machines. A machine that the ox runs is a water wheel. The water wheel gets water from a deep well. Abdul the farmer does not have an ox. Abdul's farm will not have any water.

Question: Will Abdul's farm go without water?
Answer: Yes.

Paragraph 10, Logic Type 1, Inference Supplied

The Sphynx waited at the gate of the old city. He asked every new visitor a riddle. The Sphynx would kill himself if the visitor gave the right answer. Today a new visitor gave the right answer to the Sphynx. The Sphynx then killed himself.

Question: Did the Sphynx kill himself?
Answer: Yes.

Paragraph 11, Logic Type 3, No Inference Supplied

If you eat too much yellow food it can hurt your body. People should eat only one ear of corn each day. If you eat more your body will be hurt because your hair falls out. Sherry eats three ears of corn every day. Sherry has a baby brother.

Question: Will Sherry's hair fall out?
Answer: Yes.

Paragraph 12, Logic Type 2, Inference Supplied

When you put a fish with a turtle, the turtle will think it is food. When a turtle sees food it comes close and opens its mouth. If the food is too big, the turtle will tear it up. Betty put her fish with her turtle. The turtle ate the fish.

Question: Did the fish live with the turtle?
Answer: No.

MATERIALS FOR EXPERIMENT I

Set 2

Filler 1

When a man enters the great city, he is called to see the king. The man is sent to the castle tower if he does not give the king a gold ring. Simon entered the great city without a gold ring. Simon was sent to the castle tower.

Question: Was Simon sent to the castle tower?
Answer: Yes.

Filler 2

A poor food diet can make people sick. A poor food diet is when you have only one or two kinds of food to eat. When a person gets sick the stomach swells up. Meg has a food diet of only bread and water. Meg has a red dress.

Question: Will Meg's stomach be alright?
Answer: No.

Paragraph 1, Logic Type 2, No Inference Supplied

An animal with brown and green spots can hide from its enemies. If an animal can hide from its enemies then it can walk far to find food. With enough food an animal can live a long life. The fox does not have brown and green spots. A small river runs through the woods.

Question: Did the fox live only a short life?
Answer: Yes.

Paragraph 2, Logic Type 3, Inference Supplied

Talking is needed so lions can work together. Lions talk to each other by using the same language. By using the same language they know what other lions are saying. Many lions must work together to catch enough food. If they do not work together they will not catch enough food. There are ten lions together. Each lion speaks a different language. The lions will not catch enough food.

Question: Will the lions catch enough food?
Answer: No.

Paragraph 3, Logic Type 1, No Inference Supplied

Every year some birds lose their feathers. The birds can not fly until their feathers grow back. During this time doctors catch them to see if they are sick. Sick birds are taken away and killed. Canadian geese lose their feathers every year. Canadian geese have long thin necks.

Question: Do the sick Canadian geese get better?
Answer: No.

Paragraph 4, Logic Type 3, No Inference Supplied

All over the world animals are used to run machines. An ox is an animal used to run machines. A machine that the ox runs is a water wheel. The water wheel gets water from a deep well. Abdul the farmer does not have an ox. Abdul works the farm with his brother.

Question: Will Abdul's farm go without water?
Answer: Yes.

Paragraph 5, Logic Type 2, Inference Supplied

When Mrs. Brown needs milk she asks Johnny to go to the store. When Johnny goes to the store he walks by Mike's house. When he walks by Mike's house he stops to talk to Mike's sister. Mrs. Brown does not need any milk today. Johnny will not talk to Mike's sister today.

Question: Will Johnny talk to Mike's sister today?
Answer: No.

Paragraph 6, Logic Type 1, Inference Supplied

Every Christmas there is a winter party. The iceboats have a race on the lake as long as the ice is one foot thick. All winter the ice on the lake was half a foot thick. The iceboat races were not held this year.

Question: Were the iceboat races held this year?
Answer: No.

Paragraph 7, Logic Type 7, No Inference Supplied

When you put a fish with a turtle, the turtle will think it is food. When a turtle sees food it comes close and opens its mouth. If the food is too big, the turtle will tear it up. Betty put her fish with her turtle. The turtle had black spots.

Question: Did the fish live with the turtle?
Answer: Yes.

Paragraph 8, Logic Type 1, No Inference Supplied

The Sphynx waited at the gate of the old city. He asked every new visitor a riddle. The Sphynx would kill himself if the visitor gave the right answer. Today a new visitor gave the right answer to the Sphynx. The old city had walls.

Question: Did the Sphynx kill himself?
Answer: Yes.

Paragraph 9, Logic Type 3, Inference Supplied

If you eat too much yellow food it can hurt your body. People should eat only one ear of corn each day. If you eat more your body will be hurt because your hair falls out. Sherry eats three ears of corn every day. Sherry's hair will fall out.

Question: Will Sherry's hair fall out?
Answer: Yes.

Paragraph 10, Logic Type 2, Inference Supplied

When people do not eat breakfast they get very hungry later in the day. When they get hungry later in the day they eat too much food. Too much food makes people fat. Bobby does not like to eat breakfast. Bobby will get fat.

Question: Will Bobby get fat?
Answer: Yes.

Paragraph 11, Logic Type 3, No Inference Supplied

Children do not like to do hard work. They become very unhappy. Cleaning an attic is very hard work. When people are unhappy with their work they do a poor job. Susan and Mary are ten years old. They had to clean the attic. Susan and Mary are sisters.

Question: Did Susan and Mary do a good job?
Answer: No.

Paragraph 12, Logic Type 1, Inference Supplied

The sister of Karnak died and went to the Valley of the Dead. Karnak went there to bring her back to the Land of the Living. Karnak had to lead her back but he could not look at her. If he turned to look at her, she would be lost forever. Karnak was so happy he turned to help her out. His sister was lost forever.

Question: Was the sister of Karnak lost forever?
Answer: Yes.

APPENDIX B

HYPOTHESES TO BE TESTED IN EXPERIMENT I

Given:

y_1 = encoding time;

y_2 = the number of correct responses to a query about an inference;

y_3 = time to produce a correct response to a query about an inference;

x_1 = age; and

x_2 = assignment of stimulus materials (0 = No inference supplied -- 1 = Inference supplied).

H_1: $\bar{y}_1$ for $x_2 = 0$ is greater than $\bar{y}_1$ for $x_2 = 1$.

H_2: $\bar{y}_3$ for $x_2 = 1$ is equal to $\bar{y}_3$ for $x_2 = 0$.

H_3: $\bar{y}_2$ for $x_2 = 1$ is equal to $\bar{y}_2$ for $x_2 = 0$.

H_4: $R_{y.1} > 0$, and $b_1 < 0$, where $y = y_1$.

H_5: $R_{y.1} > 0$, and $b_1 > 0$, where $y = y_2$.

H_6: $R_{y.1} > 0$, and $b_1 < 0$, where $y = y_3$.

H_7: $y_{1,x_2=0} - y_{1,x_2=1}$ for x_1 = greater observed value is greater than $y_{1,x_2=0} - y_{1,x_2=1}$ for x_1 = lesser

observed value. Expressed differently, $R^2_{y.1} > 0$, and $b_1 > 0$, where $y = y_{1,x_2=0} - y_{1,x_2=1}$.

H_8: $y_{2,x_2=1} - y_{2,x_2=0}$ for x_1 = lesser observed value is greater than $y_{2,x_2=1} - y_{2,x_2=1}$ for x_1 = greater observed value. Expressed differently, $R^2_{y.1} > 0$, and $b_1 < 0$, where $y = y_{2,x_2=1} - y_{2,x_2=0}$.

APPENDIX C

MATERIALS FOR EXPERIMENT II

Filler 1

A poor food diet can make people sick. A poor food diet is when you have only one or two kinds of food to eat. When a person gets sick the stomach swells up. Meg has a food diet of only bread and water. Meg has a red dress.

Question (Actual): Does Meg have a green dress?
Answer: No.

Question (Inference): Will Meg's stomach be alright?
Answer: No.

Filler 2

When a man enters the great city, he is called to see the king. The man is sent to the castle tower if he does not give the king a gold ring. Simon entered the great city without a gold ring. Simon wanted to see the animal fair.

Question (Inference): Was Simon sent to the castle tower?
Answer: Yes.

Question (Actual): Did Simon enter the great city?
Answer: Yes.

Paragraph 1, Logic Type 2

When people do not eat breakfast they get very hungry later in the day. When they get hungry later in the day they eat too much food. Too much food makes people fat. Bobby does not like to eat breakfast. Bobby wears black shoes.

Question (Inference): Will Bobby get fat?
Answer: Yes.

Question (Actual): Do hungry people stop eating?
Answer: No.

Paragraph 2, Logic Type 3

Talking is needed so lions can work together. Lions talk to each other by using the same language. By using the same language they know what other lions are saying. Many lions must work together to catch enough food. If they do not work together they will not catch enough food. There are ten lions together. Each lion speaks a different language. The lion is king in his land.

Question (Actual): Are there ten lions together?
Answer: Yes.

Question (Inference): Will the lions catch enough food?
Answer: No.

Paragraph 3, Logic Type 2

When you put a fish with a turtle, the turtle will think it is food. When a turtle sees food it comes close and opens its mouth. If the food is too big, the turtle will tear it

up. Betty put her fish with her turtle. The turtle had black spots.

Question (Actual): Will the turtle go away from food?
Answer: No.

Question (Inference): Did the fish live with the turtle?
Answer: No.

Paragraph 4, Logic Type 2

When Mrs. Brown needs milk she asks Johnny to go to the store. When Johnny goes to the store he walks by Mike's house. When he walks by Mike's house he stops to talk to Mike's sister. Mrs. Brown does not need any milk today. Johnny has a new, brown and white dog.

Question (Actual): Will Johnny talk to Mike's sister today?
Answer: No.

Question (Inference): Does Mike get milk for Mrs. Brown?
Answer: No.

Paragraph 5, Logic Type 3

If you eat too much yellow food it can hurt your body. People should eat only one ear of corn each day. If you eat more your body will be hurt because your hair falls out. Sherry eats three ears of corn every day. Sherry has a baby brother.

Question (Actual): Will too much yellow food hurt your body?
Answer: Yes.

Question (Inference): Will Sherry's hair fall out?
Answer: Yes.

Paragraph 6, Logic Type 4

Every year some birds lose their feathers. The birds can not fly until their feathers grow back. During this time doctors catch them to see if they are sick. Sick birds are taken away and killed. Canadian geese lose their feathers every year. Canadian geese have long thin necks.

Question (Inference): Do the sick Canadian geese get better?
Answer: No.

Question (Actual): Can the birds fly without their feathers?
Answer: No.

Paragraph 7, Logic Type 1

Every Christmas there is a winter party. The iceboats have a race on the lake as long as the ice is one foot thick. All winter the ice on the lake was half a foot thick. The winter party is held in the mountains.

Question (Actual): Is there a winter party every Christmas?
Answer: Yes.

Question (Inference): Were the iceboat races held this year?
Answer: No.

Paragraph 8, Logic Type 1

The sister of Karnak died and went to the Valley of the Dead. Karnak went there to bring her back to the Land of the Living. Karnak had to lead her back but he could not look at her. If he turned to look at her, she would be lost forever. Karnak was so

happy he turned to help her out. His sister was very beautiful.

Question (Inference): Was the sister of Karnak lost forever?
Answer: Yes.

Question (Actual): Did Karnak die in the Valley of the Dead?
Answer: No.

Paragraph 9, Logic Type 2

An animal with brown and green spots can hide from its enemies. If an animal can hide from its enemies then it can walk far to find food. With enough food an animal can live a long life. The fox does not have brown and green spots. A small river runs through the woods.

Question (Inference): Did the fox live only a short life?
Answer: Yes.

Question (Actual): Do brown and green spots hide an animal?
Answer: Yes.

Paragraph 10, Logic Type 1

The Sphynx waited at the gate of the old city. He asked every new visitor a riddle. The Sphynx would kill himself if the visitor gave the right answer. Today a new visitor gave the right answer to the Sphynx. The old city had walls.

Question (Inference): Did the Sphynx kill himself?
Answer: Yes.

Question (Actual): Did the Sphynx ask riddles?
Answer: Yes.

Paragraph 11, Logic Type 3

Children do not like to do hard work. They become very unhappy. Cleaning an attic is very hard work. When people are unhappy with their work they do a poor job. Susan and Mary are ten years old. They had to clean the attic. Susan and Mary are sisters.

Question (Actual): Is cleaning an attic hard work?
Answer: Yes.

Question (Inference): Did Susan and Mary do a good job?
Answer: No.

Paragraph 12, Logic Type 3

All over the world animals are used to run machines. An ox is an animal used to run machines. A machine that the ox runs is a water wheel. The water wheel gets water from a deep well. Abdul the farmer does not have an ox. Abdul works the farm with his brother.

Question (Actual) Are animals not allowed to run machines?
Answer: No.

Question (Inference): Will Abdul's farm go without water?
Answer: Yes.

APPENDIX D

HYPOTHESES TO BE TESTED IN EXPERIMENT II

Given:

y_1 = encoding time;

y_2 = the number of correct responses to a query about an inference;

y_3 = time to produce a correct response to a query about an inference;

x_1 = age; and

x_2 = instruction condition (1 = Integrative -- 0 = Verbatim).

H_9: $R^2_{y.2} > 0$, and $b_1 > 0$, where $y = y_2$.

H_{10}: $R^2_{y.2} > 0$, and $b_1 < 0$, where $y = y_3$.

H_{11}: $R^2_{y.2} = 0$, where $y = y_2$.

H_{12}: $R^2_{y.1} > 0$, and $b_1 < 0$, where $y = y_1$.

APPENDIX E

RAW DATA FOR EXPERIMENT I

Student	Age[c]	Encoding Time[a]		Correct Response[b]		Response Time[a]	
		IS	NIS	IS	NIS	IS	NIS
Group 1							
1	121	17.46	18.80	6	4	3.17	2.40
2	121	25.97	29.61	6	4	4.26	4.65
3	115	16.97	19.73	6	5	4.50	4.00
4	118	15.80	17.13	6	3	2.47	2.27
5	133	7.44	9.24	5	3	1.60	1.80
6	122	21.43	24.47	6	5	3.67	3.00
7	126	16.37	17.63	6	5	2.80	2.72
8	133	12.93	14.10	3	2	1.93	1.90
9	136	12.93	13.80	6	4	1.70	1.84
10	143	25.03	26.50	5	2	3.59	3.10
11	143	12.43	12.77	6	3	2.80	2.60
12	135	15.53	15.87	6	2	4.73	2.80
13	145	13.27	15.04	6	5	2.77	2.16
14	140	14.17	14.97	6	4	2.10	1.65
15	152	16.63	18.06	5	5	3.04	3.24
16	146	16.06	17.16	5	4	2.28	1.80
17	149	9.73	10.33	5	4	1.48	1.55
18	175	16.90	18.70	6	3	2.90	2.27
19	146	12.33	14.10	5	4	1.84	2.05
20	157	12.37	13.23	2	3	1.50	3.07
21	147	13.37	14.63	4	4	1.60	2.30
22	164	25.67	26.67	5	3	3.40	3.20
23	163	16.13	19.03	6	4	3.03	3.36
24	163	10.67	10.73	6	4	1.70	1.66
25	162	11.40	12.16	4	3	1.60	2.33
26	168	12.70	13.66	6	3	2.67	2.53
27	160	17.83	18.43	4	2	2.45	4.80
28	167	16.93	17.33	5	4	1.92	1.85
29	177	14.00	16.00	5	2	1.92	1.60

RAW DATA FOR EXPERIMENT I (CONTINUED)

Group 2							
30	116	16.20	14.20	5	5	1.56	2.56
31	124	24.53	23.37	5	4	2.36	3.25
32	120	19.40	19.03	4	2	4.00	3.80
33	115	16.83	14.80	5	4	1.88	3.85
34	132	18.26	24.10	4	4	1.60	4.75
35	133	18.23	18.67	6	4	2.27	4.85
36	126	13.33	13.50	5	2	2.16	3.00
37	134	16.33	14.57	5	3	1.68	2.60
38	133	15.80	14.80	5	3	2.68	2.27
39	135	17.40	16.26	4	5	2.45	3.64
40	145	21.33	19.13	3	2	2.20	2.80
41	139	12.20	9.37	6	0	0.83	----
42	139	11.43	11.67	4	2	1.85	2.30
43	155	14.43	13.17	5	4	2.04	2.90
44	159	12.03	11.90	6	3	1.33	1.27
45	154	10.23	9.83	6	5	1.93	2.04
46	144	12.33	10.83	5	4	2.00	3.45
47	152	21.57	21.00	6	4	3.27	6.30
48	144	10.70	10.37	6	4	1.17	1.75
49	150	17.43	20.10	6	5	1.97	3.80
50	148	12.27	13.06	6	4	1.13	2.05
51	148	12.43	12.13	4	2	2.06	2.80
52	158	12.63	11.66	6	2	1.63	2.10
53	164	9.00	9.17	6	2	1.33	1.60
54	171	32.86	33.03	4	2	3.15	4.90
55	163	10.60	10.60	5	2	1.80	2.10
56	168	16.90	15.57	3	0	2.47	----
57	164	11.17	11.13	5	3	1.96	2.60
58	168	12.86	13.00	4	2	2.10	2.70

[a]Measured in seconds

[b]For inference queries

[c]Measured in months

APPENDIX F

RAW DATA FOR EXPERIMENT I -- NUMBER CORRECT BY LOGIC TYPE

	IS			NIS		
	Logic Type			Logic Type		
Student	1	2	3	1	2	3
Group 1						
1	2	2	2	1	2	1
2	2	2	2	1	2	1
3	2	2	2	1	2	2
4	2	2	2	0	2	1
5	2	2	1	1	1	1
6	2	2	2	2	2	1
7	2	2	2	1	2	2
8	2	0	1	0	2	0
9	2	2	2	1	2	1
10	2	2	1	0	1	1
11	2	2	2	1	1	1
12	2	2	2	0	1	1
13	2	2	2	2	2	1
14	2	2	2	1	2	1
15	2	1	2	2	2	1
16	2	2	1	2	1	1
17	1	2	2	2	1	1
18	2	2	2	1	1	1
19	1	2	2	2	1	1
20	1	0	1	1	2	0
21	1	2	1	1	2	1
22	2	2	1	1	1	1
23	2	2	2	1	2	1
24	2	2	2	1	2	1
25	1	2	1	0	2	1
26	2	2	2	0	2	1
27	1	2	1	0	1	1
28	2	2	1	1	2	1
29	2	1	2	0	1	1

RAW DATA FOR EXPERIMENT I -- NUMBER CORRECT...(CONTINUED)

Group 2

30	2	1	2	1	2	2
31	2	2	1	2	1	1
32	1	2	1	1	0	1
33	2	2	1	2	2	0
34	2	0	2	2	1	1
35	2	2	2	2	1	1
36	2	2	1	1	0	1
37	2	2	1	1	1	1
38	2	2	1	1	1	1
39	1	1	2	1	2	2
40	1	1	1	0	1	1
41	2	2	2	0	0	0
42	2	1	1	2	0	0
43	2	2	1	1	2	1
44	2	2	2	1	1	1
45	2	2	2	2	2	1
46	2	1	2	2	1	1
47	2	2	2	1	1	2
48	2	2	2	1	2	1
49	2	2	2	2	2	1
50	2	2	2	1	2	1
51	2	1	1	1	1	0
52	2	2	2	1	0	1
53	2	2	2	1	1	0
54	2	0	2	0	1	1
55	2	2	1	2	0	0
56	1	2	0	0	0	0
57	2	2	1	0	2	1
58	1	2	1	0	1	1

RAW DATA FOR EXPERIMENT I -- ENCODING TIME BY LOGIC TYPE

	IS			NIS		
	Logic Type			Logic Type		
Student	1	2	3	1	2	3
Group 1						
1	17.3	16.9	18.2	20.9	15.6	19.9
2	26.2	24.1	27.6	32.0	25.7	29.6
3	17.7	16.5	16.7	22.1	14.4	22.7
4	14.4	17.8	15.2	19.8	13.5	18.1
5	6.9	7.6	7.8	8.9	7.8	11.0
6	18.3	21.7	24.3	22.8	23.3	27.3
7	16.4	16.4	16.3	16.8	16.4	19.7
8	12.5	13.5	12.8	13.2	12.7	16.4
9	12.8	13.2	12.8	13.3	12.0	16.1
10	24.3	24.5	26.3	25.1	23.6	30.8
11	13.0	12.8	11.5	10.2	12.6	15.5
12	15.2	15.3	16.1	16.4	14.4	16.8
13	13.0	14.0	12.8	16.5	13.0	15.6
14	13.7	14.6	14.2	14.7	12.9	17.3
15	16.3	16.6	17.0	19.1	15.5	19.6
16	14.4	17.2	16.6	18.6	13.7	19.2
17	9.2	10.4	9.6	10.5	9.7	10.8
18	16.2	19.4	15.1	22.0	13.5	20.6
19	11.3	13.4	12.3	13.5	11.0	17.9
20	12.4	12.4	12.3	14.0	11.2	14.5
21	13.3	13.4	13.4	13.3	13.3	17.3
22	27.0	29.3	20.7	27.3	20.4	32.3
23	14.7	17.9	15.8	20.9	16.9	19.3
24	10.1	10.4	11.5	9.9	9.7	12.6
25	9.7	14.0	10.5	12.5	11.6	12.4
26	12.7	12.7	12.7	12.8	12.1	16.1
27	17.4	16.5	19.6	19.7	15.6	20.0
28	14.3	17.9	18.6	17.9	14.1	20.0
29	12.9	16.9	12.2	14.3	16.3	17.4

RAW DATA FOR EXPERIMENT I -- ENCODING TIME...(CONTINUED)

Group 2

30	17.2	14.4	17.0	14.2	15.0	13.4
31	26.0	21.1	26.5	22.4	22.2	25.5
32	18.7	16.5	23.0	19.2	19.7	18.2
33	14.9	14.0	21.6	14.3	14.8	15.3
34	19.7	18.0	17.1	23.0	28.3	21.0
35	16.9	16.4	21.4	17.9	18.0	20.1
36	13.1	12.3	14.6	12.9	14.1	13.5
37	15.9	13.4	19.7	13.5	14.8	15.4
38	14.5	15.2	17.7	13.9	15.3	15.2
39	16.7	16.4	19.1	17.1	17.0	14.7
40	15.5	21.2	27.3	20.0	18.7	18.7
41	11.5	10.9	14.2	8.2	9.6	10.6
42	12.4	9.9	12.0	12.8	10.0	12.2
43	12.8	14.7	15.8	12.2	13.5	13.8
44	11.6	10.7	13.8	11.9	13.0	10.8
45	9.4	9.4	11.9	9.7	10.1	9.7
46	12.3	12.0	12.7	10.2	11.2	11.1
47	22.4	18.9	23.4	21.4	21.5	20.1
48	10.4	9.5	12.2	10.7	10.6	9.8
49	16.5	15.3	20.5	19.5	21.6	19.2
50	10.4	12.2	14.2	13.9	12.6	12.7
51	11.0	12.1	14.2	13.2	12.6	10.6
52	10.9	12.6	14.4	11.2	12.4	11.4
53	8.7	7.7	10.6	9.2	9.1	9.2
54	37.1	29.7	31.8	28.0	36.4	34.7
55	9.0	10.7	12.1	12.0	10.1	9.7
56	15.7	15.8	19.2	15.7	15.3	15.7
57	11.0	10.0	12.5	11.5	11.4	10.4
58	12.4	11.2	15.0	12.8	13.6	12.6

RAW DATA FOR EXPERIMENT I -- RESPONSE TIME BY LOGIC TYPE

	IS			NIS		
	Logic Type			Logic Type		
Student	1	2	3	1	2	3
Group 1						
1	2.5	2.5	4.5	2.8	2.7	1.4
2	4.7	3.8	4.9	4.2	5.3	3.8
3	2.7	3.6	7.4	7.8	2.6	3.5
4	2.3	1.9	3.2		2.6	1.6
5	2.1	1.3	1.2	1.4	2.4	1.6
6	3.4	3.1	4.5	3.1	3.0	2.8
7	2.8	2.7	2.9	2.8	2.4	3.0
8	2.1		1.6		1.9	
9	1.7	2.0	1.4	3.2	1.6	1.0
10	3.7	3.7	3.2		2.2	4.0
11	2.9	2.8	2.7	3.4	3.2	1.2
12	8.3	3.0	2.9		4.0	1.6
13	2.2	3.1	3.0	2.2	2.4	1.6
14	2.0	2.0	2.3	1.6	1.8	1.4
15	3.6	2.8	2.3	3.9	3.3	2.0
16	2.3	2.0	2.7	1.9	1.8	1.6
17	1.3	1.5	1.8	1.6	1.4	1.6
18	2.2	4.1	2.4	3.4	1.6	1.8
19	1.2	1.9	2.1	2.5	1.8	1.4
20	1.2		1.8	3.8	2.7	
21	0.8	1.9	1.8	2.2	2.7	1.6
22	3.6	3.5	2.8	3.6	3.4	2.6
23	2.4	3.2	3.5	3.6	4.1	1.6
24	1.7	1.7	1.7	2.0	1.6	1.4
25	1.4	1.9	1.2		2.8	1.4
26	2.6	2.9	2.5		2.3	3.0
27	2.4	2.9	3.0		4.4	5.2
28	2.1	2.0	1.4	2.6	2.1	0.6
29	1.5	1.8	2.4		2.2	1.0

RAW DATA FOR EXPERIMENT I -- RESPONSE TIME...(CONTINUED)

Group 2

30	1.7	1.4	1.5	1.6	2.5	2.8
31	2.9	2.1	1.8	3.3	3.2	3.2
32	3.0	4.4	4.2	5.0		2.6
33	1.8	1.9	2.0	4.0	3.7	
34	1.8		1.4	3.4	9.8	2.4
35	2.5	2.7	1.7	6.2	2.6	4.4
36	2.4	1.8	2.4	3.4		2.6
37	1.6	1.8	1.6	2.0	2.6	3.2
38	2.8	2.8	2.4	2.4	2.0	2.4
39	2.2	2.4	2.6	2.4	3.9	4.0
40	2.8	1.4	2.4		3.0	2.6
41	0.8	0.8	0.9			
42	2.1	1.4	1.8	2.3		
43	2.8	1.7	1.2	1.8	3.6	2.6
44	1.5	1.1	1.4	0.8	1.4	1.6
45	1.9	1.9	2.0	2.6	2.5	1.2
46	1.5	1.4	2.8	2.3	3.2	2.0
47	2.7	3.3	3.8	8.4	3.8	6.5
48	1.4	1.1	1.0	1.2	2.3	1.2
49	2.1	2.0	1.8	3.7	4.6	2.4
50	1.2	1.0	1.2	0.8	2.8	1.8
51	2.2	1.0	2.8	3.2	2.4	
52	1.4	1.9	1.6	1.4		2.8
53	1.5	1.1	1.4	1.4	1.8	
54	3.5		2.8		4.2	5.6
55	1.6	2.0	1.8	2.1		
56	2.6	2.4				
57	2.0	2.1	1.6		3.0	1.8
58	2.6	1.7	2.4		2.8	2.6

APPENDIX G

RAW DATA FOR EXPERIMENT II

Student	Age[a]	Encoding Time[b]	Correct Response[c]	Response Time[b]
Integrative Instructions				
1	111	16.20	7	2.74
2	111	31.82	7	4.00
3	125	35.77	10	5.82
4	121	22.02	5	5.48
5	122	19.20	5	4.36
6	126	27.48	6	3.83
7	123	17.90	6	2.90
8	129	20.00	11	3.96
9	132	19.37	5	3.52
10	124	11.87	7	1.69
11	121	11.43	5	2.36
12	138	18.38	4	2.75
13	139	9.92	6	1.23
14	141	9.63	7	2.06
15	140	12.40	6	2.07
16	139	11.07	5	1.96
17	156	13.05	8	2.00
18	155	16.45	7	3.53
19	153	14.48	9	1.67
20	148	12.18	9	2.00
21	155	13.75	6	2.57
22	152	16.62	8	3.35
23	156	27.20	4	3.70
24	165	20.40	9	2.82
25	164	10.68	5	1.52
26	139	15.25	6	2.10
27	163	10.12	7	1.77
28	164	13.67	8	2.15
29	165	18.75	3	3.67
30	170	14.69	3	3.73

RAW DATA FOR EXPERIMENT II (CONTINUED)

Verbatim Instructions				
31	132	16.57	6	2.60
32	112	14.50	8	2.98
33	116	14.20	9	3.23
34	116	14.73	9	2.11
35	114	16.83	4	2.75
36	123	24.22	5	4.88
37	128	22.17	4	3.55
38	129	13.38	6	2.73
39	131	26.02	7	4.57
40	125	24.30	7	4.23
41	156	8.80	7	4.06
42	140	31.55	9	4.82
43	148	25.33	5	4.12
44	147	17.80	3	2.93
45	148	14.97	3	3.07
46	151	15.98	7	2.31
47	151	10.83	7	1.63
48	148	25.73	9	4.33
49	154	16.62	12	2.63
50	154	12.78	8	1.73
51	156	19.73	9	3.24
52	158	13.67	4	1.65
53	162	14.28	2	2.10
54	167	15.80	10	2.16
55	159	21.35	7	2.34
56	156	9.53	4	1.35
57	161	16.13	8	2.48
58	158	15.70	6	2.83
59	155	9.88	8	1.83

[a]Measured in months

[b]Measured in seconds

[c]For inference queries only

APPENDIX H

RAW DATA FOR EXPERIMENT II -- CORRECT RESPONSE BY LOGIC TYPE

Student	Logic Type 1	Logic Type 2	Logic Type 3
Verbatim Group			
1	3	1	2
2	2	4	2
3	4	2	3
4	3	3	3
5	1	1	2
6	2	1	2
7	2	1	1
8	1	3	2
9	2	3	2
10	2	2	2
11	2	3	2
12	3	3	3
13	0	3	2
14	0	3	0
15	1	0	2
16	2	1	4
17	1	4	2
18	3	3	3
19	4	4	4
20	3	3	2
21	2	4	3
22	1	1	2
23	0	1	1
24	2	3	4
25	2	2	3
26	2	2	1
27	4	3	1
28	2	1	3
29	3	2	3

RAW DATA FOR EXPERIMENT II --
CORRECT RESPONSE...(CONTINUED)

Integrative Group			
30	2	2	3
31	2	4	1
32	3	3	4
33	1	2	2
34	1	3	1
35	2	3	1
36	3	1	2
37	3	4	4
38	3	1	1
39	3	2	2
40	2	3	0
41	2	1	1
42	1	3	1
43	2	4	1
44	2	3	1
45	2	2	2
46	3	2	3
47	2	3	1
48	4	2	3
49	2	4	3
50	2	2	2
51	4	3	1
52	1	2	1
53	2	4	3
54	2	2	1
55	2	1	3
56	4	1	2
57	2	3	3
58	1	1	1
59	0	1	2

RAW DATA FOR EXPERIMENT II -- ENCODING TIME BY LOGIC TYPE

Student	Logic Type 1	2	3
Verbatim Group			
1	16.6	15.7	17.4
2	13,8	13.1	16.6
3	14.4	12.4	15.9
4	14.1	14.5	15.7
5	17.0	16.0	17.6
6	22.2	24.9	25.6
7	22.8	20.2	23.5
8	12.7	12.9	14.6
9	25.4	22.2	30.6
10	23.4	24.9	24.7
11	8.4	7.4	10.7
12	32.4	30.1	32.2
13	25.5	27.3	23.2
14	18.2	16.8	18.5
15	15.9	14.1	15.0
16	16.2	15.3	16.5
17	11.3	10.2	11.1
18	26.8	22.6	27.8
19	16.5	15.3	18.1
20	13.1	11.9	13.4
21	20.0	19.1	20.1
22	13.7	13.1	14.3
23	14.7	13.5	14.3
24	16.1	14.8	16.6
25	22.6	19.4	22.2
26	9.7	9.1	9.8
27	16.7	13.8	18.0
28	16.0	14.6	16.5
29	9.6	9.3	10.9

RAW DATA FOR EXPERIMENT II -- ENCODING TIME...(CONTINUED)

Integrative Group

30	16.6	15.2	16.9
31	33.0	28.0	34.5
32	36.5	32.5	38.3
33	22.6	20.7	22.8
34	20.1	18.6	18.9
35	29.2	25.4	27.9
36	19.3	15.7	18.7
37	21.1	17.8	21.2
38	18.6	20.0	19.6
39	12.7	11.1	11.9
40	10.0	12.3	12.1
41	17.3	18.2	19.7
42	9.7	9.4	10.7
43	9.8	8.8	10.4
44	13.4	11.5	12.4
45	11.7	9.8	11.8
46	12.8	13.1	13.4
47	17.7	13.9	17.8
48	14.4	13.8	15.4
49	13.3	10.5	12.8
50	13.5	13.6	14.3
51	16.8	16.7	16.5
52	27.6	24.4	29.7
53	19.6	20.3	21.4
54	10.9	10.6	10.6
55	15.2	14.5	16.2
56	10.4	9.9	10.1
57	13.8	13.0	14.3
58	20.3	16.7	19.4
59	15.0	13.4	15.8

RAW DATA FOR EXPERIMENT II -- RESPONSE TIME BY LOGIC TYPE

Student	Logic Type 1	2	3
Verbatim Group			
1	2.5	1.4	2.5
2	2.6	3.1	3.1
3	3.6	3.1	2.6
4	1.8	2.5	3.1
5	4.0	2.0	2.5
6	3.7	11.0	3.0
7	3.3	4.4	3.2
8	1.6	2.8	3.4
9	5.2	4.5	2.9
10	3.9	3.5	4.4
11	3.8	8.6	3.0
12	4.1	6.1	9.4
13		6.6	5.3
14		3.1	
15	4.0		2.6
16	2.5	1.8	2.3
17	0.6	2.0	1.3
18	4.2	2.7	6.4
19	2.1	2.9	2.6
20	1.6	1.4	1.8
21	3.6	3.1	3.3
22	1.8	1.6	1.6
23		1.8	2.4
24	2.5	1.9	2.3
25	2.2	1.9	2.7
26	2.2	1.0	1.2
27	2.5	2.4	2.2
28	2.9	2.6	2.6
29	1.9	1.9	1.8

RAW DATA FOR EXPERIMENT II -- RESPONSE TIME...(CONTINUED)

Integrative Group

30	3.1	2.4	2.7
31	5.4	3.6	2.6
32	6.4	4.6	5.9
33	6.4	4.8	5.7
34	2.8	6.0	2.6
35	3.6	4.0	2.6
36	2.4	4.0	3.0
37	4.0	3.8	4.0
38	3.6	4.8	2.6
39	1.7	1.5	1.9
40	1.9	2.4	
41	3.2	2.0	2.6
42	1.0	1.4	0.8
43	1.9	2.3	1.2
44	2.1	2.2	1.6
45	1.6	2.3	2.0
46	2.3	2.2	1.6
47	2.3	4.9	1.8
48	1.7	1.3	2.0
49	2.1	1.9	1.6
50	2.3	2.7	2.7
51	3.3	2.8	3.6
52	3.0	4.9	2.0
53	2.5	2.4	3.5
54	1.5	1.8	1.0
55	2.0	1.2	2.2
56	1.9	1.6	1.5
57	1.8	1.9	2.1
58	4.4	3.6	3.0
59		2.6	4.3