DOCUMENT RESUME

ED 240 512 CS 007 499

AUTHOR Allen, JoBeth
TITLE Inference: A Research Review.
PUB DATE [80]
NOTE 31p.
PUB TYPE Information Analyses (07D)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Age Differences; *Cognitive Development; *Cognitive Processes; Literature Reviews; Reading Ability; *Reading Comprehension; *Reading Research; *Reading Skills; Research Methodology
IDENTIFIERS *Inferences; Theory Development

ABSTRACT

Since E.L. Thorndike's landmark 1917 study of the complexity of reading comprehension, inferential research has generally focused on either inference as a developmental process or the nature of inferences made during reading. In his 1930 study, R. W. Tyler established that inference could be objectively measured. S. G. Paris conducted several studies in the 1970s showing that readers' age and ability level influenced their inferencing skills, and indicating that young children are capable of producing inferences but usually do not do so spontaneously. Examinations of the nature of inferences have not been as conclusive as developmental studies. K. T. Goetz's 1977 study of high school students found no main effects for the importance or salience of the reading material. A 1979 study by P. D. Pearson, J. Hansen, and C. Gordon revealed that prior knowledge has a simple effect for inferable questions but no effect for explicitly stated questions. In a major contribution on the nature of inferencing, R. Tierney, C. A. Bridge, and M. Cera discovered that good readers made more causal and conditional connections between propositions while poor readers more often overgeneralized. One point of unanimity in reading research is that inferential ability is vital to mature readers. (MM)

# INFERENCE: A RESEARCH REVIEW

JoBeth Allen

Assistant Professor

Kansas State University

Inference: A Research Review

As interest in the topic of the unobservable part of the reading process, comprehension, has grown, several new areas of reading research have emerged. The present paper will examine one of these factors, inference, which was an almost unheard of topic before 1970. Calfee and Drum (1979) note that

> Comprehension is a complicated matter; it can be virtually synonomous with thinking. Trying to analyze the process of comprehension is an interesting challenge. We propose here two broad categories of comprehension tasks -- transliteral and inferential.

While many authors agree that there is such a skill as drawing inferences, agreement on a definition is more difficult. A basic definition, representing many writers who do not deal with hierarchies of inference, comes from Strange (1980).

> A question/answer relationship is textually implicit if both question and answer are derived from the text, but there is no expressed grammatical cue tying the questions to the answer (p. 395).

Frederiksen (1979) has developed much more specific and elaborate definitions about inference and types of inferences. His broad definition, however, adds specificity to Strange's concept.

> The fifth and highest level of processing operates entirely on propositions, generating new propositions that are given, for example, from prior discourse, from discourse context, or from previously acquired knowledge about the world. Any such processes will be referred to here as inferences.

In elaborating on the nature of inferences, Trabdsso (1980) has defined two ways in which inferences are made. The first involves text

connecting: finding relationships between events in the text. The second is slot-filling: adding missing information to enable text connecting. Trabosso reiterates the difficulty of studying inference reported by many researches; "...the making of an inference is a highly automatic and largely unconscious process...(but) is not a simple or obvious process" (p. 3). In spite of difficulties with observations, measurement and even definitions, inference has been the topic of much research in the past ten years. That research will be analyzed in this paper.

In 1917, Edward L. Thorndike conducted a landmark study in reading comprehension. The study was significant because it looked at the complexity of comprehension for the first time; up to that point reading had been considered to be "...a rather simple compounding of habits" (Thorndike, 1917, p. 425). Thorndike gave two hundred sixth-grade pupils a series of simple paragraphs and asked them to write the answers to questions about them. The questions included examples of main idea, factual recall and inference. Thorndike was impressed with the extensive variety of answers to his questions, and found that the responses did not fit neatly into categories of errors. Rather, he drew conclusions about the nature of "correct reading" based on the analysis of responses (Thorndike, 1917):

> In correct reading (1) each word produces a correct meaning, (2) each such element of meaning is given a correct weight in comparison with the others, and (3) the resulting ideas are examined and validated to make sure that they satisfy the mental set or adjustment or purpose for whose sake the reading was done (p. 428).

He went on to describe certain errors as failures to perform one of these operations. Thorndike concluded that correct reading was very closely tied to the reasoning process; thus, his title in reporting the

research is "Reading as Reasoning: A Study of Mistakes in Paragraph Reading."

Before leaving the Thorndike study, it is interesting to note that the three factors Thorndike describes might be viewed by present theorists as representative of the interactive nature of reading. Thorndike found that each word must produce a correct meaning (text factor), which the reader must weigh correctly in relation to other words (reader factor), and then validate according to the purpose for reading (process factor).

The reasoning factor identified by Thorndike was supported in several studies designed to analyze factors of comprehension. Feder (1938) factor analyzed the responses of 700 college students on "comprehension maturity" tests. He identified two distinct levels of response: reading for information and reading for inference. Comprehension factors were examined by F. B. Davis in two separate studies. The first study (Davis, 1944), conducted in 1941 with college freshmen, concluded that the two most important factors in reading were memory for word meanings and reasoning in reading (combining the categories of weaving ideas and drawing inferences). The second study was a much more carefully controlled and planned inspection of the reading process. In it, Davis (1968) administered a series of nine tests to 1,100 twelfth-grade students. He analyzed eight possible factors of comprehension to determine what percent of the unique variance each contributed. His categories and findings are summarized in Table 1. The percentages are averages from within-day and across-day matrices.

Table 1

DAVIS CATEGORIES AND AVERAGE PERCENTAGES OF COMPREHENSION FACTORS

| Categories | Percent Unique Variance |
|---|---|
| 1. recalling word meanings | 32 |
| 2. drawing inferences about word meanings | 3.5 |
| 3. factual recall | 10 |
| 4. weaving together ideas | 5 |
| 5. drawing inferences from context | 20.5 |
| 6. recognizing author's purpose, tone, etc. | 11 |
| 7. identifying author's technique | 5.5 |
| 8. following structure of passage | 13.5 |

Davis again interpreted his data to support word meanings and drawing inferences as the two factors accounting for over half of the unique variance.

Spearritt (1972) used Davis' 1968 study and applied the statistical analysis of maximum likelihood. He found support for the word knowledge factor, but felt that three other factors, drawing inferences from context, recognizing a writer's purpose, and following the structure of a passage, were so closely related that they should be considered as a single factor: reasoning.

Although these studies used categories identified as inference ability, the first study to focus exclusively on inference was conducted in 1930 by a professor at Ohio State University (Tyler, 1930). He was concerned with the objective measurement of all educational goals. His department had adopted the goal that each student should have "An ability to draw inferences

from facts, that is, to propose hypotheses." To measure students' ability to infer, Tyler tested sixty-six students in a beginning zoology course. The students took two tests in the same day, one requiring production of an inference, the second requiring recognition of an inference. Students read a series of factual statements; they were instructed to "Write... the one inference which seems to you most reasonable in attempting to interpret the facts" (Tyler, 1930, p. 476). A sample paragraph and logical inference were provided. Papers were collected, and the students were given the same items in multiple choice form. One might expect a rather large interference between the inference generated and the multiple-choice items. However, Tyler found a correlation of only .38. His interpretation of this data was that "the ability to select the most reasonable inference from a given list is not the same as the ability to propose an original inference" (p. 477-8). Tyler did not consider the possibility that inferences drawn in the first measure might have interferred with the ability of students to select a correct inference.

Tyler deleted the multiple choice format in favor of the generated inference, and developed a procedure for evaluating such inferences. He sorted inferences into five "equal-appearing intervals," which he had each member of the education faculty use for sorting samples of student inferences. He found three instructors to be as reliable as the whole department in determining "the quality of the original inferences...proposed by the student" (p. 479). He further validated the procedure with new faculty members, finding a correlation of their judgement with that of existing members' to be .96. In another analysis, Tyler found that his method of testing inference had a very low correlation (r=.29) with a test of factual

knowledge over the same material. He concluded that inference was not only a separate skill, but that it could be measured and "the merit of an inference is not the purely subjective quality often claimed" (p. 479).

Inference did not appear again in the literature as a separate concern until the seventies. The only mention of inferential ability was in studies such as Davis' (1944), where it was considered as part of the hierarchy of reading skills. Even these studies dealt with mature readers, rather than with children learning the basics of the reading process. As researchers began to look at young readers, they first focused on more observable behaviors such as word recognition for lists of words, nonsense words, syllables and even letters. Rate was easily measured, as were eye movements and other physical characteristics of the reader. It was not until the behavioral wave of research subsided that the complexities of comprehension were examined. Reading research was given added impetus from such fields as phycholinguistics, artificial language and psychology. The seventies produced intense discussion and theory about the nature of the reading process; the theories are presently being tested through research. The studies evolving from this atmosphere pertinent to inference are reported here.

Several interesting studies of children's inferential skills were conducted by Paris (Paris & Upton, 1974; Paris, 1975; Paris & Lindauer, 1976). Tierney and Cunningham (1980) summarize the first study by noting that "Paris (1975) found young children are less able to relate their own background experiences in the process of inferencing" (p. 46). This conclusion contrasts with the opinion of those who feel that young children do not have sufficient

background experiences to make the needed connection. Paris pursued this factor in his next study (Paris & Upton, 1974). Paragraphs were read to twelve subjects at each of six grade levels, kindergarten through fifth grade. Pupils were asked to respond to four literal and four inferential questions. The most interesting conclusion, apart from anticipated main affects for grade and question type ($p. < .01$), was that younger children seemed to be capable of drawing inferences. What separated them in ability from older children was that they did not do so spontaneously.

In order to explore this conclusion, Paris and Lindauer (1976) conducted an experiment using subjects from first, third and fifth grades. A total of forty-eight students read sentences where an instrument was either explicit (The man dug a hole with a shovel) or implicit (The man dug a hole.) They were then tested on the four implicit and four explicit sentences by being given the cue (which had been stated in only half of the sentences) and asked to recall the sentence. Analysis showed a significant difference between grades one and three, and one and five, but not between three and five, indicating that the development of the skill occurred in the primary grades. By fifth grade, students performed equally well on explicit and implicit instrument sentences. Paris and Lindauer performed a second experiment using sixteen each kindergarten, second and fourth graders. In this design, all sentences implied an instrument, but did not state it. Questions were asked using as recall cues explicit information (the stated subject, verb or object of the sentence) or implicit (the instrument). Again, younger children did not do as well on implicit-cued sentences as on explicit ones.

The third experiment reported by Paris and Lindauer attempted to remedy what they hypothesized as a lack of spontaneous production rather than an inability to deal with inferential material. The researchers had ten first graders act out the implicit sentences. The children were instructed to perform the actions for "The man dug a hole." The results were dramatic. Overall recall was sigificantly improved. More importantly, the difference between explicitly cued and implicitly cued sentences was eradicated, with subjects recalling 72% of the explicit-cue sentences and 70% of the implicit-cue sentences. Paris and Lindauer concluded that young children were capable of drawing inferences, but that they had no spontaneous strategy. This observation has been supported in other areas of learning, such as memory for lists (Smirnov & Zinchenko, 1969). It is unfortunate that Paris and Lindauer did not use a large enough sample to report their results in terms of significance; and that no one has sought to replicate this study using controls. The implications for teaching inference skills seem promising.

A number of studies have dealt with variations of the developmental process identified by Paris. Hildyard (1976) had sixteen first, third and fifth graders listen to passages, and sixteen third and fifth graders read passages silently. Measure were taken of implications (defined as explicit inferences requiring no world knowledge) and for inferences, both those related to world knowledge and contradictory to world knowledge. A "marked developmental difference" was found for implications and for those inferences that contradicted previous world knowledge. Reading facilitated implication; listening facilitated inference related to world knowledge.

Mandler and Johnson (1977), in their classic study of story structure and its effect on recall, examined the types of inferences children made in terms of their appropriateness. Their subjects were first and fourth grade students, and adults. The data in Table 8, Proportion of Subjects Producing Additions During Recall, Total Number of Additions, and Proportion Occuring in each Classification, is of particular interest. All subjects were approximately equal in the proportion of "reasonable elaborations" (1st - .36; 4th - .31; adult - .35). There were developmental trends in two other categories, however. Emphatic or redundant additions which were consistent with the original text increased with age (1st - .15; 4th - .41; adult - .49). Irrelevant or structural "fillers" which were not text-based decreased with age (1st - .40; 4th - .29; adult - .16). Mandler and Johnson interpreted these data in terms of young children's need to fill in missing story structure; it seems equally important to note that the skill of making appropriate inferences undergoes developmental changes in the years between first and fourth grades.

In an unpublished study done in 1977, Stein (1978) examined first and fifth graders' interpretations and recall of stories with very discrepant information (e.g., a mean, selfish fox helps a weak bear catch and cook a fish). When asked for exact recall, first graders recalled only the story line, not the discrepant description of the fox. Fifth graders recalled both the story and the description; some even made some inferences about the fox's motives. When asked to "fill in anything that is missing," first graders again gave fairly verbatim recall, with no mention of incongruities. Fifth graders made inferences as to the reasons for the incongruous actions,

and sought to provide resolution to the story. While her report contained no statistical data, these findings have been replicated in other studies (Stein & Glenn, 1975).

In one such study, Nezworski, Stein, and Trabasso (1979) looked for differences between kindergarten and third-grade students, testing a total of 144 children. Their material was written in two forms; the control passage presented a negative protagonist, while the added information passage gave an explanation of why the protanogist acted as reported, making that person less negative. Results of the questioning of each type of passage are reported in Table 2.

Table 2

PROPORTION OF CORRECT INFERENCES MADE TO PROBE QUESTIONS

| Inference type | Control | | Added information | |
|---|---|---|---|---|
| | K | 3 | K | 3 |
| causal | .36 | .38 | .64 | .87(*) |
| consequence | .33 | .33 | .82 | .96(*) |
| reaction | .89 | .97(*) | .88 | .97(*) |

(*) significant at the .05 level

It appears that the low proportion of correct responses in the causal and consequence categories on the control passage indicates a uniform difficulty in dealing with ambiguity for first and third graders. However, the older students were better able to make use of added information in resolving ambiguity.

A developmental difference at an even younger level was reported by Poulsen, Kintsch, Kintsch, and Premack (1979). They found that four and

six year old children can both make up stories from pictures presented in logical order. However, when the pictures were presented in random order, the younger children reverted to simply labeling each picture. The older children tried to make sense of the pictures as presented, and made inferences about thoughts and emotions in order to tie a story line together. This finding is consistent with Stein's 1977 study in its developmental difference, but a perplexing discrepency exists. The "older" students in the Poulsen, et al. study were six; the younger children in Stein's study were first grade, probably six or seven years old. The age that Stein reported as unable to resolve incongruity showed itself to be quite able to do so in the picture-story telling task presented by Poulsen et al. One possible explanation is that in the Stein study, the story was read to subjects; therefore, they did not have the text for reference. Subjects in the Poulsen et al. investigation had concrete representation of the incongruity, and could ponder and process with more ease. Also, the required task would tend to facilitate inference in the story-telling condition. Telling a story, for a child who understands the construct of a story (which both four and six year old children in the study did), requires the logical connection of ideas and events. In the case of the incongruous pictures (those in scrambled order), inferences were required to create a cohesive story. In contrast, the Stein task was only loosely focused on resolving ambiguity, with instructions to fill in missing information. It is quite possible that with specific inference probes, these younger children could have produced logical inferences. The work of Paris and Lindauer (1976) supports this hypothesis.

Another major category of inferential research, besides its developmental nature, has to do with the nature of inferences made. Much of the current examination of inference is based on schema theory proposed by reading theorists (Anderson, 1977; Rumelhart & Ortony, 1977; Spiro, 1977). Various systems for analyzing inferences have been devised, such as Collins, Brown and Larkin's (1977) Progressive Refinement Model, which "entails examining inference holistically, within the framework of the reader's construction, evaluation, and perhaps revision of a scenario for the text" (Tierney, Vaughn, & Bridge, 1979). Frederiksen's Taxonomy of Text-Based Inferences has been used in several research designs as a system for categorizing types of inferences (Frederiksen, 1977; Frederiksen, Frederiksen, Humphrey, & Ottesen, 1978). An excellent, highly usable taxonomy was proposed by Warren, Nicholas and Trabasso (1979). The following group of studies adds to these taxonomies empirical data on the nature of inference.

Goetz (1977) conducted a study of 184 subjects divided into a ninth/tenth grade group and an eleventh/twelfth grade group. Groups were further separated into high and low verbal ability sections. The material read was of two types: an important version and an unimportant version. One story, for example, presented the reader with the information that a woman was late to the airport because she stopped to talk to her son. In the unimportant version, the narrative describes the outcome of this diversion to be less time spent in the airport. In the important version, the talk causes her to miss her flight, which crashes; thus the time with her son saved her life. Other text variables included salience and

explicitness, described by Goetz in the following passage:

> Salience of the premise was varied by changing the amount of detail or emphasis given to the material that cued the specific target inference. Explicitness of the target was varied in order to provide a control condition in which the target had been explicitly stated (p. 45).

There were six sets of eight passages with various combinations of importance, salience and explicitness, each from 500 to 520 words in length. The experimenter carefully matched these fiction passages, and arranged them to avoid an order affect.

Subjects were asked five questions about each passage. The first was always an inference question, the last an "importance" question. Interspersed were basic premise questions and questions used to control for general memory. Subjects also rated each question as 1) an exact quote, 2) paraphrase, 3) directly implied, or 4) consistent with story. An interesting design strategy had the students perform a word recognition test after reading the passage and before answering the questions; this test served as a buffer for short-term memory as well as a control for readability of the passages (albeit a limited control).

Main effects were reported for grade level and verbal ability. There was no main effect for importance; however, there was an interaction between importance and explicitly stated items. Goetz reports this interaction:

> Simple effects analysis of this interaction revealed that the important passages produced significantly higher scores than the unimportant passages for the implied condition (Important = .81, Unimportant = .74) $p < .01$. Thus the importance of the target inference had a significant effect on the probability that the inference would be made.

Salience was also involved in an interaction; low verbal ability students

inferred significantly more highly salient passages. Students seemed aware that they were making inferences, since they rated significantly more explicitly stated propositions as having been stated. In summary, it appears that inferences are more likely to be made when it is important that they be made.

A study designed to examine when and how inferences are made was conducted by Spiro and Esposito (1977). They required twenty educational psychology majors to read 250 word vignettes and answer questions. In the passages, A "pragmatically implied" B, and variable C lessened the extent of that implied relationship. One group of students read the passages with C coming after the A-B relationship; the other group read C before A-B. The control group had no C statement. Seven questions were asked for each passage, with the following distribution:

1) asked if B had been mentioned

2) asked a yes/no question of B

3)-6) other questions about the passage, designed to mask the intent of the study (not analyzed)

7) C information.

Subjects further rated the certainty of their responses on a six point scale.

There were significantly ($p < .015$) more errors when C was read after A-B than when C occurred before. In the C-after condition, subjects often reported that the inference was not in the text (B-mention errors) or that the opposite was true (B-incorrect errors). Spiro and Esposito concluded that "predictable information is superficially processed and not stably

represented in memory." While it appears likely from this experiment that readers are more likely to make errors after the inference has been made, then contradicted, design limitations lessen the generalizability of the findings. The authors report sample sizes of only four students in each of the experimental groups, and eight in the control group. While they report that group differences accounted for 23.7% of the reported variance, they did not specify whether this analysis was of the experimental versus control, between the two experimentals, or among all three. Four remains an almost inconsequential sample size. Perhaps an even more damanging problem surfaces in examining the instructions given the students and the time allowed for reading. Subjects were instructed to read and understand each passage, and asked not to look back. However, they were given nine minutes for each 250 word passage. Some must have filled the eight minutes with rereading, if only out of boredom. Others might have processed the incongruity of the C-after condition, and reread for resolution.

Schreiner and Shannon (1980) examined "The Recall of Explicit and Implied Propositions by Good and Poor Readers Using Three Types of Assessment Procedures." They defined good readers as eighth grade students reading at a mean grade equivalent of 10.0, based on the California Achievement Test; poor readers were from the same grade reading at a mean grade equivalent of 5.0. The subjects were randomly assigned to code groups FR (free recall), AR (aided recall) or MC (multiple choice). Three "natural expository passages" were read; they were reported as having readabilities of 3.0, 4.0 and 5.0 on the Fry graph (although the Fry does not measure to tenths).

Main effects were found for ability, explicitness and question type. Good readers performed better than poor readers, with means of 2.85 and 1.45 respectively. Explicit propositions were recalled a mean number of times 2.47; the mean for implicit items was 1.59. Means for multiple choice, aided recall and free recall were 2.32, 1.40 and .80, respectively. The authors reported surprise that even the high group recalled only 52% of the total propositions; however, this finding is consistent with other research (Tierney, Bridge, & Cera, 1978-79).

A major limitation of the study is the classification of good and poor readers, and subsequent testing based on that classification. It was not reported whether the scores from the CAT were overall reading score, vocabulary, or comprehension section. If comprehension was a large part of the determinant score, then the poor performance on comprehension questions was to be expected. However, a matter for greater concern is the readability level of the passages used. All passages were at an independent level for the good readers. The poor readers, on the other hand, may have been reading at instructional and even frustration levels, since their <u>average</u> reading ability was 5.0. This failure to match reading ability with passage difficulty is a crucial limitation.

The role of prior knowledge on a reader's ability to make inferences was the focus of a study conducted by Pearson, Hansen, and Gordon (1979). Their population was "slightly above average" second graders, differentiated by strong and weak schemata for knowledge about spiders. Recall was measured by "wh-" questions, which were judged to be more ecologically valid than free recall. The two groups did not differ significantly on

average reading ability scores, as measured by the Metropolitan Achievement Test, or on IQ (weak Schema mean IQ = 114.8, strong group = 120.4). The subjects were placed in schema groups based on their scores on an eight-question test of knowledge about spiders. Students answering five, six, or seven (mean = 5.8) were designated as having strong schemata. Students answering four questions were dropped from the study; the remaining groups were significantly different at the .001 level.

To these well-defined groups, Pearson et al. presented a basal selection (Lyons and Carnahan, 1972) rewritten to include extra information about spiders, and a narrative line. The story had a readability level of 2.8 on the Spache Readability Formula. Students read the passage once, silently, and could not refer to it during questioning. Examiners did not help with unknown words. Explicit and inferential knowledge were measured during questioning.

Results indicated a simple effect for prior knowledge on inferrable questions ($p < .025$). There was not a significant effect for prior knowledge on explicitly stated questions. Main effects for strength of prior knowledge and question type, both significant at the .01 level, are reported in Table 3.

Table 3

EFFECTS OF PRIOR KNOWLEDGE AND QUESTION TYPE

| | explicit questions | implicit questions | total |
|---|---|---|---|
| strong schema | 4.7 | 2.8 | 7.5 |
| weak schema | 3.8 | 1.0 | 4.8 |
| average | 4.25 | 1.9 | |

In their discussion of the results, Pearson et al. note that the findings are not consistent with the prediction that prior knowledge would affect both inferential and literal recall. However, it seems that the lack of significant differences on the explicit material is actually a validation of the reading ability/intelligence control so carefully built into the design. The study showed that both groups learned stated facts equally well. The more difficult process of combining facts with prior knowledge (inferencing) did prove to be dependent on level of prior knowledge.

One of the most widely referred to studies of the nature of inference was performed by Connie Bridge as dissertation research, and reported by Tierney, Bridge, and Cera (1978-79). The study was concerned with the extent and type of inferential processing of written, non-narrative discourse. Subjects were thirty-six third grade readers, divided into groups of good and poor readers on the basis of Stanford Achievement Test scores and teacher judgement. Children read orally a passage about dinosaurs, read a buffer passage to control for short-term memory, and responded to free and probed recall. Probe questions were based on free recall responses, and avoided leading readers to insights they had not generated. Frederiksen's system of text analysis for Text-Based Inferences was used to categorize responses; his eight classes of inferential operations were the basis for comparison of types of inference between the two groups.

Major findings included the well-documented point that good readers answer significantly more questions than poor readers, whether the questions are analyzed as a whole or for explicit and inferential types separately.

Good readers tend to do everything in the reading task better than poor readers; that is why their SAT scores and teachers' opinions single them out as good readers. There was no significant difference for good or poor readers in the proportion of inferred to explicit recalls. (There is a discrepency in the reporting of this result. In the abstract of Bridge's dissertation, it was stated that approximately 45% of the recalls were inference. In the Tierney, et al. report, the proportion of inferred to explicit recalls was statistically reported as .51 for all readers.) The third major finding was that during free recall, both groups recalled about twice as many explicit propositions as inferred. In the probed condition, more inferences than literal recalls were recorded.

Possibly the most important contribution this study makes is about the nature of inferences made by good and poor readers. The discussion section explains:

> ...good readers generated more causal and conditional connections between propositions, while poor readers tended to substitute more general concepts for the specific terms used in the passage. The good readers' use of causal and conditional inferences seemed to add to the organization and cohesiveness of their recalls. The poor readers' use of superordinated concepts seems to result in a loss of accuracy and specificity (p. 564).

This finding indicates the need for more specific criteria for evaluating inferences. The Warren et al. system is more accurate in assigning inferences on the basis of correct inferences, as well as on their basis in text. Warren et al. (1979) point out the necessity of applying a "relevancy criterion" to certain categories.

There are several problems affecting the generalizability of the study. There was no matching of groups, nor reporting of group means

for such relevant data as intelligence or verbal ability. The standardized test was not reported; groups were assigned according to percentiles (good = above sixty-eighth percentile, poor = below thirty-eighth percentile; these are not equivalent). It was reported that the experimenters dropped those who scored below 84% word recognition on the experimental material. That is a terribly low word recognition criterion; it is most likely that many of the poor readers were reading at a frustration level. Did their necessary attention to word recognition and analysis interfere significantly with their ability to comprehend? Did the fact that the passage was read orally interfere with the comprehension of either group?

In discussing the findings, the authors hypothesized that "maybe with probing the good readers were willing to extend their recalls beyond the explicit." It also seems likely that students, conditioned by school, were tuned into the "right answer syndrome" which emphasizes the recall of literal information (Durkin, 1978-79). There may be another factor influencing the reversed proportion of explicit/inferred recalls in the probed condition. It is likely that the questioners asked questions posed to elicit inferences, since that was the focus of the study. It is also logical to assume that the good readers, especially, had already recalled most of the explicit information they were capable of remembering; inference was the next level of processing (Taba, 1965).

In 1979, Tierney and Bridge performed further analysis on Bridge's data. They examined Frederiksen's Text-Based Inference categories, and found no statistical differences across categories. They then reexamined the responses of nine poor and nine good readers according to five additional categories. These included three of Frederiksen's Functional

Contexts of Inference: generalization, deletion and integration. Also included were vanDijk's macro-rules for extension and connection. No significant differences were found for generalization, extension or connection. Results were significant for the finding that poor readers deleted more propositions than good readers ($p < .05$), and that good readers integrated more propositions than poor readers ($p < .01$). Tierney and Bridge conclude that "the ability to integrate propositions is essentially what differentiates 'control' of inferred processes by good and poor readers" (p. 32). While this seems to be an oversimplification, even an overgeneralization from the data, it is certainly a variable worth investigating. Further research should differentiate between correct and incorrect (or appropriate and inappropriate) inference categories to simplify interpretation of data.

A final category of investigation concerning inference is the more pragmatic research which asks how inferential ability can be improved. This research will be presented briefly, since it is peripherally related but not the central focus of this paper. Nardelli (1957) failed to find significant gains in sixth grade pupils who participated in a short unit of instruction on drawing inferences. Wong (1980) found that learning-disabled students were better able to draw inferences when prompted recalls were used; however, no gain in actual ability was measured. Tierney and Cunningham (1980) report that an unpublished doctoral dissertation (Gordon, 1979) compared three treatments on increasing inferencing. Treatment one sought to build prior knowledge and awareness of text structure; the second taught inferencing strategies directly; the third treatment was actually a control utilizing a "language-related"

curriculum. Better transfer was reported for direct instruction, although this was not the result hypothesized. Although their results with college students may not generalize, Steingart and Glock (1979) found imagery to have a more facilitating effect than mere repetition for making inferences. Second graders studied by Hansen and Pearson (1980) made more inferences, and the learning transferred better, when they were part of a group where inferential questions were asked regularly. This group was superior to a "strategy" group that was trained in the process of inferencing from their own experiences.

This paper has been an attempt to analyze two main branches of inferential research, inference as a developmental process, and the nature of inferences made during reading. Inference has been identified as a unique component of comprehension (Davis, 1944 & 1968; Spearritt, 1972). Tyler (1930) established that inference was a skill that could be objectively measured. When several different ages and ability levels of readers were examined, differing patterns of inference emerged. Paris conducted several experiments showing that young children are capable of producing inferences, but that they usually do not do so spontaneously (Paris & Upton, 1974; Paris, 1975; Paris & Lindauer, 1976).

While almost every study of inference has shown that older children spontaneously produce more inferences (as do better readers), and that most readers give more literal recalls than inferential ones, other significant differences have been noted. Many of these differences deal with type or quality of inference as a developmental variable. Hildyard (1976) found that there was a "marked developmental difference" in the ability to draw inferences that contradicted previous world knowledge.

Younger children produced more irrelevant inferences, while older readers tended to have more appropriate additions in Mandler and Johnson's 1977 study. Third graders proved to make better use of added information in drawing inferences to solve ambiguity than did kindergarteners (Nezworski, et al.). This younger age group, however, proved capable of producing inferences to solve ambiguity when presented with scrambled-order pictures, a task that proved too difficult for four year old subjects (Poulsen, et al.).

Examinations of the nature of inferences made by readers of connected discourse have not been as conclusive as the developmental studies. With the hypotheses that importance, salience and explicitness would affect the reader, Goetz (1977) found no main effects for importance or salience. It did appear, from reported interactions, that inferences are more likely to be made when they are important to the story; salience was reported to enhance the inferencing of students with low verbal ability. Inference appeared to be superficially processed, without stability in memory, in a study by Spiro and Esposito (1977). Prior knowledge seems to have a facilitating affect on inferential processing (Pearson, et al.). In a major contribution on the nature of inferencing, Tierney, et al. (1978-79) discovered that good readers made more causal and conditional connections between propositions; poor readers more often overgeneralized, Tierney and Bridge (1979) later reported that the key to good reading was the ability to integrate propositions.

Here and there, in bits and pieces, inference is being established as a measurable, definable skill. Perhaps the only point of unanimity is that inferential ability is vital in the mature reader. Those who are devoting their energies to producing mature readers have accepted the

challenge presented by Thorndike over half a century ago: "It is not a small or unworthy task to learn 'what the book says'" (Thorndike, 1917, p. 434).

## References


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

Calfee, R. C. & Drum, P. A. How the researcher can help the reading teacher with classroom assessment. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol. 1). Hillsdale, N.J.: Erlbaum, 1979.

Collins, A., Brown, J. S., & Larkin, K. M. Inference in text understanding (Tech. Rep. No. 40). Urbana, Ill.: University of Illinois, Center for the Study of Reading, 1977.

Davis, F. B. Fundamental factors of comprehension in reading. Psychometrika, 1944, 9, 185-197.

Davis, F. B. Research in comprehension in reading. Reading Research Quarterly, Summer, 1968, 3, 499-545.

Durkin, D. What classroom observations reveal about reading comprehension instruction. Reading Research Quarterly, 1978-79, XIV, 481-533.

Feder, D. D. Comprehension maturity tests - a new technique in mental measurements. Journal of Educational Psychology, November 1938, XXIX, 597-606.

Frederiksen, C. H. Discourse comprehension and early reading. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol. 1). Hillsdale, N.J.: Erlbaum, 1979.

Frederiksen, C. H. Inference and the structure of children's discourse. Quoted in J. C. Harste & R. F. Carey (Eds.) New perspectives on comprehension. 1979 (3, US ISSN 0193-4759).

Frederiksen, C. H., Frederiksen, J. D., Humphrey, F. M., & Ottesen, J. Discourse inference: Adapting to the inferential demands of school texts. Quoted in J. C. Harste & R. F. Carey (Eds.) New perspectives on comprehension. 1979 (3, US ISSN 0193-4759).

Goetz, E. T. Inferences in the comprehension of and memory for text. Urbana, Ill.: University of Illinois, Center for the Study of Reading, 1977. (ERIC Document and Reproduction Service No. ED 150 548).

Gordon, C. J. The effects of instruction in metacognition and inferencing on children's comprehension abilities. Unpublished doctoral dissertation, University of Minnesota, 1979. Quoted in R. J. Tierney & J. W. Cunningham. Research on teaching reading comprehension (Tech. Rep. No. 187). Urbana, Ill.: University of Illinois, Center for the Study of Reading, November 1980.

Hansen, J., & Pearson, P. D. The effects of inference training and practice on young children's comprehension, April 1980. (ERIC Document Reproduction Service No. ED 186 839).

Hildyard, A. Children's abilities to produce inferences from written and oral material. Unpublished doctoral dissertation, University of Toronto, 1976.

Mandler, J. M., & Johnson, N. S. Remembrance of things parsed; Story structure and recall. Cognitive Psychology,, 1977, 9, 111-151.

Nardelli, R. R. Some aspects of creative reading. Journal of Educational Research, March 1957, L, 495-508.

Nezworski, T., Stein, N. L., & Trabasso, T. Story structure versus content effects on children's recall and evaluative inferences, June 1979. (ERIC Document Reproduction Service No. ED 172 187).

Paris, S. G. Integration and inference in children's comprehension and memory. In F. Restle et al. (Eds.) Cognitive Theory (Vol. 1). New York: John Wiley and Sons, 1975.

Paris, S. G., & Lindauer, B. K. The role of inference in children's comprehension and memory for sentences. Cognitive psychology, 1976, 8, 217-227.

Paris, S. G., & Upton, L. B. Children's comprehension of implicit and explicit information in paragraphs, 1974. (ERIC Document Reproduction Service No. ED 099 104)

Pearson, P. D., Hansen, J. & Gordon, C. The effect of background knowledge on young children's comprehension of explicit and implicit information, March 1979. (ERIC Document Reproduction Service No. ED 169 521)

Poulsen, D., Kintsch, E., Kintsch, W., & Premack, D. Children's comprehension and memory for stories. Journal of Experimental Child Psychology, December 1979, 28, 379-403.

Rumelhart, D. E., & Ortony, A. The representation of knowledge in memory. In R. C. Anderson et al. (Eds), Schooling and the acquisition of knowledge. Hillsdale, N. J.: Erlbaum, 1977.

Schreiner, R., & Shannon, P. The recall of explicit and implied propositions by good and bad readers using three types of assessment procedures. Paper presented at the Annual IRA Convention, St. Louis, May 1980.

Smirnov, A. A., & Zinchenko, P. I. Problems in the psychology of memory. In M. Cole & I. Maltzman (Eds.), A handbook of contemporary Soviet psychology. New York: Basic Books, 1969.

Spearritt, D. Identification of subskills of reading comprehension by maximum likelihood factor analysis. Reading Research Quarterly, 1972, VIII, #1, 92-111.

Spiro, R. J. Remembering information from text: Theoretical and empirical issues concerning the state of schema reconstruction hypothesis. In R. C. Anderson et al. (Eds), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Erlbaum, 1977.

Spiro, R. J., & Esposito, J. J. Superficial processing of explicit inferences in text, December 1977. (ERIC Document Reproduction Service No. ED 150 545).

Stein, N. W. How children understand stories: A developmental analysis, March 1978. (ERIC Document Reproduction Service No. 153 205).

Stein, N. L., & Glenn, C. G. A developmental study of children's recall of story material. Paper presented to Biennial Meeting for the Society for Research in Child Development, Denver, April 1975.

Steingart, S. K., & Glock, M. D. Imagery and the recall of connected discourse. Reading Research Quarterly, 1979, XV, #1, 66-82.

Stronge, M. Instructional implications of a conceptual theory of reading comprehension. The Reading Teacher, January 1980, 33, #4, 391-397.

Thorndike, E. L. Reading as reasoning: A study of mistakes in paragraph reading. The Journal of Educational Psychology, 1917, VIII, #6, 323-332.

Tierney, R., & Bridge, C. A. The functions of inferences: an extended examination of discourse comprehension. In M. T. Kamil & A. J. Moe (Eds.), Reading research: Studies and applications. Clemson, S.C.: Reading Conference, 1979, 129-133.

Tierney, R., Bridge, C. A., & Cera, M. The discourse processing operations of children. Reading Research Quarterly, 1978-79, XIV, #4, 539-573.

Tierney, R., & Cunningham, J. W. Research on teaching comprehension, November 1980. (ERIC Document Reproduction Service No. ED 195 946).

Tierney, R., Vaughn, J. L., & Bridge, C. A. Toward Understanding comprehension: An examination of systems for analyzing inferences. In J. C. Horste & R. F. Carey (Eds.) New perspectives on comprehension. Monograph in Teaching and Learning, 1979 (3, US ISSN 0193-4759)

Trabosso, T. On the making of inferences during reading and their assessment, January 1980. (ERIC Document Reproduction Service No. ED 181 429).

Tyler, R. W. Measuring the ability to infer. Educational Research Bulletin, November 1930, IX, 475-80.

Warren, W., Nicholas, D., & Trabasso, T. Event chains and inferences in understanding narratives. In R. Freedle (Ed.) New directions in discourse processing. Norwood, N.J.: Ablex, 1979.

Wong, B. Y. L. Activating the inactive learner: Use of questions / prompts to enhance comprehension and retention of implied information in learning disabled children. Learning Disabilities Quarterly, Winter 1980, 3, #1, 29-37.