DOCUMENT RESUME

ED 291 062 CS 009 021

AUTHOR Marzano, Robert J.
TITLE A Study of Inference in Standardized Reading Test Items and Its Relationship to Difficulty.
INSTITUTION Mid-Continent Regional Educational Lab., Aurora, CO.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Apr 87
CONTRACT 400-86-0002
NOTE 22p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Deep Structure; *Inferences; *Language Processing; Learning Strategies; Linguistic Theory; Objective Tests; Reading Comprehension; *Reading Tests; *Standardized Tests; *Test Items
IDENTIFIERS *Metropolitan Achievement Tests; Test Analysis

ABSTRACT

To study the relationship between inferences made on standardized reading tests and item difficulty, 50 items on the reading comprehension section of the Metropolitan Achievement Test were analyzed independently in this study by two raters using four general categories of inferences: (1) reference inferences, (2) between proposition inferences, (3) source inferences and (4) metalinguistic inferences. The items were in standard reading comprehension test format (reading passages followed by multiple choice items based on the passage). An inference was operationally defined as the mental process of inducing or deducing, as cued by the test item, information not explicitly stated in the test. No significant relationship was found between inference types and item difficulty; however, a significant relationship was found between a general measure of raw amount of information processed and item difficulty. Although most reading skill hierarchies assume that inferential cognitive operations are inherently more difficult than non-inferential operations, these findings suggest that either inferential cognitive operations are not inherently more difficult than non-inferential cognitive operations, or that the inferential cognitive operations have been internalized at the level of automaticity for school-aged test takers. (Twenty references are attached and two tables of data are included.) (NH)

A STUDY OF INFERENCE IN STANDARDIZED READING TEST ITEMS AND ITS RELATIONSHIP TO ITEM DIFFICULTY

by

Robert J. Marzano

Mid-continent Regional Educational Laboratory

April, 1987

This publication is based on work sponsored wholly, or in part, by the office of Education Research and Improvement, Department of Education, Under Contract #400-86-0002. The publication's content does not necessarily reflect the views of OERI, the Department or any other agency of the U.S. Government.

Abstract

Most reading skill hierarchies assume that inferential cognitive operations are inherently more difficult than non-inferential operations, yet there has been little empirical study of the relationship between inference types and item difficulty on standardized reading test items. In this study 50 items from the reading comprehension section of the Metropolitan Achievement Test were analyzed for four general categories of inference. Item difficulty (p values) was then regressed on inference types. No significant relationship was found between inference types and item difficulty, however, a significant relationship was found between a general measure of raw amount of information processed and item difficulty. These findings suggest that either inferential cognitive operations are not inherently more difficult than non-inferential cognitive operations or that the inferential cognitive operations in this study have been internalized at the level of automaticity for school aged test takers.

Although there has been a considerable amount of study of the different types of inferences made while reading different forms of written discourse (Crothers, 1979; Kintsch, 1979; Warren, Nickolas and Trabasso, 1979) there has been little research done on the types of inferences made on standardized reading tests and their relationship to item difficulty.

In a study of surface level linguistic features and their relationship to item difficulty on standardized reading tests, Drum, Calfee and Cook (1980) found that such surface structure elements as word length, propositional density and syntactic density were significantly related to item difficulty and accounted for as much as three fourths of the variance. Commonly surface level linguistic measures are associated with non-inferential cognitive operations used to create a micro-structure representation of information explicitly stated (Kintsch, 1979). Under this interpretation of surface level linguistic features, the Drum, Calfee and Cook findings would seem to imply that inferential cognitive operations are not strongly related to item difficulty in reading test items. This conclusion is supported by DiStefano and Valencia (in press) who found no significant relationship between the type of reading question (literal versus inferential) and item difficulty on items administered by the National Assessment of Educational Progress. However, DiStefano and Valencia did not study different types of inferences and included very few

inference questions among their sample of items. Hence, the small sample of inferential items and the collapsing of all inference types into one category could have masked a relationship between certain inference types and item difficulty.

The apparent lack of relationship between inferential cognitive operations and item difficulty is not consistent with most hierarchies of cognition, especially those that deal with the processing of linguistic information (Rosenshine, 1980). That is, since most models of the reading process either implicitly or explicitly assert that the process of answering inferential questions is more difficult than the process of answering literal questions one would assume that the number and type of inferences required on a reading test item would be a stronger predictor of an item's difficulty than the number and type of non-inferential cognitive operations. This was the basic finding of Hillocks and Ludlow (1984) in their study of student responses to questions (both literal and inferential) from relatively long narrative and expository passages. In fact, Hillocks and Ludlow found that different inference types could be arranged in a hierarchic fashion relative to difficulty.

Given the discrepant findings of the Hillocks and Ludlow versus the Drum, Calfee and Cook and the DiStefano and Valencia studies, a question as yet unanswered is "to what extent are specific inference types related to item

difficulty on standardized reading tests?" The purpose of this study was to answer that question. More specifically, this study sought to answer the research question: "What is the relationship between the type of inferences in standardized test items and item difficulty?"

METHOD

Fifty items from the reading section of the Metropolitan Achievement Tests, Intermediate Level, Form JS (Prescott, Balow, Hogan and Farr, 1978) were analyzed for four general categories of inference each with subcategories. The items were in standard reading comprehension test format (reading passages followed by multiple choice items based on the passage).

An inference was operationally defined as the mental process of inducing or deducing, as cued by the test item, information not explicitly stated in the text. The four general categories of inferences studied were: 1) reference inferences, 2) between proposition inferences, 3) source inferences and 4) metalinguistic inferences.

Inference Categories

Reference inferences are those in which a reader must infer that a word, phrase, or a syntactic cue in an item refers to a specific word, proposition or set of

propositions in the reading passage accompanying the item. For example, assume a reading test item were written in the following way:

The young girl in the story was late for:

a) lunch

b) school

c) a tea party

d) baseball practice

If the passage to which the item applied did not use the term "young girl" but referred to her by name (e.g., Jana) and forms of the third person pronoun (e.g., she, her), the reader would have to infer that "young girl" referred to Jana.

There are a number of models for and ways of describing the different types of reference inferences that can be made (Halliday and Hasan, 1976; Meyer, 1975; Turner and Greene, 1977). In this study four subcategories of reference inferences were analyzed:

1) reference by syntax (The syntactic structure of an item signals information in the text.)
2) reference by synonym (A synonym is used in the item for a term in the text.)
3) reference by general term (A superordinate term is used in the item for a subordinate term in the text.)

4) reference by specific term (A subordinate term is used in the item for a superordinate term in the text.)

Pronomial reference, perhaps the most common form of reference in oral and written discourse, was not included in the analysis because virtually every occurrence of a pronoun used in a test item was accompanied by a pronoun in the text. Consequently the reader was not required to make an inference in the item to get back to the text because the relationship between the pronoun and its antecedent would have already been established as a result of reading the text.

Between proposition inferences occur when the reader must infer a relationship between propositions which is not explicitly stated in the text. For example, assume that an item were written in the following way:

The young girl was late for baseball practice because:

a) she had to finish her paper route
b) she had to stay after school
c) she was not feeling well
d) she didn't feel like practicing that day.

Here the reader must make a connection between the proposition "the young girl was late for school" and one of

the four propositions listed in the alternatives. The connection the reader must make is one of causality. Presumably, in the text to which this item refers, one of the four alternative propositions was stated as a cause for the proposition in the stem. Meyer (1975) refers to such relationships between propositions as rhetorical predicates. Halliday and Hasan (1976) refer to them as conjunctives. If there were no explicit linguistic signal in the text (e.g., use of the conjunction <u>because</u> between the two propositions) the reader would have to infer that the proposition in the stem and the correct alternative did, in fact, have a causal relationship.

Again there are a number of ways to describe the different types of relationships and, consequently, inferences that can be made between propositions. Marzano, Hagerty, Valencia and DiStefano (1987) have determined that between proposition relationships commonly described in most propositionally based systems of language analysis can be classified into four major categories: causal relationships, additive relationships, comparative relationships and temporal relationships. These can exist between propositions explicitly stated in the text and between those not stated in the test. In other words, two propositions may have a relationship and both are explicitly stated in text or only one is stated in the text. These two characteristics (type of relationship between propositions and implicit or explicit presence in the text) were

collapsed in this study to create two general categories of between proposition inferences:

- inferences requiring the reader to identify a causal, additive, temporal or comparative relationship between two propositions explicitly stated in the text which do not have an explicit linguistic marker signaling the relationship.

- inferences requiring the reader to identify a causal, additive, temporal or comparative relationship between two proposition one of which is not explicitly stated in the text.

Source inferences are those in which the reader must infer some characteristic about the author or the intention of the author from reading the text. For example, an item which referred to information the author must have known or could not have know would require an inference about "source." Inferences about source include aspects of "theme" as described by Halliday (1967) and "staging" as described by Grimes (1972).

Metalinguistic inferences are those which require the reader to know some specific characteristics and conventions of written discourse. For example, an item which assumes the reader knows that a story will generally include a

setting, an initiating event, a climax and a conclusion requires metalinguistic inferences. De Beaugrande (1980) has identified eight types of metalinguistic structures common to written discourse. These are: descriptive, argumentative, literary, poetic, scientific, didactic and conversational. Van Dijk (1980) has identified four types of metalinguistic structures all of which are covered by de Beaugrande's categories. Within the present study, inferences about any of the above structures were coded as metalinguistic.

Analysis of Items

The fifty items on the reading comprehension section of the Metropolitan Achievement Tests were analyzed independently by two raters using the inference categories described above. Each item was scored in a dichotomous fashion (presence of inference type versus lack of presence of inference type) for each of the four general categories of inference. Inter-rater reliabilities on the initial analysis ranged from .82 for source inferences to .96 for reference inferences as measured by Pearson product moment correlations. Although these results indicated substantial agreement, all disagreements were submitted to a third rater. The third rater's agreement with one of the primary raters was accepted as the correct coding.

Analysis of Data

Item difficulty (item p values) was regressed on the four general inference types using a stepwise multiple regression analysis. Table 1 reports the means and standard deviations for each of the five variables in the equation.

Table 1 here

Table 2 contains the regression coefficients and F values for the variables in the equation.

Table 2 here

RESULTS

Table 1 indicates that the fifty items contained many inferences of all types. Reference inferences were the most frequent, occurring in about nine out of ten items. Source inferences were the least frequent. As Table 2 indicates none of the inference types were significant (.05 level) predictors of item difficulty within the multiple regression equation. The multiple R for the equation was .33 and had a probability of .55.

The lack of significance of the multiple R was interpreted as an indication that the overall amount of inferences made on a reading test item is not a significant

TABLE 1

Means and standard deviations for variables in the equation.

Variable	N	M	SD
Item difficulty (p value)	50	57.20	15.61
Reference	50	.90	.61
Between Proposition	50	.43	.63
Source	50	.07	.25
Metalinguistic	50	.10	.40

TABLE 2

Regression coefficients, F values and significance levels.

Variable	B	Beta	F	Sig
Reference	.962	.038	.034	.855
Source	-.636	-.010	.003	.958
Metalinguistic	-11.189	-.289	2.17	.153
Between proposition	-5.027	-.202	.92	.346
Constant	59.673		74.602	.000

predictor of item difficulty. The lack of significance of any predictor variable in the equation was interpreted as an indication that no single type of inference included in this study is significantly related to item difficulty. In other words, the findings of this study imply that the number and type of inferences made on reading comprehension test items have little relationship to the difficulty of items.

To test whether the difficulty of the items was a function of non-inferential rather than inferential cognitive operations, two other predictor variables were entered into the equation: 1) passage length, and 2) depth of answer. Passage length is a commonly used to measure the amount of raw information and surface complexity, both syntactic and semantic, of written and oral information (Lee and Canter, 1971; O'Hare, 1972). Hence, passage length can be considered to be a general measure of many of the surface level characteristics studied by Drum, Calfee and Cook (1980). Depth of answer is an adaptation of Meyer's (1975) notion of hierarchic propositional structure within written discourse and Christensen's (1968) notion of sentence weights, to describe levels of subordination among sentences within paragraphs. It was one of the primary non-inferential measures used in the DiStefano and Valencia (in press) study. As used in this study, depth of answer can be considered to be a measure of the amount of non-inferential cognitive processing one performs to identify superordinate

and subordinate relationships among propositions explicitly stated in a text.

When these two variables were entered into the equation the multiple R was raised to .59 (which was still not significant) solely on the predictive strength of the variable, passage length, which had a bivariate correlation of -.47 with item difficulty and was the only significant predictor of item difficulty within the equation.

DISCUSSION

The results of the present study were not consistent with those of Hillocks and Ludlow (1984) but were generally consistent with those of Drum, Calfee and Cook (1980) and DiStefano and Valencia (in press). The amount of inference in general and the specific types of inferences were not found to be significant predictors of item difficulty. Rather, item difficulty was more a function of the raw amount of information which must be processed. This would be reflected in such gross measures of surface level linguistic characteristics as passage length.

These findings can be explained by Johnson-Laird's (1983) theory that task difficulty, presumably within any domain, is primarily a function of the amount of information that must be processed and not the inherent difficulty of the cognitive operations performed on the information. In other words, inferential cognitive operations are not

inherently more difficult than non-inferential operations. Task difficulty, then, is not a function of the types of thinking involved but the sheer amount of information that must be processed and the number of alternatives that must be kept in working memory.

These findings can also be explained using LaBerge and Samuels (1974) notion of automaticity and its relationship to task difficulty. They state that once a set cognitive operations have been internalized--learned at the level of automaticity-- they require little of the capacity of working memory and consequently are not a major factor relative to the difficulty of tasks in which they are used. This position is also taken by Anderson (1983) who states that skill or procedural learning progresses through at least three stages with the last being the autonomous stage--that at which the procedure can be executed with little or no conscious attention. Relating the theory of automaticity to the present study, we might conclude that the difficulty of reading comprehension items (and, presumably, other types of items) is a function of the extent to which the cognitive processes involved have been internalized and can, consequently, be executed automatically. It might be the case, then, that the inferential cognitive operations involved in reading comprehension test items are inherently more difficult than non-inferential operations, however, those inferential operations have simply by internalized to the level of automaticity by school aged test takers.

If either of these interpretations is correct, it would imply that skill hierarchies (and, consequently, the distinction between inferential and literal items) have little practical validity as applied to standardized reading test items. Further research must be done to reconcile the discrepancy between this study and that of Hillocks and Ludlow. Perhaps the hierarchy of cognitive operations they identified is valid for relatively long blocks of discourse and/or for certain types of discourse. In other words, perhaps skill hierarchies are not independent, invariant constructs but change depending on the type and amount of information processed and the level of skill of the person engaged in the task.

REFERENCES

Anderson, J. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Christensen, F. (1963). A generative rhetoric of the sentence. College Composition and Communication. 14, 155-161.

Crothers, E.J. (1979). Paragraph structure inferences. Norwood, NJ: Ablex.

de Beaugrande, R. (1980). Text, discourse and process. Norwood, NJ: Ablex.

DiStefano, P. & Valencia, S. (in press). Reading comprehension: Factors that influence performance in narrative and expository passages. Journal of Educational Measurement.

Drum, P., Calfee, R.C. & Cook, L.K. (1980). The effects of surface structure variables on performance in reading comprehension. Reading Research Quarterly. 16, 486-513.

Grimes, J. (1972). The thread of discourse. Ithaca, NY: Cornell University Press.

Halliday, M. (1967). Notes on transitivity and theme in English. Journal of Linguistics. 3, 37-81.

Halliday, M. & Hasan, R. (1976). Cohesion in English. London: Longman.

Hillocks, G. & Ludlow, L.H. (1984). A taxonomy of skills in reading and interpreting fiction. American Educational Research Journal. 21 (1), 7-24.

Johnson-Laird, P. (1983). Mental models. Cambridge, MA: Harvard University Press.

Kintsch, W. (1979). The representation of meaning in memory. Hillsdale, NJ: Erlbaum.

LaBerge, D. & Samuels, S. (1974). Toward a theory of automatic information processing in reading comprehension. Cognitive Psychology. 6, 293-323.

Lee, L. & Canter, S. (1971). Developmental sentence scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. Journal of Speech and Hearing Disorders. 36, 315-340.

Marzano, R, Hagerty, P., Valencia, S. & DiStefano, P. (1987). Reading diagnosis and instruction: Theory into practice. Englewood Cliffs, NJ: Prentice-Hall.

Meyer, B.J. (1975). The organization of prose and its effects on memory. Amsterdam: North Holland Publishing Co.

O'Hare, F. (1973). Sentence combining: Improving student writing without formal grammar instruction. Urbana, IL: National Council of Teachers of English.

Prescott, G.A., Balow, I.H., Hogan, T.P. & Farr, R.C. (1978). Metropolitan achievement tests. New York: The Psychological Corporation.

Rosenshine, B. (1980). Skill hierachies in reading comprehension. In R.J. Spiro, B.C. Bruce & W.F. Brewer (Eds.). Theoretical issues in reading comprehension. Hillsdale, NJ: Erlbaum.

Turner, A. & Greene, E. (1977). The construction and use of a propositional text base. Boulder, CO: The Institute for the Study of Intellectual Behavior, University of Colorado.

van Dijk, T. (1980). Macrostructures. Hillsdale, NJ: Erlbaum.

Warren, W.H., Nickolas, D.W. & Trabasso, T. (1979). Event chains and inferences in understanding narratives. In R.O. Freedle (Ed.) New directions in discourse processing II. Norwood, NJ: Ablex.