DOCUMENT RESUME

ED 317 614 TM 014 732

AUTHOR Nagy, Philip
TITLE Assessing Thinking Skills in Social Problem Solving.
PUB DATE Apr 90
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Cognitive Ability; Elementary Education; *Elementary School Students; Foreign Countries; Grade 3; Grade 6; *Group Discussion; Intermediate Grades; *Problem Solving; Program Evaluation; *Protocol Analysis; Schemata (Cognition); *Social Problems; Student Evaluation; Thinking Skills
IDENTIFIERS *Ontario

ABSTRACT

The purpose of this report is to explore an analysis of discussions, among groups of elementary school children, of a social problem. The intent of the research is to contribute to the advancement of methods for program and student assessment, particularly toward goals not usually evaluated by traditional testing programs. The analysis method used is an adaptation of schema theory set against a background of recent research on the solving of ill-structured problems. The major contribution of this study is in operationalizing the imposition of a schema framework on protocol data. The particular data that form the example analyzed come from an ongoing curriculum project comparing methods of teaching thinking skills in the classroom. Subjects were 3rd and 6th graders in 18 classes at 8 rural elementary schools in 3 school board jurisdictions in southwestern Ontario. Groups of five students each were taken from class and asked to discuss for 10 minutes a potentially iniquitous family situation involving allowances and household chores and what might be done about the problem; 76 discussions were recorded. The analysis showed patterns of differences among protocols, expressed as aspects of typical, enriched, and impoverished schemata. Many of the patterns in the data could be tentatively linked to the instructional treatments. Although the context is group discussion, much of the proposed methodology is applicable to problem-solving protocols of individuals. (Author/TJH)

Assessing Thinking Skills in Social Problem Solving

Philip Nagy

Department of Measurement, Evaluation, and Computer Applications
The Ontario Institute for Studies in Education

Presented at the Annual Meeting of the American Educational Research Association, April, 1990: Boston.

Abstract

Assessing Thinking Skills in Social Problem Solving

The purpose of this report is to explore an analysis of discussions, among groups of elementary school children, of a social problem. The intent of the research is to contribute to the advancement of methods for program and student assessment, particularly toward goals not usually evaluated by traditional testing programs. The analysis method used is an adaptation of schema theory set against a background of recent research on the solving of ill-structured problems. The major contribution of this study is in operationalizing the imposition of a schema framework on protocol data. The particular data which form the example analysed come from an ongoing curriculum project comparing methods of teaching thinking skills in the classroom. The analysis showed patterns of differences among protocols, expressed as aspects of typical, enriched, and impoverished schemata. Many of the patterns in the data could be tentatively linked to the instructional treatments. Although the context is group discussion, much of the proposed methodology is applicable to the problem-solving protocols of individuals.

Introduction

The purpose of this report is to explore an analysis of discussions, among groups of elementary school children, of a social problem. The intent of the research is to contribute to the advancement of methods for program and student assessment, particularly toward goals not usually evaluated by traditional testing programs. The analysis method used is an adaptation of schema theory set against a background of recent research on the solving of ill-structured problems. The major contribution of this study is in operationalizing the imposition of a schema framework on protocol data. The particular data which form the example analysed come from an ongoing curriculum project comparing methods of teaching thinking skills in the classroom. Although the context is group discussion, much of the proposed methodology is applicable to the problem-solving protocols of individuals.

Background

The literature related to this study comes from several fields of research: first, storage in memory, in which interest has grown from memory for nonsense syllables to that for more complex phenomena (Kintsch, 1974); second, problem solving, in which interest has grown beyond simple problems with clear-cut solutions to ill-structured problems with complex solutions (Frederiksen, 1984); third, assessment, in which it is increasingly recognized that the more complex goals of education lack appropriate assessment methods (Archbald and Newman, 1988); and fourth, the teaching of thinking, in which a consensus seems to be forming that some form of direct teaching is appropriate (Resnick & Klopfer, 1989). While all four areas will be dealt with, the emphasis in what follows will be on the first two.

Schema Theory

In attempting to understand memory for materials such as stories or series of events, researchers have posited complex entities whose role it is to act as organizing principles in memory. Anderson, Spiro and Anderson (1978) describe schema theory, one version of such an organizing principle, similar to frames (Minsky, 1975) or scripts (Schank and Abelson, 1977). Schemata act as mental structures that incorporate general knowledge, and are more abstract than the particulars of a given situation. Interpretation in terms of a schema involves matching elements in a specific situation to generic characterizations, slots, or placeholders. Anderson et al report a study in which subjects read a restaurant story and a grocery-store story involving purchase of the same foods; the expectation was that the subjects' internal restaurant schemata would impose more structure on the subjects' memories, with a resulting increase in memory. The results confirmed their expectation, and support the hypothesis that complex material is stored by complex mechanisms.

The script or schema concept appears in many variations to accommodate different aspects of memory, and different types of stored material. Abelson (1981) proposes both strong and weak scripts, which differ in whether they imply sequencing or causality within the story-line. Anderson (1984) also supports both a strong and weak view of the role of schemata, but describes the contrast differently; Anderson's distinction is whether particular elements (e.g., a salad or dessert in a restaurant) are required, or merely likely. With such adaptations, a schema view of stored memories provides a flexible perspective from which to view complex data.

Cognitive theorists have debated the psychological status of schemata. Abelson (1981) argues that schemata have psychological reality, rather then being merely organizers for the convenience of researchers. On the other hand, Alba

and Hasher (1983) have reviewed the research for and against schema theory, and argue that the evidence shows that stored memories are richer than the highly selected subset predicted by schema theory. This suggests that it might be appropriate to view schema theory as a method of imposing order on complexity, not necessarily involving any strong assumptions concerning the nature of human memory. This perspective, adapted for the present research, gives the method the status of a portrayal technique, an heuristic device useful for imposing order on data.

As a device for imposing order and examining differences, schema theory holds promise. For example, Schallert (1982) notes that schemata evolve; that is, they become more elaborate and specific with experience. This suggests that an examination of the details of story-lines across individuals might be used to highlight differences in specificity or sophistication, differences which in turn might be linked to age, experience, or, in the present example, training. There is a difficulty with such a perspective, however -- how to decide what constitutes a more complete or more sophisticated version of a schema. For example, Horton and Mills (1984) reviewed the literature on human learning from a schemata perspective, using a levels-of-processing framework. They concluded that such an approach is plagued by the lack of an independent definition of depth of processing. Thus, a present limitation to the technique is reliance on subjective decisions concerning the adequacy or quality of particular pieces of data.

Problem Solving

The focus of this study is an "ill-structured" problem; children discuss a potentially iniquitous family situation involving allowances and household chores, and what might be done about it. Frederiksen (1984) summarized Simon's

original distinction between well-structured and ill-structured problems. The characteristics of ill-structured problems include greater complexity, less definite criteria for deciding if a solution has been reached, lack of complete information, absence of a "legal move generator", and no convenient list of accepted procedures. They also have higher verbal content and are more context dependent. Most "real-life" problems would be classified as ill-structured. While methods for the analysis of well-structured problems (e.g., logic puzzles, chess) have tended to be based in the artificial intelligence literature, taking the form of production rules (if... then statements) or flowcharts, such methods have not proved easily adaptable to rich verbal contexts. Voss and Post (1988) noted that the method chosen for the analysis of ill-structured problems reflects the theoretical concerns of the investigators. Three examples demonstrate the variety of theoretical concerns and approaches used.

Larkin (1980) has worked in the area of physics and algebra problems, and is primarily concerned with the teaching of problem solving. The types of problem with which she is concerned exhibit some characteristics of both well-structured (e.g., definite answer, accepted methods) and ill-structured (e.g., verbal content) problems. She has found that large-scale units such as Schank's scripts are useful in the analysis of problem solving in such domains.

Voss, Greene, Post & Penner (1983) offer a second method of analysing protocols. The problem they set for their subjects was the lack of productivity of the Soviet agricultural system. Their main concern, as discussed in retrospect by Voss and Post (1988), was the development of a framework for understanding the problem solving process. They categorized statements as one of several types of "goal structure operators" or "reasoning structure operators". Goal statements deal with relatively global moves in the discussion,

such as identification of major issues and subproblems within the Soviet system. Reasoning statements deal with the analysis provided by the respondent within the structure of these subproblems.

Finally, Lawrence (1988), also concerned with basic understanding of the problem solving process (Voss and Post, 1988), presents yet another method for analysing ill-structured problems, this time in the context of judicial decision making. Lawrence's basic model consists of elaborate if... then statements. She spends eonsiderable effort on the need for an analysis system to capture a priori perspectives ("frames of reference"), which correspond, according to Voss and Post (1988), to the magistrates' courtroom schemata.

Voss and Post's (1988) linking of methodology to theoretical framework is germane. The motivating concern for the present study is to expand the arsenal of assessment devices available at the school level. Thus, as will be seen, when faced with the choice between richness of detail and operational simplification, we have chosen the latter.

Assessment and Thinking

There is considerable dissatisfaction with the impact of traditional (i.e., multiple-choice) standardized testing programs on school curricula. Nagy, Traub and MacRury (1986), in a review of this literature, point out the danger that what is most easily assessed tends to become most important. At the same time, there is a movement toward the teaching of "higher-order" thinking skills (Resnick and Klopfer, 1989). Despite theoretical progress (e.g., Nickerson, 1989), there is some antagonism between the teaching and assessment of thinking skills and traditional standardized testing. Calls for improvement in the assessment of thinking skills (Haertel, 1986; Archbald and Newman, 1988; Stiggins, 1988) tend to be calls for development of technologies beyond the

multiple-choice item. While most educators would agree that the ability to carry on a discussion is an important outcome of schooling, it is not easily assessed, and, one might surmise, is often not emphasized in the curriculum. The present study is an attempt to assess the quality of discussions, thus contributing to the promotion of one aspect of higher order thinking in the curriculum.

The students who provided the data for this investigation are participating in a three-year study of the efficacy of direct teaching of thinking skills, based on a model developed by Beyer (1987). Since the data were collected after only one year of the project, the primary import of this study lies in the methodology used for the assessment of the group discussions, rather than in a definitive test of Beyer's model.

Method

Sample

The subjects come from eight rural elementary schools in three school board jurisdictions in Southwestern Ontario. Nine schools are part of the larger study; scheduling problems prevented data collection in one school. Three schools were chosen within each board from volunteer schools, subject to demographic constraints (e.g., avoidance of schools with high turnover, or schools with large numbers of split-grade classes). Once chosen, the three schools within each board were assigned randomly to one of the three treatments. Within each school one Grade 3 and one Grade 6 class were chosen for participation. Thus, there were three classes within each grade-treatment combination, a total of 18 classes. Students within each treatment were essentially equal in average score obtained on an administration of the Vocabulary and Reading subtests of the Canadian Tests of Basic Skills. To collect the discussion data, students were taken from class

in groups of about five and asked to discuss a problem for ten minutes. Across the two grades and three treatments, 76 such discussions were recorded.

Treatment

Data were collected at the end of one year of a three-year project, which is intended to compare three treatments. The treatments were administered to all students during three selected science units over the course of the year. Apart from subject matter content, which varied across the two grades, the units focused on or required the skills of observing, classifying, and problem solving respectively. With some minor variations, the units were taught in October (1988), February and April (1989), and took approximately 12-15 weeks. (Exact scheduling was in the hands of the 18 teachers.) The group discussion data were collected in May and June.

The Experimental groups (six classes, three from each grade) were taught by a method based on the work of Beyer (1987), in which a thinking skill is introduced and defined, rules are developed for the use of the skill, guided practice is provided and gradually removed, and transfer is explicitly taught for. These six teachers were supported by a peer coaching component of the project (Showers, 1984). Peer coaching was the major feature of the second set of six classes, called the Coaching groups. These six teachers used a traditional method of teaching science content, but had the same peer coaching support as the Experimental groups. The Control group used traditional methods of teaching science, and had no peer coaching support. All 18 teachers were supported with substantial amounts of training and in-service. The missing school was one of the Coaching schools.

The full project of which this analysis is part contains substantial amounts of curriculum monitoring, in the form of classroom observation and teacher

interviews. After one year, the experimental teachers are doing what they were asked to do; the Experimental treatment is clearly distinguishable, in practice, from the other two. After this first year, the Coaching treatment is not as distinguishable from the Control treatment. This can be attributed to the fact that volunteers had to be sought at the school level rather than the individual teacher level, since the project required participation of six classes in the same school over the three years. One of the main prerequisites for the success of coaching is that the teachers entering into such a relationship do so voluntarily. In practice for the first year, this meant reliance on a close professional relationship between one Grade 3 and one Grade 6 teacher in each school. These relationships, naturally, have varied.

Instrument

Groups of students were taken from class and presented, both orally and in writing, with the following situation (note that neither age nor gender is specified): "There are two children in the Puzzlewich family. One child is called Pat and the other is B.J. Both of the children receive the same allowance. Pat is involved in many after school activities such as music lessons, ringette, church choir, and youth group. B.J., however, just attends youth group once a week. Mrs. Puzzlewich is always asking B.J. to do extra chores around the house. She NEVER asks Pat to help out. B.J. complains to the mother that it is unfair to have to do all of the chores and yet receive the same allowance as Pat. 'I want an increase in allowance.' The Mother says, 'You aren't paid for chores. Your allowance is just for being part of this family. You may not have an increase in allowance.' What do you think?"

Groups of students varied in the enthusiasm with which they tackled the question. Discussions varied from two to twelve minutes. Although it had been

the intention that the interviewer remain out of the discussion as much as possible, most groups required substantial prompting. In extreme cases, the interviewer spoke more than all the students in the group. The amount of prompting required has been treated as a variable in the analysis.

Initial Analysis

The methodology evolved during the course of the data analysis. The eventual products are two: a method of tracking the degree of cohesion in the discussion -- the extent to which it was a conversation among the group rather than five children taking turns talking to the one adult; and a two-level category system for the statements made, organized to reveal the basic collective schema of each group with respect to family fairness. Methods of identifying indicators of typical, impoverished, or enriched schemata were developed, and comparisons of their occurrence made across treatments and grades. What follows is a reconstructed version of progress in the analysis; most cul-de-sacs are omitted, although some are discussed for their instructive value.

Cohesion. A reading of the transcripts reveals many indicators of cohesion in the group discussions: use of other individuals' names, starting an utterance with *or*, *but* or *and*; voicing specific agreement or disagreement with an earlier speaker. After exploring several of these as indicators of cohesion, a system making use of all of them, as well as any other evidence, was devised. All utterances which could be clearly linked to an earlier one were simply counted. Intervening prompts by the interviewer were excluded, as were responses only to the interviewer's prompts. Personal anecdotes or comments were ignored unl they led the group back to the topic.

Schemata. The analysis began with development of an elaborate category system for the statements made. The procedure was ad hoc; each time a statement was encountered that did not fit existing categories, a new category was created. Some categories, in the end, served no useful function, others served to simplify the data, and still others formed the basis for the eventual development of the schema theory perspective.

Those categories that served no purpose, usually due to lack of frequency, included the connectives described in the previous paragraph, statements of facts from the case, statements of assumptions about the case (e.g., that Pat probably pays the fees for his/her activities); humour and fantasy; lapses into incoherence or self-contradiction; and comments on the progress of the discussion.

Those categories that helped to keep the data simple included:

1. General prompts from the interviewer, intended to get the ball rolling;
2. Specific prompts from the interviewer, intended to get more information on a student's point;
3. General agreement from a student, usually just chiming in;
4. Specific agreement from a student, directly to another's previous statement;
5. Personal anecdotes -- these were initially subcategorized as related or unrelated to the discussion, but this distinction did not prove useful.
6. Details, which were usually expanding on a point beyond a level judged useful for the intended analysis (e.g., adding to a list of chores that Pat could do when time permitted).

Those categories that formed the core of the schema view included the following:

1. Unfairness statements -- 11 different categories were created to

accommodate the variety of statements as to why the situation was unfair to B.J.;

2. Proposed solutions -- 40 different categories were created to categorize the variety of actions suggested by the students;
3. Cautions -- 9 categories were created to deal with statements expressing cautions, usually about why another's suggestion might not work;
4. Age statements -- 7 categories were required to deal with the variety of expression of the relationship between age, responsibility, and allowance;
5. Value positions -- 49 categories were required to deal with the variety of value positions expressed.

Second Analysis

After the initial analysis, several problems became apparent. A system with more than 100 categories would have some difficulty in producing a useful analysis. Many of the categories were used by only a couple of groups or individuals within the groups, and many captured very subtle distinctions in meaning. Boundaries among the five major categories were unclear: age and unfairness statements were, in fact, types of value statements; due to the nature of conversational language, solutions and positions could not always be distinguished (e.g., students often began with something like "You could..."); many of the statements categorized as value statements stretched the definition of the term, resulting in the last listed category being very much a catch-all; and given that the value category was loose, most of the caution statements could be reconstrued as varieties of value positions. Considerable collapsing and rearranging was both necessary and relatively easy.

Given these problems, a second sorting, with some amalgamation and deletion, of these 116 categories yielded six Value categories and seven Solution

categories as follows:

Value-A -- statements showing acceptance of responsibility for tasks, and awareness of the broader family context;

Value-B -- statements showing a disregard for responsibilities, including statements that Pat ought to help only when convenient;

Value-C -- statements that the family ought to operate on a monetary basis;

Value-D -- statements showing awareness of age, and its impact;

Value-E -- statements about the feelings of anyone in the family;

Value-F -- statements that both chores and extracurricular activities have value for the individual engaged in them;

Solution-A -- solutions which involve differential allocation of allowance or non-monetary awards;

Solution-B -- solutions which involve achieving, by a variety of means, a balance between the story characters of activities, chores, and rewards;

Solution-C -- a catch-all for unlikely or irrelevant proposals;

Solution-D -- weaker solutions involving fairness when convenient;

Solution-E -- solutions involving unilateral action by B.J.;

Solution-F -- solutions which involve emphasis on a process, such as discussion or keeping records, or setting up a schedule;

Solution-G -- more responsible solutions involving family cooperation and sharing costs.

Development of Schemata

Differences in quality of the responses are evident in the above listings of the six Value categories and seven Solution categories. Two possible paths for subsequent analysis seemed possible: one, categorize all discussions as enriched, typical, or impoverished, on the basis of all thirteen categories, and examine

patterns across grades and treatments; or two, identify enriched, typical, and impoverished treatments on the basis of each category, and examine patterns across grades and treatments. Even a cursory examination of the data demonstrated that the first path would be impractical; the evidence across the thirteen categories was not consistent enough. Therefore, the latter path was chosen.

For initial purposes, the thirteen statement categories were subjectively rated for quality as follows:

	Value	Solution
Typical	A	A, B
Impoverished	B, C	C, D
Enriched	D, E, F	E, F, G

How this categorization was applied to the data, and the subsequent evolution of the analytic method, follows in the *Results* section of this report.

Data Display

The analysis reported herein is exploratory. Precisely what is being counted in the various results below has evolved, and the developed categories presently lack precise definition. Categorizations also lack any inter-rater reliability estimates. A casual examination of the data suggests that statistical tests (c.f., analyses of variance) would be appropriate. However, a more detailed examination suggests that such analysis might not be supportable; such tests were not done. Most results are reported as box-and-whisker plots (Tukey, 197). In box-and-whisker diagrams, the dots at the left and right ends of each display represent the minimum and maximum values; the dot within the box represents the median, and the box itself encloses the middle one-half of the cases. In some

situations, different tendencies can be identified depending on whether one focuses on the box, the bulk of the distribution, or the whiskers, the outliers.

Results

Cohesion and Related Variables

Three variables were examined which might bear a relationship to the ability of the groups to sustain a discussion:

1. Cohesion -- the number of statements made that respond directly to or follow directly from a previous statement. This variable is plagued by dependence on overall length of the discussion; however, when cohesion was expressed as a fraction of length, the result was found to be misleading due to the great variety in the length variable.
2. Prompts -- the number of prompts required from the interviewer to get the discussion started or to keep it going. Note that interviewers were not instructed to be consistent in deciding when a prompt was required.
3. Length -- the total number of speaker changes, excluding the interviewer.

Figure 1 contains box-and-whisker displays for these three variables. Within the Grade 3 cohesion results, there is a slight trend favouring the Experimental group. In Grade 6, the trend favours the Coaching group. In both grades, the Coaching groups required more prompts to keep the discussion going. Finally, in Grade 6, the Control group's average production was substantially shorter than the other groups.

****

Insert Figure 1 about here

****

Schemata

A tabulation of percentage responses within each grade across the six Value

categories and seven Solution categories is reported in Table 1. (The raw frequencies, not reported, reveal that the Grade 6 groups made proportionally 19% more statements than the Grade 3 groups.) Approximately 72% of the Grade 3 responses and 61% of the Grade 6 responses are captured in the Typical Schemata. The corresponding figures for the Impoverished Schemata are 11% (Grade 3) and 13% (Grade 6); and for the Enriched Schemata, 17% (Grade 3) and 26% (Grade 6). Interestingly, two of the views dubbed Impoverished, Value-C and Solution-D, are more common among Grade 6 than Grade 3 students. For example, there were 16 disagreements expressed with the mother's statement that allowance ought to be just for being a member of the family -- all from Grade 6 students. This is probably best explained by increased self-centredness resulting from the approach of adolescence, or by increased recognition of the importance of money. At the same time, the tentative and subjective nature of these categorizations ought not be overlooked. For example, some of the Solution-E unilateral actions, such as spending more time at friends' homes to avoid chores, might well be classified differently by another analyst. As any parent or teacher would attest, growth toward adolescence can be a tortuous path.

****

Insert Table 1 about here

****

The frequency of statements which were categorized as Value-A, Solution-A, or Solution-B (the typical perspective) was examined across groups. Four measures were created. The first three were simple frequencies within the three just-named categories. These might be considered as measures of the amount of discussion related to what has been dubbed the "typical" view of family fairness: acceptance of responsibility for tasks, and awareness of the broader family

context; suggestions of solutions which involve differential allocation of allowance or non-monetary awards; or suggestions for solutions which involve achieving, by a variety of means, a balance between the children in the story of activities, chores, and rewards. The fourth measure was an attempt to capture the breadth of discussion. In the original categorization system, there were 13 categories which eventually were collapsed into the two categories, Solution-A and Solution-B. Thus, the fourth variable was the number of these original 13 categories touched upon by each group. Figure 2 contains the data on the typical fairness schema.

****

Insert Figure 2 about here

****

As can be seen, there are differences across groups in the extent to which they express sentiments dubbed part of the typical view of family fairness. These differences, however, do not fall into a simple pattern. For the expression of Value-A sentiments, a slight pattern favouring the Experimental over the Control groups can be seen in both grades, with the Coaching groups somewhat intermediate. Within Solution-A, a less regular pattern can be seen favouring the Coaching groups. There is an irregular pattern within Solution-B, but examining Solution-A and Solution-B together, it is possible to perceive a slight tendency for Experimental groups to prefer Solution-B, balancing activities, chores, and rewards over Solution-A, giving differential reward for differential chores. The strength of this tendency, however, should not be exaggerated. The Solution-Breadth variable reveals no compelling patterns.

Frequencies within the Impoverished and Enriched Schemata are too low to allow the use of box-and-whisker diagrams. Table 2 contains these data,

expressed as average occurrence per group discussion. The across-grade differences are as interesting as those across treatments. The Grade 6 students have made proportionally more than twice as many of the Enriched Schemata statements, as would be expected, but they have also made more of the Impoverished Schemata statements. These statements are essentially of two kinds: irrelevant or unworkable (e.g., increase both allowances or sneak money to B.J.) and selfish (e.g., help out when you happen to be home). The first come largely from Grade 6, in a ratio of about 3:2, while the second come largely from Grade 3, in a ratio of about 3:1.

****

Insert Table 2 about here

****

There are as well some discernible treatment effects in Table 2. In Grade 6, the Enriched perspective seems to be evident about twice as often in the Experimental and Coaching groups as in the Control group. The pattern in Grade 3 also seems to favour the Experimental group. More detailed examination, not evident in Table 2, reveals other features. First, there were 100 statements from all groups that had to do with B.J. taking matters into his/her own hands. Of these, 85 said merely that s/he ought to get involved in other activities. However, of the 15 statements that said more than that (e.g., get a job, disappear after school, go on strike), 13 came from the Experimental Grade 6 group. Second, statements across both grades calling for setting up of a process for resolution of the problem (family conferences, point systems, schedules) came from the Experimental group in a ratio of more than 2:1. Finally, the most sophisticated and mature concepts of family responsibility (e.g., activities for children cost parents money, everyone ought to contribute to the common good)

came from the Experimental groups in the ratio of about 3:1.

Discussion

This analysis rests on some assumptions concerning the nature of the data which need to be discussed before dealing with the specifics of the analysis. Schrag (1988) has argued that there is no way of assessing the thinking required for a task unless we know what tools the thinker has available. Since thinking processes are not directly observable, they must be inferred from observation of the relationships between input and output,. in this case between B.J. and Pat's family problem and the recorded discussions. This inferential difficulty is a commonplace; one needs to accept a reasonable amount of inference in cognitiv research. In the present case, one needs to accept that *generally* the typical student's statements are an adequate representation of the typical student's thinking.

Perhaps more contentious is the acceptance that in *particular* each student's speech adequately represented his/her thought processes. That is, what effect did the group have on the ability of each individual to think out the issue and express an opinion? There is no evidence available on the question of whether some individuals felt compelled to either remain silent or voice passive agreement when faced with the expressed opinions of more assertive classmates. This is a real limitation to the data available, which will, in future work be dealt with through individual post-interviews. At this time, however, we can examine data that is somewhat relevant, the length of each discussion. Wi ive students per group, a discussion length of 10 means that the average student took advantage of the opportunity to speak twice. Seven of the 76 discussions (5 from Control groups) were shorter than 10, while 57 were longer than 20.

While there are no data on individual behaviour within the group, it seems safe to conclude that, while some students might have been unduly reticent, a substantial majority probably took the opportunity to express their views.

The purpose of this research has been to explore the potential of schema theory for distinguishing between group problem solving protocols. The method that has evolved requires identification of aspects of typical, impoverished, and enriched schemata within each protocol, rather than identification of entire protocols as either typical, impoverished, or enriched. The method has identified differences across protocols, which, in turn, have been linked to the different treatment groups. This linkage is not particularly strong or consistent, but there is no reason to expect otherwise. First, the ability to generate a thoughtful discussion in a social context has to be considered as far transfer from the treatments, which focused on the thinking skills of observing, classifying and problem solving in the context of science content. Second, the data were collected after only one year of such treatment. Even the most ardent proponents of direct teaching methods (e.g., Beyer, 1987) recommend a multi-year, multi-context treatment before expecting real gains. Detailed investigation of the outcomes of the thinking skills curriculum, currently in progress, requires a different methodology, involving longer instructional times, a multivariate set of outcome measures, and many more classrooms.

The thinking skills curriculum project has served as a vehicle for the examination of the feasibility of a theoretically grounded method of comparing solutions to ill-structured problems. It is legitimate to assess the success of the analytic method used, and what might be required by way of refinements.

One difficulty of the method, identified in the literature (Horton and Mills, 1984), is that the categorization of statements is at root subjective. What one

would like to consider as a deeper level of processing could as easily be construed simply as more like the sentiments adults would like to see children express. The problem with this feature of the method is demonstrated by the fact that the Grade 3 students appeared less selfish, in aspects of their protocols, than the Grade 6 students. One might choose to define growth in perception of the situation empirically by accepting what might be a natural outcome of adolescence. A better face may be put on the situation by substituting "assertive" for "selfish". Or, one might choose to consider what is desirable from the adult perspective as a valid curricular goal, and take the Grade 6 results as undesirable. Whatever the decision, it is beyond the scope of this paper. The problem remains, however, of having to distinguish level of moral development (however defined) from level of cognitive processing.

A second difficulty of the method is that, given the desirability of the category system evolving from the data, where one starts is important. When to open a new category is an arbitrary decision, based on a subjective view of the history of the analysis. Whatever "category width" might mean in this context, effort needs to be spent in holding it somewhat constant. In the present study, the analyst was blind to treatment, but not to grade; the Grade 3 protocols were analysed first. This has an unknown effect on the evolution of the category system. Tied in with the obvious issue of simple inter-rater reliability, already mentioned, it would seem important for different analysts to analyse the data in different orders. The issue is somewhat simplified by the possibility that an already-created category system might be imposed on the data, but there still remains the difficulty of valid and reliable creation of that first set.

It is legitimate to ask what has been accomplished by an imposition of schema theory that might not have been done from a more traditional perspective, such

as a relatively theory-free development of a "marking scheme". First, there is ample evidence that teachers require assistance in assessing higher-order outcomes of instruction (Haertel, 1986; Stiggins, 1988). Neither the identification of what constitutes higher-order thinking nor the development of appropriate marking systems is a trivial task. Both require development and imposition of a theoretical framework. Pursuit of notions of typical, impoverished and enriched story-lines for complex situations is appropriate to such a task. Second, there are calls from those studying the assessment practice of teachers (e.g., Stiggins and Bridgeford, 1985) for more focused methods. A method allowing comparison of practice with a well-developed image of what might be qualifies as a focused method, both for research and for the improvement of practice. Finally, the method proposed generalizes over contexts. Variations are being developed in ongoing work for both analysis of essay responses to social problems and individual responses to practical ill-structured situations.

Footnote. This project was supported by an Ontario Ministry of Education grant to Middlesex, Oxford, and Perth County Boards of Education. I wish to thank the students and staffs of these boards for their assistance and cooperation. I also wish to thank Patricia Allison and Lynn Watson for the data collection.

References

Abelson, R. P. (1981). Psychological status of the script concept. American Psychologist, 36, 715-729.

Alba, J. W., & Hasher, L. (1983). Is memory schematic? Psychological Bulletin, 93, 203-231.

Anderson, R.C. (1984). Some reflections on the acquisition of knowledge. Educational Researcher, 13(9), 5-10.

Anderson, R.C., Spiro, R.J. & Anderson, M.C. (1978). Schemata as scaffolding for the representation of information in connected discourse. American Educational Research Journal, 15, 433-439.

Archbald, D.A., & Newmann, F.M. (1988) Beyond standardized testing. Reston VA: National Association of Secondary School Principals.

Beyer, B.K. (1987). Practical strategies for the teaching of thinking. Toronto: Allyn & Bacon.

Frederiksen, N. (1984). Implications of cognitive theory for instruction in problem solving. Review of Educational Research, 54, 363-407.

Haertel, E. (1986). Choosing and using classroom tests: Teachers' perspectives on assessment. A paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Horton, D. L., & Mills, C. B. (1984). Human learning and memory. Annual Review of Psychology, 35.

Kintsch, W.A. (1974). The representation of meaning in memory. New York: John Wiley & Sons.

Larkin, J. H. (1980). Teaching problem solving in physics: The psychological laboratory and the practical classroom. In D. T. Tuma & F. Reif (Eds.), Problem solving and education: Issues in teaching and research. Hillsdale,

NJ: Lawrence Erlbaum Associates.

Lawrence, J.A. (1988) Expertise on the bench: Modelling magistrates' judicial decision-making. In M.T.H. Chi, R. Glaser, and M.J. Farr (Eds.). The nature of expertise. Hillsdale, NJ: Lawrence Erlbaum Assoc.

Minsky, M. A. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), The psychology of computer vision. New York: McGraw-Hill.

Nagy, P., Traub R. & MacRury, K. (1986). Strategies for evaluating the impact of province-wide testing. Toronto: Ministry of Education.

Nickerson, R:S. (1989). [Special Issue]. Educational Researcher, 18(9).

Resnick, L. & Klopfer, L. (Eds.). (1989). Toward the thinking curriculum: Current cognitive research. ASCD yearbook. Association for Supervision and Curriculum Development: Alexandria, VA.

Schallert, D.L. (1982). The significance of knowledge: A synthesis of research related to schema theory. In (W. Otto & S. White, Eds.). Reading expository material. Toronto: Academic Press.

Schank, R. C., & Abelson, R. P. (1977). Scripts, plans, goals and understanding. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schrag, F. (1988). Thinking in school and society. New York: Routledge.

Showers, B. (1984). Peer Coaching: A Strategy for Facilitating Transfer of Training. Center for Educational Policy and Management, Eugene, Oregon. 97403

Stiggins, R. J., & Bridgeford, N.J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22 (4), 271-286.

Stiggins, R.J. (1988). The nature and quality of teacher-developed classroom assessments. A paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.

Tukey, J. W. (1977). Exploratory data analysis. Toronto: Addison-Wesley.

Voss, J.F., Greene, T.R., Post, T.A., & Penner, B.C. (1983). Problem solving skill in the social sciences. In G.H. Bower (Ed.). The psychology of learning and motivation: Advances in research theory. (pp 165-213). New York: Academic Press.

Voss, J.F. & Post, T.A. (1988). On the solving of ill-structured problems. In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.). The nature of expertise. (pp. 261-285). Hillsdale, NJ: Lawrence Erlbaum.

Table 1

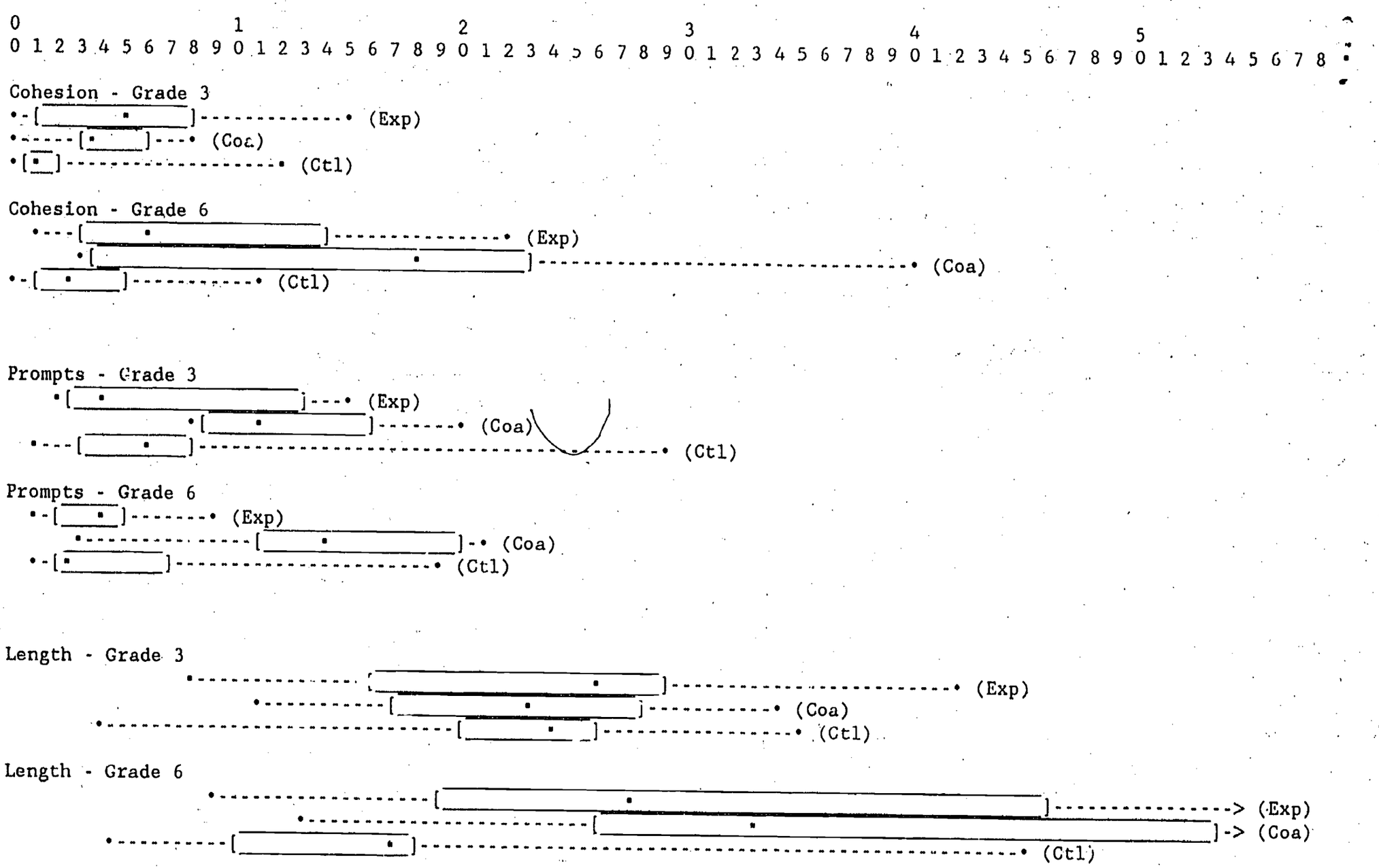
Percentage of Statements by Value and Solution Category

	Grade 3	Grade 6
Value Categories		
Typical		
A - accept responsibility	11.0	12.4
Impoverished		
B - disregard responsibility	0.0	1.1
C - operate on financial basis	4.3	5.9
Enriched		
D - awareness of age	3.5	2.4
E - concern for feelings	0.6	2.4
F - value of effort	4.5	6.3
Solution Categories		
Typical		
A - differential reward	19.2	18.9
B - balance of activities	41.7	29.5
Impoverished		
C - weak solutions	3.5	1.3
D - fairness when convenient	3.5	4.9
Enriched		
E - unilateral action	3.5	8.2
F - emphasis on process	2.5	2.9
G - sharing and cooperation	2.0	3.8

Table 2

Average Occurrence, per Group Discussion, of Statements from Impoverished and Enriched Schemata

	Grade 3	Grade 6
Impoverished		
Experimental	2.3	2.3
Coaching	1.8	4.3
Control	1.5	1.6
Enriched		
Experimental	3.4	6.6
Coaching	1.5	6.0
Control	2.4	3.3



Grade 3 - Exp, Coa, Ctl, N = 9, 9, 13; Grade 6 - Exp, Coa, Ctl, N = 18, 9, 18

Figure 1
Box-and-Whisker Diagrams of Cohesion and Related Variables

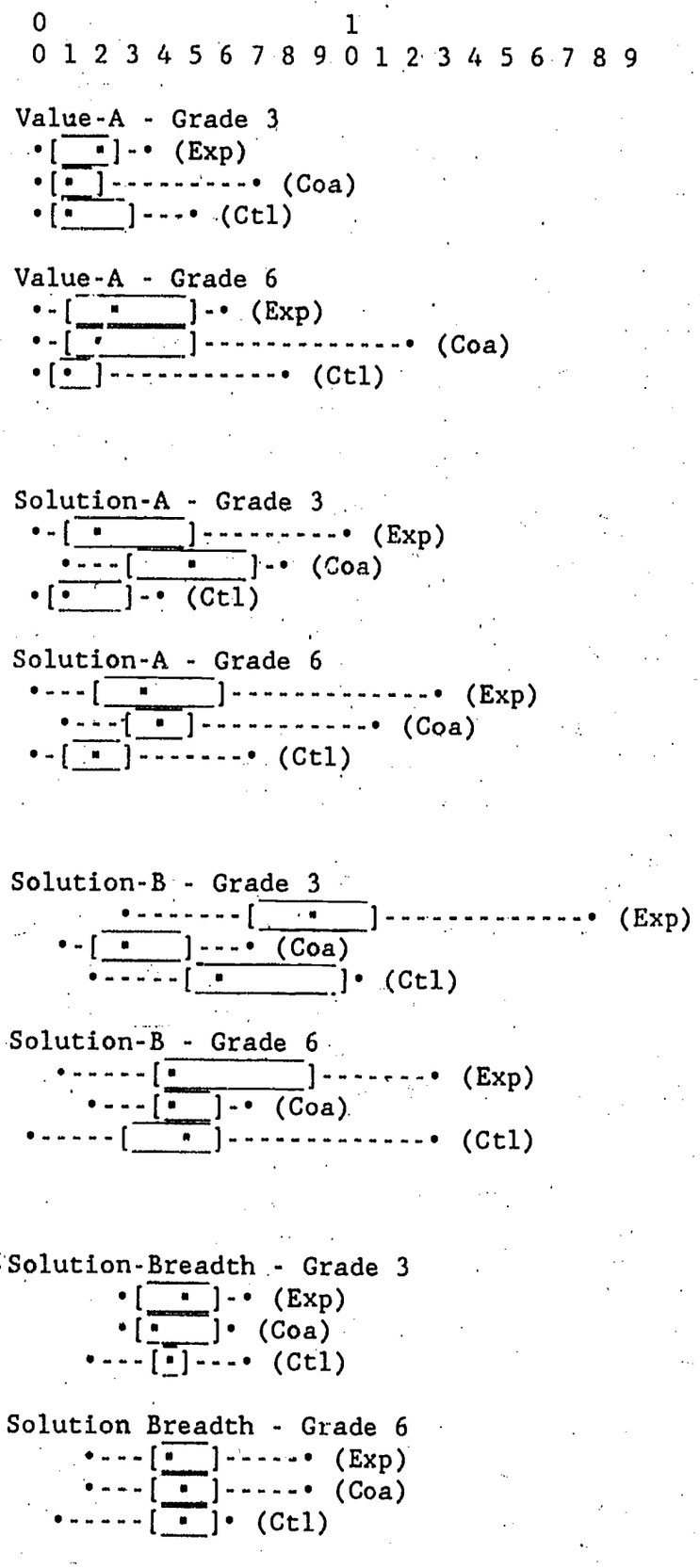


Figure 2
Box-and-Whisker Plots for "Typical" Schemata Variables