DOCUMENT RESUME

ED 334 199 TM 016 447

AUTHOR Ross, John A.; Cousins, J. Bradley
TITLE Patterns of Student Growth in Reasoning about Multivariate Correlational Problems.
INSTITUTION Ontario Inst. for Studies in Education, Peterborough. Trent Valley Centre.
SPONS AGENCY Ontario Dept. of Education, Toronto.
PUB DATE Apr 91
NOTE 36p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Age Differences; *College Students; Comparative Analysis; *Correlation; *Elementary School Students; Elementary Secondary Education; Foreign Countries; Higher Education; Individual Development; Multivariate Analysis; Problem Solving; *Secondary School Students; *Student Development; *Thinking Skills; Trend Analysis
IDENTIFIERS Canada

ABSTRACT

Previous studies of the development of correlational reasoning have focused on the interpretation of relatively simple data sets contained in 2 X 2 tables. In contrast, this study examined age trends in subjects' responses to problems involving more than two continuous variables. The research is part of a multi-year project to conceptualize student development on correlational reasoning skills and to enhance performance through instructional interventions. In Studies 1, 2, and 3, instruments were progressively developed and administered to small samples of subjects in grades 4, 5, 7, 9, 10, and 11 and postgraduate education (n=20 in each grade in each study). The resulting multidimensional profiles of student growth showed that subject performance in reasoning about correlational problems increased slowly and weakly with age. Study 4 found that students' ability to solve multivariate correlational problems could be improved through a modest amount of instruction. Similarities between findings from research concerning 2 X 2 data problems and research on multivariate continuous data problems were attributed to the existence of central conceptual structures, the scope of which remains to be determined. There are four tables and two figures of study data. (Author/SLD)

PATTERNS OF STUDENT GROWTH IN REASONING ABOUT MULTIVARIATE CORRELATIONAL PROBLEMS

Dr. John A. Ross and Dr. J. Bradley Cousins

The Ontario Institute for Studies in Education,
Trent Valley Centre
Box 719, 150 O'Carroll Avenue
PETERBOROUGH, Ontario
CANADA K9J 7A1

A paper presented at the annual conference of the American Educational Research Association, Chicago, 1991.

Removed for Blind Review[1]

Maynes, F., & Ross, J. (1984). Integrating school curricula with social concerns. *Alberta Journal of Educational Research, 30* (3), 179-193.

Robinson, F., Ross, J., & White, F. (1985). *Curriculum development for effective instruction.* Toronto: OISE Press.

Ross, J. (1986). Cows moo softly: Acquiring and retrieving a formal operations schema. *European Journal of Science Education, 8* (4), 389-397.

Ross, J. (1988a). Improving social-environmental problem solving through coopertive learning. *American Educational Research Journal, 25* (4), 573-591.

Ross, J. (1988b). Controlling variables: A meta-analysis of training studies. *Review of Educational Research, 58* (4), 405-437.

Ross, J. (1990). Learning to control variables: Main effects and interactions of two rule governed approaches. *Journal of Research in Science Teaching, 27* (6), 523-540.

Ross, J., & Maynes, F. (1981). *Geographic Thinking Skills.* Toronto: OISE Press.

Ross, J., & Maynes, F. (1983a). Teaching problem solving: An instructional design strategy. *Canadian Journal of Education, 8* (2), 155-173.

Ross, J., & Maynes, F. (1983b). Experimental problem solving: An instructional improvement field experiment. *Journal of Research in Science Teaching, 20* (6), 543-556.

Ross, J., & Robinson, F. (1987). The use of rule structures in teaching experimental design to secondary school students. *Science Education, 71* (4), 571-589.

----

1. The research was funded by the Ontario Ministry of Education through a grant to the Ontario Institute for Studies in Education. The views expressed here are not necessarily those of the Ministry of Education. Floyd Robinson contributed to the conceptualisation of the project and Anne Hogaboam-Gray contributed to all phases of the activities.

Abstract

Previous studies of the development of correlational reasoning have focussed on the interpretation of relatively simple data sets contained in 2 X 2 tables. In contrast the research reported here examined age trends in subjects' responses to problems involving more than two continuous variables. In Studies 1, 2 and 3 instruments were progressively developed and administered to small samples of subjects ranging from grades 4 to post-graduate (n = 20 in each grade). The resulting multi-dimensional profiles of student growth showed that subject performance in reasoning about correlational problems increased slowly and weakly with age. Study 4 found that students' ability to solve multivariate correlational problems could be improved through a modest amount of instruction. Similarities between findings from research into 2 X 2 data problems and research on multivariate continuous data problems were attributed to the existence of central conceptual structures, the scope of which remain to be determined.

Patterns of Student Growth in Reasoning About Multivariate Correlational Problems

Correlational reasoning skills are an essential part of many school subjects, especially Science, Geography and Math. Correlational problems are also relevant to real life tasks in which the main difficulty is dealing with illusory correlation--recognizing that despite superficial appearances based on limited data, certain variables are not related. Adult performance in recognizing noncorrelations is poor (Ross & Lepper, 1980; Smedslund, 1963), even for adults with expert knowledge of the domains in which specific correlational problems are embedded (Chapman & Chapman, 1982; Jenkins & Ward, 1965). One of the most important contexts for coming to grips with spurious correlations arises when claims are made about relationships between racial membership and measures of good citizenship; stereotyping may be the result of incorrect correlation (Jennings, Amabile, & Ross, 1982).

The research reported in this article is part of a multi-year project to conceptualize student development on correlational reasoning skills and to enhance performance through instructional interventions. The portion of the research reported here had several purposes: (a) to review previous attempts to conceptualize student growth on correlational reasoning tasks, (b) to develop a technique for assessing students' reasoning in mulitivariate problems involving continuous variables, and (c) to identify patterns of growth under conditions of existing instruction.

Review of Previous Research

Correlational problem solving involves finding the degree of association between two or more variables which cannot be physically manipulated by the problem solver. In instructional contexts, reasoning about correlation might more properly be labelled regression because it typically refers to the search for causal relationships. Correlational problems are intuitively similar to experimental problems and for this reason science educators assign both types to the same category of objectives--integrated science processes. Some learning theorists have also put both types in the same category. For example, Piaget included both in formal operations, the highest level of cognitive functioning, arguing that correlational reasoning is prerequisite to experimental reasoning (Inhelder & Piaget, 1958).

There is substantial empirical evidence that performance on the two types is independent. For example, when Yap & Yeany (1988) attempted to construct a learning hierarchy of formal operations and integrated process skill items, they found that correlational reasoning did not fit into the hierarchy while the remaining constructs did. Similarly, a factor analysis of a large set of formal operations and integrated process items by Baird & Borich (1987), and a re-analysis of the data by Roth (1989), found that correlational items did not load on the same factor as the experimental items.

A computer search of research data banks (Resources in Education, Current Index to Journals in Education, Ontario Educational Research Information Service and Dissertation Abstracts) and subsequent manual branching identified 18 studies that investigated the development of correlational reasoning skills in multi-age samples. To be included in the review the studies had to meet three criteria: First, the study had to contain a task requiring the performance of correlational reasoning, (i.e., a task in which subjects had to find the degree of association between two or more variables which could not be physically manipulated by the subject). Excluded by this criterion were a large number of studies in which subjects had to interpret data produced by conducting an experiment; for a review of this literature, see Authors, (1988b). Also excluded was a smaller set of studies examining the development of key statistical concepts such as averaging. On first examination these studies (e.g., Gal, Rothschild and Wagner, 1990) appeared to be relevant in that subjects were asked to find a relationship between one variable (e.g., grades of students) and another (e.g., membership in Class A or Class B), but the recommended technique for finding the answer (calculate the average for each group and compare) reduces the problem to two data points. Second, the study had to provide an explicit coding scheme for assessing performance on the correlational task. Excluded by this criterion were investigations (e.g., Koslowski & Okagaki, 1986) which examined the extent to which various factors, including correlational evidence, were used by subjects in making causal attributions. Third, the study had to provide data for several age groups, preferably (but not necessarily) with rigorous statistical tests to substantiate claims about age trends. Table 1 describes selected characteristics of the studies included in the review.

Table 1 About Here

All but one of the studies were cross-sectional; only Niemark (1975) contained longitudinal evidence. The samples extended from the earliest school years through university and almost all found significant age trends. The pattern was not entirely linear; for example, some researchers found no significant differences between the performance of students in grade 7 and grade 10 (Authors, 1981; Shaklee & Mims, 1981), grade 7 and college (Lawson, 1982), grade 10 and grade 12 (Padilla, McKenzie & Shaw, 1986) and grade 10 and college (Shaklee & Mims, 1981).

With a few exceptions (noted in Table 1) subjects were assigned simple problems involving two dichotomous variables. In a few instances subjects were required to sort the data provided by the stimulus into categories that were meaningful to them (Adi, Karplus, Lawson & Pulos, 1978; Inhelder & Piaget, 1958; Lovell, 1961; Niemark, 1975) or plot the information on a graph (Curtis, 1985; Wavering, 1989). In the remainder, stimulus data were presented in an appropriate data processing framework: a table, a graph or a spatial arrangement in which concrete representations of individual cases (e.g., mice, plants, fish) were displayed in a two-way classification analogous to a 2 X 2 table. The task assigned to subjects was to find out if a relationship existed between two variables and, in most studies, to provide a rationale for their decision. In 12 studies, data given to subjects provided significant relationships exclusively; subjects were required to exclude as well as include relationships in six studies (Adi et al., 1978; Inhelder & Piagst, 1958; Kuhn & Brannock, 1977; Kuhn & Ho, 1977; Lawson, Adi & Karplus, 1979; Lovell, 1961).

Most researchers applied a qualitative rating scheme to assess strategies used to solve correlational problems; growth was defined in terms of a hierarchy of levels of performance rather than in terms of meeting or failing to meet a single criterion. The source of the coding schemes (the last column of Table 1) was difficult to determine from the descriptions given in the studies, but it appeared that less than half the investigators made an attempt to derive their coding schemes constructively from the data rather than imposing a priori schemes. The dominant paradigm in these studies was Piagetian; the majority of researchers attempted to report their observations of subject performance within universal stages of growth in logical thinking. Although there was considerable variation in the levels reported, a modal coding scheme could be derived containing five categories of response to a problem involving two dichotomous variables (e.g., healthy/diseased plants, with/without bug spray):

- Level 0 was a residual category containing uninterpretable responses and meandering thoughts that appeared to go in no particular direction.

- Level 1 consisted of responses based on a single cell of a 2 X 2 table; e.g., if there were many healthy plants that received bug spray, then plant health and insecticide are related.

- Level 2 consisted of responses based on two cells; e.g., if there were many healthy plants and few diseased plants that received bug spray, then plant health and insecticide are related.

- Level 3 consisted of responses based on comparing confirming and disconfirming cases across all four cells; e.g., if the sum of the healthy plants that received bug spray plus the diseased plants that did not (confirming cases) exceeded the sum of the diseased plants that received the spray plus the healthy plants that did not (disconfirming cases), then plant health and insecticide are related.

- Level 4 consisted of responses based on calculating percentages and ratios across all four cells; e.g., if the percentage of healthy plants is higher within the group that received bug spray than it was among the group that did not, then plant health and insecticide are related.

Variations from this scheme largely consisted of making finer distinctions within each of the categories (e.g., Adi et al., 1978) or adding additional levels to accommodate subject inconsistencies between trials (e.g., Kuhn & Ho, 1977). Non-Piagetian schemes were developed by Authors (1985), Curtis (1985) and Wavering (1989).

Previous studies of correlational reasoning, viewed as a body of research, suffer from several shortcomings. The first concerns the limited scope of the task assigned to subjects. Most investigators focused on simple two variable problems, yet students are more likely to encounter multivariate situations, both in school settings and in out-of-school contexts. So too, the tasks assigned to subjects tended to be limited to dichotomous variables, even though a large proportion, if not the majority, of data that students will encounter are continuous. In addition the majority of studies did not provide data sets which called for the finding of a nonrelationship. The ability of students to reject false positives was rarely tested, even though the ability to debunk false correlational claims is an important and challenging responsibility of literate adults and is one that is not easily learned by students. Donnelly & Welford (1989), for example, found that most 15-year-olds were

> ...not able to operate criteria for the elimination of "relationships"....The constant emphasis on relationships, and the limited attention to unrelated data, has an effect here, so that the possibility of the absence of a relationship is resisted (or) pupils may have difficulty deciding when an "effect" is so embedded in random variation or operates in so narrow a range as to be neglected.

The second concern is that previous researchers in this domain have given little attention to how subjects represent correlational problems. In most instances subjects were given data within a frame that facilitated the search for relationships; rarely were they asked to organize and select information themselves. Since problem representation (the identification of the elements of the problem and their relationship) has come to be recognized as a critical determinant of success in other problem solving domains, the relative neglect of this issue in correlational reasoning research is anomalous.

A third concern is that the developmental trends described by earlier studies may not be able to accommodate individual differences. When constructing developmental profiles there is an inevitable trade-off between highlighting central tendencies and categorizing individual variations. Virtually all previous investigators constructed "single strand" profiles of growth. That is, the various dimensions of growth were combined into levels and arranged in a single sequence. In some studies deviations from the central trend were addressed in increasingly complex subcategory systems, but even in these the maintenance of a linear scheme assumes an unwarranted regularity. In contrast, a multi-dimensional profile describing levels on each dimension would allow for finer distinctions in describing the performance of individual students.

A final concern, affecting only a portion of previous studies, relates to the source of data for conceptualizing student growth. Many of these conceptions were developed a priori through logical analysis of task demands and imposed on the data. In contrast, constructivist contributions to the understanding of student reasoning suggest that ideas about development should be derived from student data.

An attempt was made to reduce these deficiencies in the research reported here. Multidimensional profiles of student growth in correlational reasoning, and a technique for assessing growth, were progressively developed through four data collections.

Study 1

The specific purpose of Study 1 was to devise a technique for assessing performance on a multivariate correlational task and to use the technique to develop an intitial conception of student growth.

Method

Subjects. The subjects were multi-age convenience samples drawn from four school districts in central Ontario. There were 20 respondents (half male, half female) in each of grades 4, 7, 10, Ontario Academic Credit (OAC, a college preparation year) and university undergraduates who had taken one course in statistics.

Instrument. The review of previous research provided three criteria for specifying the complexity of the task to be performed by students: (a) the task should be multivariate rather than bivariate; (b) the task should include continuous, as well as dichotomous, variables; (c) subjects should be required to reject, as well as accept, purported relationships. To these a fourth criterion was added: (d) the task should make limited knowledge demands on subjects because tasks which combine heavy skill and knowledge demands create problems of cognitive overload that reduce performance. In addition, there is evidence that skill items that make limited knowledge demands are more effective in detecting the effects of instruction (Authors, 1988b).

Correlational reasoning instruments used by previous researchers were examined but none met all four criteria for task complexity. Consequently a new instrument was developed. The instrument provided subjects with data for 18 cases; for each case there was information on a child's age, sex and how far the child swam in two minutes. The following directions were given:

> Mrs. Campbell, the swimming coach, tests the swimming team every week. Here are the results from last Saturday (distance in 2 minutes).
>
> (a) Does swimming have anything to do with age? Write a sentence that says how they might be related.
>
> (b) Find out if swimming is related to age. Use only the information given in the the table. Show your work.

Procedure. The instrument was administered by students' teachers, with the exception of the grade 7 sample which was tested by the school principal in the presence of the teacher. Students were told that the results would not count toward their final grade and that findings would be used for research. No time limits were assigned; most students finished in less than 20 minutes.

The responses were analyzed in a 4-step procedure resulting in the development of a multidimensional profile of growth:

1. All the responses were read by the authors to obtain a general impression of the data. A subsample (three from each grade) was selected to be representative of each grade and of the range in the total sample. The subsample responses were roughly ordered from poorest to best using intuitive criteria.

2. The subsample responses were examined (beginning with a comparison of the highest and lowest) to identify differences in performance. For each dimension the responses in the subsample were arranged as levels in a hierarchy. At this stage the investigators tried to be as inclusive as possible. For some dimensions there appeared to be levels that were logically compelling even though they had not explicitly appeared in the responses; these were added to the developing hierarchies in italics to differentiate them from observed responses.

3. The remaining responses in the total sample were used to test the adequacy of the profiles developed from the subsample. Dimensions were added, deleted and revised and intermediary levels were added from the larger set of data. Levels that were logical but not observed were dropped. The schemes were reduced in complexity by selecting only the most frequently occurring levels and by combining dimensions and levels whenever possible.

4. The profiles were reviewed by testing out specific hypotheses about growth derived from previous studies. For example, responses that included graphs were examined in terms of the patterns of student errors in constructing graphs identified by Wavering (1989).

Results and Discussion

The analysis of the first data collection resulted in an initial conception of growth that provided a series of levels within dimensions describing increasingly sophisticated responses to the task. Since these profiles were revised in subsequent data collections only an example will be given here. The final versions are given in Figure 4 at the end of Study 3.

Figure 1 gives an example of one dimension--how data were displayed or organized. The response labelled in the figure as level 1 (no data display) was the modal response for grade 4, although it appeared frequently in other grades. The next level (concrete pictures) was found in grade 4 and very rarely elsewhere. Tabular displays, the example given in the figure is a less sophisticated version, began to appear in grade 7, with more professional forms involving data summaries (averages) and control variables becoming visible in grade 10. Bar graphs were used by some grade 10 and OAC students. The scatter plots, with and without best fit lines, were virtually limited to university students.

Figure 1 About Here

The analysis revealed several deficiencies in the data collection: It was observed that not all students followed the directions in the prescribed order: many examined the cases prior to making a prediction. Many did not draw an explicit conclusion stating how the variables in the task were related, possibly because the item did not overtly call for a conclusion. They provided charts, graphs, pictures and calculations, but it was difficult to determine whether they believed the variables in the task were related and, if so, how. There was also doubt in some responses as to whether the effect of the control variable had been examined, because the relationship between age and swimming speed in the stimulus data was virtually the same for boys as it was for girls. In addition, some adjustment of the multi-grade sample was required because a large proportion of the grade 4 students submitted blank responses and expressed anxiety to their teachers about their inability to address the task. Finally, there was concern that because only a single task was administered the profiles might be too contextualized to the surface content of the task.

Study 2

The specific purpose of Study 2 was to refine the developmental profiles constructed in Study 1 by reducing some of the deficiencies in the data collection technique.

Method

Subjects. The subjects were multi-age convenience samples who were not involved in the first data collection, but were drawn from the same school systems. There were 20 respondents (half male, half female) in each of grades 5, 7, 9, 11, OAC and adults enrolled in a Master's of Education course on program evaluation.

Instruments. The swimming instrument was revised. Students made a prediction on the first page ("Are taller students faster swimmers than shorter students? Write a sentence that says how [or if] height is related to swimming speed.") and this page was collected. The second page containing the task situation and the raw data for the problem was then distributed. Students were asked to "Find out if there was a relationship between the height of students and how fast they swim. Use only the information in the table. Show your work." Then they were asked to "Write a sentence that says how (or if) height is related to swimming speed." The data provided in the stimulus were reworked so that the correct interpretation would find that height and swimming speed were positively related for boys, but not for girls. A second version of the instrument was created having exactly the same form and data distribution; the wording of the directions was identical except for the variable names. The content of the task was a hockey problem; the correct interpretation of the stimulus data would find that height and skating speed were positively related for old players, but not for new ones.

Procedure. The instruments were randomly assigned such that each was completed by 10 subjects in each grade using the same administration procedures as in Study 1. The responses were independently coded by the two principal investigators using the performance hierarchies developed in Study 1 and differences between coders were resolved through discussion.

Results and Discussion

Inter-rater reliability ranged from .72 in the grade 7 sample to .87 in the graduate student sample. Resolution of these differences and the appearance of new responses resulted in the revision of the profiles. No new dimensions emerged, but there was considerable reworking of levels within existing dimensions. Figure 2 contains an example of the Figure 1 dimension as revised by the new data: levels 1/2, 3/4 and 5/6 of the old scheme were combined; a new level focussing on the isolation of the control variable was constructed and the level labels were revised to emphasize the deep structure rather than surface features of the representations.

Figure 2 About Here

Mean scores on each of the dimensions indicated that some dimensions emerged more quickly and more completely than others, a finding compatible with the results from Study 1. None of the subjects in the first or second data collection (a total of 220 respondents, including 80 at university or in the university prep year) met the expectations set out for students at the end of grade 12 in the provincial curriculum guidelines (Ontario Ministry of Education, 1988). Even so, these scores may have been inflated by students who came up with correct or partially correct conclusions even though they had no apparent strategy for interpreting the data. The data provided in the instruments may have corresponded with some students' beliefs about height and athletic performance. Previous researchers (e.g., Chapman & Chapman, 1982; Jennings, Amabile & Ross, 1982) have found that subjects' pre-existing theories about relationships strongly influence whether they detect a correlation in a body of data.

Study 3

The specific purposes of Study 3 were to address concerns that arose in the second data collection, consolidate the developmental profiles and test the utility of the profiles with a sample of teachers.

Figure 3 About Here

Method

Subjects. The subjects were multi-age convenience samples who were not involved in earlier data collections but were from the same school boards. There were 20 respondents (half male, half female) in each of grades 5, 7, 9, 11 and OAC.

Instruments. A new instrument was produced to deal with students who were able to generate a correct conclusion from their belief systems. In Part A of the new instrument subjects were asked to predict whether the amount of vaccine administered to students is related to the number of flu cases in those schools. The data presented in Part B (shown in Figure 3) supported a counter-intuitive finding (the vaccine increased the likelihood of getting the flu) that became visible only when students controlled for a third variable (the form in which the vaccine was delivered): The number of cases of flu increased with the amount of vaccine given in pill form; in needle form there was no relationship between the amount of vaccine and the number of flu cases. In Part C (also shown in Figure 3), subjects were asked to draw a conclusion. Part A was collected before Parts B and C were distributed. The swimming instrument was revised to produce a similar finding: height was negatively related to swimming speed for girls and was unrelated to swimming speed for boys. The hockey instrument was not used.

Procedures. The instruments were randomly assigned so that each was completed by 10 subjects in each grade, using the same administration procedures as in Study 1. The responses were read and minor changes were made to the profiles. The responses were randomly assigned to 10 packages of vaccine items and 10 packages of swimming items. Each package contained one response from each of the five grades, arranged in random sequence. Grade 7 Geography teachers ($\underline{n} = 16$) received a short in-service in which the growth schemes for the correlational reasoning skills were represented as a marking scheme. After practice sessions using four of the packages, each teacher independently coded two more packages, one for each instrument; 16 packages were each coded independently by two teachers.

Means for each skill dimension were calculated and polynomial contrasts were used to assess the effect of age on performance.

Results and Discussion

The review of responses led to only minor changes in the profiles. For example, in the organizing information dimension (in Figure 2), level 4 concerned with exposing the influence of control variables was split into two levels: ordering on one variable with a control (new level

4) and ordering on two variables with a control (new level 5). The final version of the developmental profiles is shown in Figure 4 in the marking scheme format that was used in Study 3.

Figure 4 About Here

The proportion of exact agreement between pairs of teachers ranged from .58 (for concluding) to .80 (for locating data); the overall inter-rater reliability was .72. Examination of the disagreements revealed no consistent patterns and there were no responses that led to the modification of the profiles in Figure 4. The low reliability between teachers was attributed to the coders' insufficient knowledge of correlational reasoning; for example, it was clear that some had not grasped the concept of controlling variables. None had taught the skill and for most the workshop was their first introduction to this type of problem. Also, the generally low performance of the students meant that markers rarely saw examples of higher skill levels. In addition, the grade 7 teachers may have been unfamiliar with the expressions of students in grades other than their own.

Figure 4 About Here

Since there were no sex differences in performance, the responses of males and females were pooled. Table 2 displays the means and standard deviations for each grade on each skill; in this table the scores from the two items were averaged.

Table 2 About Here

The polynomial trend analysis showed that there were significant differences between age groups for the total score [$F(4,75) = 4.16, p<.004$] and for three of the four indicators: organizing information [$F(4,75) = 3.71, p<.008$], synthesizing data [$F(4,75) = 2.64, p<.040$] and concluding [$F(4,75) = 3.61, p<.010$]. There was a significant linear trend overall [$F = 6.29, p<.014$], but there was also a significant deviation from linearity [$F = 3.44, p<.013$] and there was a significant cubic trend [$F = 6.47, p<.013$]. The linear trend in the total scores came from the organizing and concluding indicators; the cubic trend came from the synthesizing and locating skill. There was also a quadratic term in the concluding skill.

Study 4

The specific purposes of Study 4 were to determine whether students' ability to solve multivariate correlational problems could be improved through modest amounts of instruction and if so, whether the instruments on which the developmental profiles were based could detect the impact.

Method

Subjects. Subjects were a convenience sample of grade 7 students. Students in 12 intact classes constituted the treatment group and were matched with a no-treatment control group of 12 classes. None of the teachers ($n = 24$) had taught correlational reasoning before and none of the students had previously been exposed to the materials. All teachers were experienced grade 7 Geography teachers from the four school districts which participated in earlier phases of the research. All teachers were volunteers.

Instruments. Student achievement was measured with two forms of the correlational reasoning instruments described in Study 3.

Procedures. Innovative instruction took the form of a treatment delivered by regular classroom teachers over nine 40-minute periods. The instructional strategy was a "rules provided" method (Authors, 1985) that has produced substantial student achievement effects in other problem solving programs (Authors, 1983a; 1983b; 1984; 1986; 1987; 1988a; 1988b; in press). Treatment teachers were given a half day in-service and detailed instructional materials.

Half the students received the medicine item as the pretest followed by the swimming item as the postest; for the remainder the order was reversed. A random sample of half the student responses was selected from each teacher; if there were less than 20 complete data sets in the class, 10 were randomly chosen. The responses ($n = 266$) were coded by a single trained tester using the scheme in Figure 4. Two random samples of items were independently coded by one of the principal investigators and additional random samples were independently coded by experienced teachers.

Student scores were aggregated to the class level and an analysis of covariance using multiple regression was conducted. Total scores and scores on each dimension were the criterion variables; pretest (covariate) and instructional treatment were the predictor variables.

Results and Discussion

The inter-rater reliability (proportion of exact agreement) between the principal investigator and trained tester was .88 on the first sample of items ($n = 48$) and .92 on the second set ($n = 24$ items). The inter-rater reliability between teachers and the trained tester averaged .81 ($n = 236$ items). These findings underscore the importance of the markers' knowledge of correlational reasoning.

The means and standard deviations are displayed in Table 3. Pretest performance was extremely low with a large number of students receiving zero, especially on the skills of organizing and synthesizing information, as found in Studies 1, 2 and 3. The posttest scores were much higher for the treatment group, but not for the control. The posttest mean for the treatment condition reached 83% and 90% of the maximum scores for locating and synthesizing information respectively, but students obtained only 56% and 45% of the maximum possible for organizing information and concluding.

The source table from the regression analysis using the total scores is displayed in Table 4. It shows that the instruments effectively distinguished posttest performance of instructed from uninstructed students, once variability associated with pretest performance had been accounted for. The regression equations for each of the dimensions (not reported) showed the same pattern.

Tables 3 and 4 About Here

General Discussion

One of the purposes of the research was to develop a technique for measuring students' correlational reasoning skills. The instruments that were produced have several attractive features: They are easily administered, taking about 15 minutes of student time to complete. They embody tasks that are comparable to the correlational problems encountered by students on a daily basis; that is, the task is multivariate; continuous as well as dichotomous variables are involved; subjects are required to exclude as well as include relationships, and the test content is familiar to most students. The instruments effectively discriminate instructed from uninstructed students and distinguish older from younger respondents.

The instruments also have some undesirable features. The tests were so difficult for the subjects in the four data collections reported in this paper that it would be unwise to use them in the Primary or Junior Divisions; even in the Intermediate Division floor effects are likely to be a problem. A possible solution might be to supplement these instruments with simpler tasks, for example, the 2 X 2 contingency table problems developed by Shaklee & Mims (1981). The inter-rater reliability of the instruments developed in the project is also a source of concern, particularly when the markers are classroom teachers with relatively little experience in teaching correlational reasoning. The level of agreement was higher between a trained tester and one of the principal investigators, but even then the level reached (89% perfect agreement) was below the reliability recorded with other problem solving instruments (e.g., the low to high 90's in Authors, 1988a).

Another purpose of the research was to identify patterns of student growth under conditions of existing instruction. Although the findings were consistent for each of the samples, the credibility of the findings is limited by the use of convenience samples drawn from four school districts in the same geographic area. It should not be assumed that the same patterns will be found in cross-sectional studies conducted in other jurisdictions.

The profiles of correlational reasoning that were developed from problems involving multiple continuous variables in these studies differed from the profiles that emerged from earlier research into problems involving two dichotomous variables. The profiles in our research provide separate heirarchies for each dimension rather than combining them into a single strand; these levels can be combined in alternate ways (although some combinations are intuitively unlikely) to accommodate broad variation in student performance. Our profiles also give overt attention to problem representations, providing a scale for assessing the complexity of subjects' frameworks for organizing correlational data. But the similarities between our schemes and the patterns of growth observed by previous researchers are greater than the differences.

To what can these similarities be attributed? The emergence of neo-Piagetian paradigms in response to empirical and theoretical developments in cognitive psychology provide a bridge between our conceptions and earlier reports of growth in correlational reasoning. Biggs & Collis (1982) propose a scheme for interpreting open ended responses to cognitive tasks that is especially helpful. The Structure of Observed Learning Outcomes (SOLO) analyses responses in terms of four dimensions.

1. *Capacity* in SOLO is concerned with the amount of working memory required; at the highest level it is the subject's ability to consider interactions among the stimulus, pieces of data and subject hypotheses. This dimension is comparable to the attention given to subjects' consideration of the mediating effects of control variables in the higher levels of our "organizing informat[illegible] dimension and to the concern with interactions among the four cells of the 2 X 2 table in profiles derived from bivariate dichotomous problems.

2. *Relating operations* in SOLO is concerned with drawing conclusions based on the interaction of the stimulus and the data, specifically with the number of different aspects of the data and the interactions among these aspects that are considered by the subject. This dimension is congruent with the attention given to two- and three-variable relationships in our "concluding" dimension and to the consideration of confirming and disconfirming evidence in profiles based on 2 X 2 problems.

3. *Consistency and closure* in SOLO balances the desire to come to closure with the desire to accommodate variation in the data; specifically, it focuses on the amount of data considered by the problem solver. In our profiles this concern relates to the dimensions "locating or selecting data" and "synthesizing information". The same attention is given in profiles developed by previous researchers which emphasize the extent to which subjects examine data in all four cells of 2 X 2 correlational problems.

4. *Structure* in SOLO involves representing the three dimensions above, but the examples given by Biggs & Collis are particularly prone to the objection that this attribute of performance may be in the mind of the observer rather than the mind of the subject. In contrast, Case & Griffin (in press) propose that there are central conceptual structures

underlying cognitive performance and that these can be represented sufficiently concretely that they correspond to what is in children's minds--examples are provided. The concern with representation is addressed in our dimension "organizing information", but it has been treated only implicitly in previous conceptions of growth in correlational reasoning based on 2 X 2 problems.

There are grounds for thinking that there is a central correlational schema underlying performance on the multiple, continuous variable correlational problems examined in Studies 1-4 and the bivariate, dichotomous problems examined by prevous investigators. What needs to be addressed in future research is the scope of the schema: Is it limited to the domain of correlational problems or is it as broad in its application as the central conceptual structures for logico-mathematical reasoning and for social cognition that Case and his colleagues are reporting (in e.g., Case & Griffin, in press) in their studies of intellectual growth in children aged 4 to 10?

Instructional Implications

The results in Studies 1, 2 and 3 showed an age trend in which growth was slow and not continuous. The final levels reached were far below the expectations of curriculum guidelines governing the schools in which the research was conducted. This finding provides further evidence of the need for a fresh approach to teaching correlational reasoning: current practice is insufficient to prepare students for the correlational problem solving tasks they will encounter in school and in out-of-school settings. The performance of the Master's of Education sample, made up of experienced teachers with undergraduate records sufficiently impressive to gain admittance to a demanding academic program, evokes concern that the student need may not be addressed. This group scored no better than secondary school students, indicating that there is a deficiency in subject content knowledge that is likely to impede instructional effectiveness.

Despite these concerns, Study 4 shows that the performance of grade 7 students can be considerably improved through an instructional intervention delivered by regular classroom teachers who are given appropriate support in the form of detailed curriculum materials and in-service training. Inquiry into instructional interventions will be the main focus of the next phase of our research on correlational reasoning. The current directions include (a) replicating the instructional improvement experiment in other grades, beginning in grade 9, (b) identifying conditions (of the student and of instruction) that influence learning of correlational reasoning skills and (c) exploring instructional treatments, particularly cooperative learning strategies and approaches involving the use of the computer as a tool for reducing the tedium of plotting data. Our thinking about instructional interventions is based upon our earlier research on problem solving in school domains and is particularly stimulated by the suggestion of Case & McKeough that higher levels of intellectual performance are acquired via socially facilitated processes and are directly teachable:

> ...children are seen as re-constructing the conceptual inventions of prior generations with the aid of the current generation, rather than as abstracting universal logical invariants from their own epistemic activity [in press: 23].

References

Adi, H., Karplus, R., Lawson, A., & Pulos, S. (1978). Intellectual development beyond elementary school VI: Correlational reasoning. *School Science and Mathematics, 78* (8), 675-683.

Baird, W., & Borich, G. (1987). Validity considerations for research on integrated science process skills and formal reasoning ability. *Science Education, 71* (2), 259-269.

Biggs, J., & Collis, K. (1982). Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome). New York, NY: Academic.

Case, R., & Griffin, S. (in press). Child cognitive development: The role of central conceptual structures in the development of scientific and social thought. In C. Hauert (Ed.), Advances in psychology - Developmental psychology: Cognitive, perceptuo-motor and neurological perspectives.

Case, R., & McKeough, A. (in press). Schooling and the development of central conceptual structures: An example from the domain of children's narrative. In F. Weiner (Ed.), Recent research on schooling and cognitive development.

Chapman, L., & Chapman, J. (1982). Test results are what you think they are. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under conditions of uncertainy: Heuristics and biases*, Cambridge: Cambridge University Press.

Curtis, J. (1985). Teaching college biology students the simple linear regression model using an interactive microcomputer graphics software program. Unpublished PhD disertation, University of Wisconsin-Madison. UMI 85-12296.

Donnelly, J., & Welford, A. (1989). Assessing pupils' ability to generalize. *International Journal of Science Education, 11* (2), 161-171.

Gal, I., Rothschild, K., & Wagner, K. (1990). Statistical concepts and statistical reasoning in school children: Convergence or divergence. Paper presented at the Annual Meeting of the American Educational Research Association, Boston.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York, NY: Basic.

Jenkins, H., & Ward, W. (1965). Judgment of contingency between responses and outcomes. *Pschological Monographs, 79* (1), 1-17.

Jennings, D., Amabile, T. & Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under conditions of uncertainy: Heurstics and biases*, Cambridge: Cambridge University Press, pp. 211-230.

Koslowski, B., & Okagaki, L. (1986). Non-Humean indices of causation in problem-solving situations: Causal mechanism, analogous effects and the status of rival alternative accounts. *Child Development, 57*, 1100-1108.

Kuhn, D., & Brannock, J. (1977). Development of the isolation of variables scheme in experimental and "natural experiment" contexts. *Developmental Psychology, 13* (1), 9-14.

Kuhn, D., & Ho, V. (1977). The development of schemes for recognizing additive and alternative effects in a "natural experiment" context. *Developmental Psychology, 13* (5), 515-516.

Lawson, A. (1982). The relative responsiveness of concrete operational seventh grade and college students to science instruction. *Journal of Research in Science Teaching, 19* (1), 63-77.

Lawson, A., Adi, H., & Karplus, R. (1979). Development of correlational reasoning in secondary schools: Do biology courses make a difference? *American Biology Teacher, 41* (7), 420-425, 430.

Lawson, A., & Bealer, J. (1984). The acquisition of basic quantitative reasoning skills during adolescence: Learning or development. *Journal of Research in Science Teaching, 21* (4), 417-423.

Lovell, K. (1961). A follow-up study of Inhelder and Piaget's "The growth of logical thinking." *British Journal of Psychology, 52* (2), 143-153.

McKenzie, D. & Padilla, M. (1986). The construction and validation of the test of graphing in science (TOGS). *Journal of Research in Science Teaching, 23* (7), 571-579.

Niemark, E. (1975). Longitudinal development of formal operations thought. *Genetic Psychology Monographs, 91*, 175-225.

Ontario Ministry of Educatic~ (1988). *Intermediate and Senior Geography Curriculum Guidelines*. Toronto: Queen's Park.

Padilla, M., McKenzie, D., & Shaw, E. (1986). An examination of line graphing ability of students in grades seven through twelve. *School Science and Mathematics, 86* (1), 20-26.

Ross, L. & Lepper, M. (1980). The perseverance of beliefs: Empirical and normative considerations. In R. Sweder (Ed.), *Fallible judgment in behavioural research*, San Francisco, CA: Jossey-Bass, 17-36.

Roth, W. (1989). Confirmatory factor analysis for validity considerations: A critique. *Science Education, 73* (6), 649-655.

Shaklee, H., & Mims, M. (1981). Development of rule use in judgments of covariation between events. *Child Development, 52*, 317-325.

Shaklee, H., & Paszek, D. (1981) *Eliciting systematic rule use in covariation judgment: The early years*. Washington, DC: National Institute of Education.

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology, 4*, 165-173.

Wavering, M. (1984). Interrelationships among Piaget's formal operations schemata: Proportions, probability, and correlation. *Journal of Psychology, 118* (1), 57-64.

Wavering, M. (1989). Logical reasoning necessary to make line graphs. *Journal of Research in Science Teaching, 26* (5), 373-379.

Yap, K., & Yeany, R. (1988). Validation of hierarchical relationships among Piagetian cognitive modes and integrated science process skills for different cognitive reasoning levels. *Journal of Research in Science Teaching, 25* (4), 247-281.

TABLE 1

Studies of Correlational Reasoning Using Multi-Age Samples

Study	Sample	Instrument	Number of Variables	Type of Variables	Definition of Growth	Coding Scheme Derived from Data
Adi et al. (1978)	n=80, gr.9 & 12	interview	two	dichotomous	levels	yes
*Curtis (1985)	n=108 undergrads, 31 grads, 16 faculty	open ended	two	continuous	levels	possibly
Inhelder & Piaget (1958)	not reported	interview	two	dichotomous	levels	yes
Kuhn & Brannock (1977)	n=80, gr.4,5,6, university	interview	four	dichotomous	levels	yes
Kuhn & Ho (1977)	n=100, gr.4,6,8,12, university	interview	four	dichotomous independent variables; continuous dependent variables	levels	yes
*Lawson (1982)	n=122, gr.7, college	open ended	two	dichotomous & continuous	criterion	no
Lawson et al. (1979)	n=507, gr.6,8,10,12, university	open ended	two	dichotomous	levels	probably
Lawson & Bealer (1984)	n=391, gr.6,8,10,12	open ended	two	dichotomous	criterion	no
Lovell (1961)	n=68, primary, secondary	interview	two	dichotomous	levels	no
McKenzie & Padilla (1986)	n=377, gr.7-12	multiple choice	two	continuous	criterion	no
Niemark (1975)	n=66, gr.4-6**	open ended	two	dichotomous	criterion	no
Padilla, McKenzie & Shaw (1986)	n=625, gr.7-12	multiple choice	two	continuous	criterion	no
Shaklee & Mims (1981)	n=103, gr.4,7,10, college	multiple choice	two	dichotomous	levels	no
*Shaklee & Paszek (1981)	n=72, gr.1-4	interview	two	dichotomous	levels	no
Wavering (1984)	n=94, gr.8,10,12	interview	two	dichotomous	levels	no
Wavering (1989)	n=150, gr.7-12	open ended	two	continuous	levels	yes
Authors (1985)	n=35, gr.4,6,8,10, adult	interview	two	continuous	levels	yes
Authors (1981)	n=1680, gr.7 & 10	multiple choice	two	continuous	levels	no

* reports an instructional intervention
** students were followed for 3 years; at the end of the study they were in grades 7-9

Table 2

Means and Standard Deviation for Correlational Reasoning Skills

From Study 3

	X Level for Each Skill							
	Organizing (0-4)		Locating (0-2)		Synthesizing (0-3)		Concluding (0-3)	
Grade	X	SD	X	SD	X	SD	X	SD
5 (n=16)	.00	.00	.06	.17	.06	.25	.06	.17
7 (n=16)	.22	.36	.41	.69	.09	.27	.31	.51
9 (n=16)	.06	.17	.38	.47	.22	.36	.28	.36
11 (n=16)	1.19	.68	1.56	.60	.63	.43	.59	.58
13 (n=16)	1.19	1.29	1.19	.87	.56	.48	1.00	1.13

Table 3

Means and Standard Deviations for Correlational Reasoning Skills

in Study 4

	Treatment (n=12)		Control (n=12)		Maximum
	Pre	Post	Pre	Post	Score
Total	1.57 (1.90)	7.86 (2.77)	1.51 (1.83)	2.09 (2.33)	12
Organize	0.25 (0.69)	2.25 (1.18)	0.22 (0.61)	0.42 (0.82)	4
Locate	0.49 (0.75)	1.79 (0.58)	0.52 (0.81)	0.75 (0.90)	2
Synthesize	0.10 (0.41)	2.48 (0.99)	0.04 (0.28)	0.23 (0.62)	3
Conclude	0.73 (0.89)	1.34 (0.98)	0.74 (0.90)	0.68 (0.87)	3

Table 4

Source Table From Regression Model for Total Scores

in Study 4

Predictor	ss	df	MS	F	p<
1. Pretest	246.08	1	0.45	0.04	ns
Error	246.08	22	11.19		
2. Treatment	193.35	1	193.35	77.03	.001
Error	52.73	21	2.51		

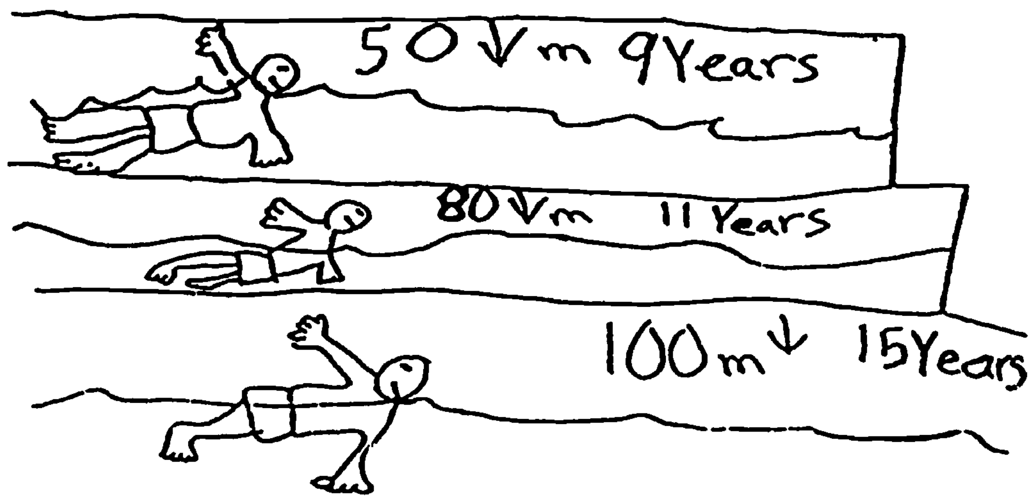
<table>
<tr><th>Level</th><th>Example: Representing data for the swimming problem</th></tr>
<tr><td>1. No Data Display</td><td>The original list of cases is left untouched</td></tr>
<tr><td>2. Concrete Picture</td><td>
</td></tr>
<tr><td>3. Tabular Display</td><td>
<table>
<tr><th>Sex</th><th>Name</th><th>Age</th><th>Distance</th></tr>
<tr><td>Girl</td><td>Sue</td><td>10</td><td>77</td></tr>
<tr><td>Girl</td><td>Karen</td><td>9</td><td>71</td></tr>
<tr><td>Boy</td><td>Mel</td><td>10</td><td>65</td></tr>
<tr><td>Boy</td><td>Lincoln</td><td>10</td><td>69</td></tr>
<tr><td>Girl</td><td>Janice</td><td>11</td><td>68</td></tr>
<tr><td>Girl</td><td>Carolyn</td><td>11</td><td>63</td></tr>
<tr><td>Boy</td><td>Bob</td><td>11</td><td>79</td></tr>
<tr><td>Boy</td><td>Colin</td><td>12</td><td>90</td></tr>
<tr><td>Boy</td><td>Jeff</td><td>12</td><td>88</td></tr>
<tr><td>Girl</td><td>Cathy</td><td>12</td><td>68</td></tr>
<tr><td>Girl</td><td>Sarah</td><td>12</td><td>70</td></tr>
<tr><td>Boy</td><td>Mike</td><td>13</td><td>95</td></tr>
<tr><td>Girl</td><td>Judy</td><td>13</td><td>82</td></tr>
<tr><td>Girl</td><td>Helen</td><td>13</td><td>61</td></tr>
</table>
</td></tr>
</table>

Figure 1: Levels of growth in organizing information: In Study 1

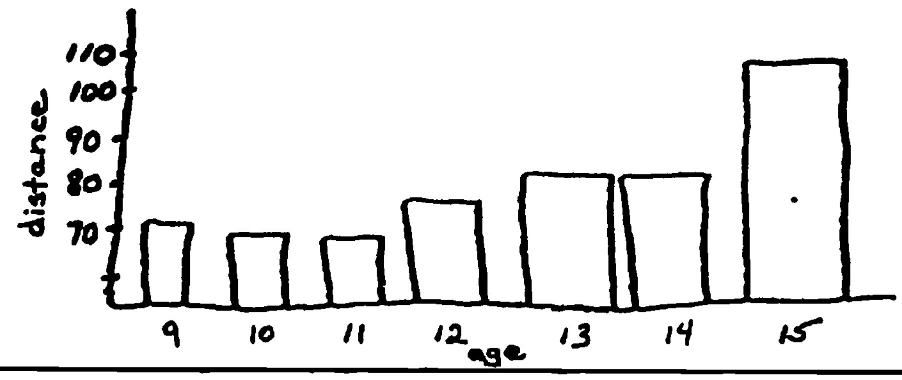
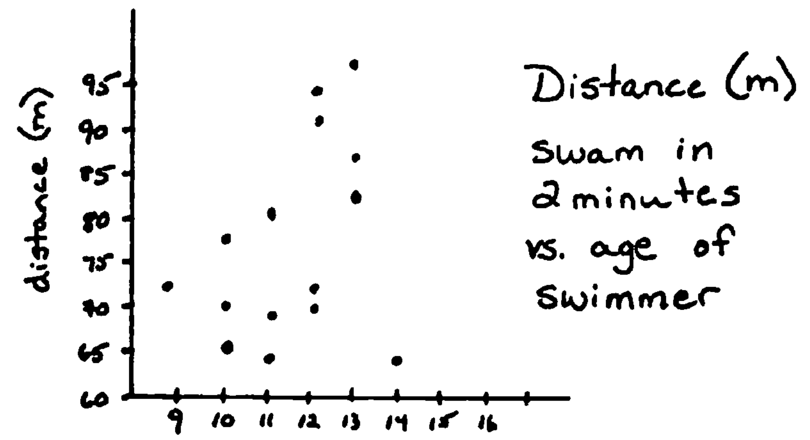
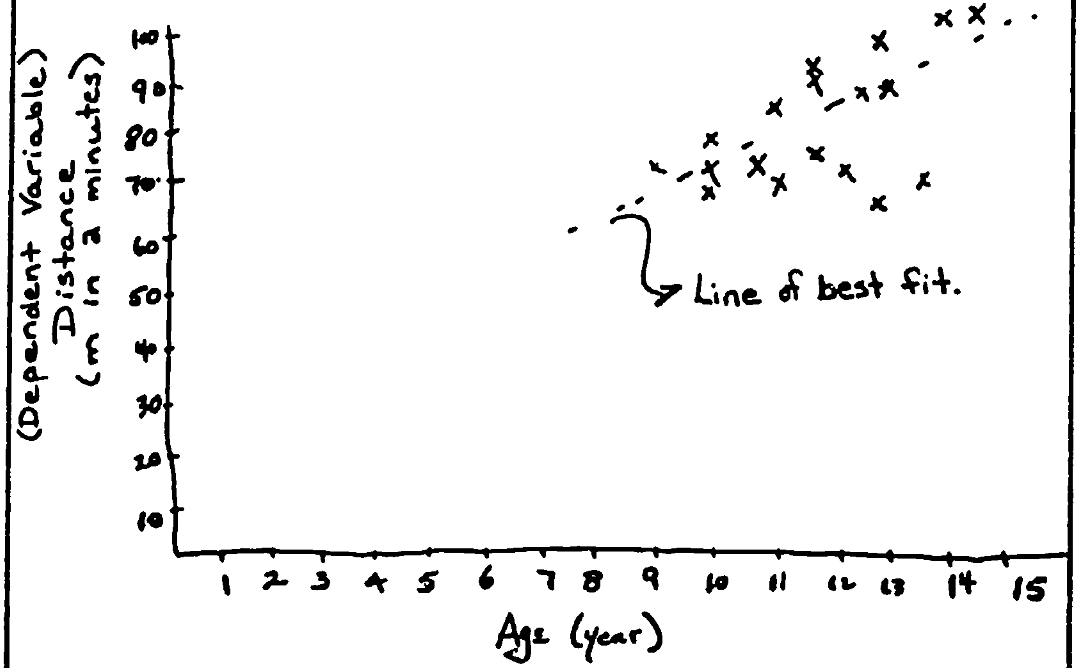
Level	Example: Representing data for the swimming problem
4. Bar Graph	
5. Scatterplot	
6. Scatterplot with best fit line	A statistical correlation may be applied to the above data, however, I do not recall the formula. Also, the information may be graphed. 

Figure 1 continued

Growth in Organizing a Correlational Problem

To organize a correlational problem is to arrange the information in a structure that makes it possible to find relationships between variables.

1. No Productive Display

   - the student does not rearrange the data or does so in an unproductive way; (e.g., the student might calculate the average of all the swimmers and the average distance swum by the whole group) or the student might draw a picture showing the superficial features of the problem (e.g., a picture of some students swimming)

2. Ordering on One Variable

   - the cases are arranged from lowest to highest on one variable; e.g., a table listing the swimmers from youngest to oldest; the table could also contain groups of swimmers rather than individual cases

3. Ordering on Two Variables

   - the cases are arranged from lowest to highest on both variables; e.g., in a line graph, a scatter plot or a bar graph ordered horizontally as well as vertically

4. Displays Exposing the Influence of Control Variables

   - the cases are arranged to expose the influence of a control variable on the relationship; e.g., a graph of age and swimming distance containing a trend line for girls and another line for boys

Figure 2: Example of a growth scheme for a correlational reasoning skill from Study 2

Name ____________________

QUESTION B

Find out if there was a relationship between amount of medicine and number of cases of flu. Use only the information in the table. Show your work.

One winter there was a lot of flu. All the students were given the same medicine to protect them from getting it. The amount of medicine given to the students was different in each school. Some students got the flu and some didn't.

School	How the Medicine was Given	Amount Given (in .00 ml)	Number of Flu Cases (per 100)
Alma	pill	0.20	14
Appleby	needle	0.28	25
Bayside	pill	0.60	16
Brant	needle	0.45	38
Central	needle	1.10	29
Churchill	pill	0.90	37
Donlands	needle	0.88	20
Drayton	pill	0.35	8
Edwards	pill	0.65	24
Elgin	pill	0.10	5
Franklin	needle	0.22	32
Frontenac	needle	0.58	18
George VI	pill	0.80	32
Halton	needle	1.10	36
Independence	needle	0.62	28
Jeffries	pill	0.45	22
Knowlton	pill	0.38	18
Lawrence	needle	0.77	37

QUESTION C: Write a sentence that says how (or if) the amount of medicine is related to whether or not you get the flu.

____

____

____

Figure 3: Example of correlational reasoning test item

1. ORGANIZING INFORMATION	
(a) Did the student have <u>no</u> productive display (i.e., no rearrangement of the data)?	0 marks
(b) Did the student have a display in which the data were rearranged from lowest to highest on 1 variable?	1 mark
(c) Did the student have a graph in which the data were rearranged simultaneously from lowest to highest on 2 variables?	2 marks
(d) Did the student have a display with the data rearranged from lowest to highest on one variable, and a control variable?	3 marks
(e) Did the student have a graph with the data rearranged from lowest to highest on 2 variables, and a control variable?	4 marks
2. LOCATING OR SELECTING DATA	
(a) Did the student randomly (or unsystematically) select a few (e.g., 3 to 4) cases?	0 marks
(b) Did the student systematically select a few (e.g., 3 to 4) strategic cases?	1 mark
(c) Did the student select enough cases (12 or more) to make a decision as to whether or not there is a relationship?	2 marks
3. SYNTHESIZING INFORMATION	
(a) Did the student make no attempt to summarize?	0 marks
(b) Did the student use a simple summary strategy (e.g., averages)?	1 mark
(c) Did the student put the data in a graph, without a trend line?	2 marks
(d) Did the student use a nontechnical way to represent the trend (e.g., draw a trend line on a graph)?	3 marks
4. DRAWING A CONCLUSION	
(a) Did the student fail to describe a correct relationship?	0 marks
(b) Did the student describe a partially correct relationship? (e.g., if the student controlled variables, one relationship is correct as shown in part d; or, e.g., if variables are not controlled, a relationship that is correct only if unusual cases are ignored; i.e., height is negatively related to swimming; medicine increases flu)	1 mark
(c) Did the student describe a correct 2 variable relationship? (e.g., height and swimming are not related or, e.g., medicine and flu cases are not related)	2 marks
(d) Did the student describe a correct 3 variable relationship? (e.g., height is negatively related to swimming for girls and height is not related to swimming for boys; or, e.g., medicine is positively related to the flu for the pill and medicine is not related to the flu for the needle)	3 marks

Figure 4: Marking scheme for correlational skills