DOCUMENT RESUME

ED 238 707 SE 043 683

AUTHOR Shaklee, Harriet; Hall, Laurie
TITLE Methods of Assessing Strategies for Judging Covariation between Events.
INSTITUTION Iowa Univ., Iowa City. Dept. of Psychology.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE [83]
CONTRACT NIE-G-80-0091
NOTE 36p.; Appendix A of SE 043 682.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cognitive Processes; Educational Research; *Evaluative Thinking; Higher Education; Individual Differences; *Learning; Mathematical Concepts; *Mathematics Education; Prediction; *Probability; Problem Solving; *Psychological Studies; Sex Differences
IDENTIFIERS *Mathematics Education Research; Rule Learning

ABSTRACT

Past research indicates poor agreement about strategies people use to assess covariation between events. This research investigates method of assessment as one possible source of this low consensus. A set of problems was developed in such a way that different judgment rules would produce different decisions about the relationships between events. College subjects judged these problems, then were asked to explain their judgment strategy. In addition, they were shown model strategies and asked to choose the one like their own strategy and the model that would be the best strategy. Subjects whose judgments indicated use of the most sophisticated strategy were quite accurate in reporting their judgment rules. Subjects using the less accurate rules most commonly reported using strategies which could not have produced the obtained pattern of problem solutions. These findings suggest that self-report is a weak basis for conclusions about sources of error in covariation judgment. (Author)

Methods of Assessing Strategies for Judging Covariation Between Events

Harriet Shaklee and Laurie Hall

University of Iowa

Running head: Judging Event Covariations

in press, Journal of Educational Psychology

2

Abstract

Past research indicates poor agreement about strategies people use to assess covariation between events. This research investigates method of assessment as one possible source of this low consensus. A set of problems was developed in such a way that different judgment rules would produce different decisions about the relationships between events. College subjects judged these problems, then were asked to explain their judgment strategy. In addition, they were shown model strategies and asked to choose the one like their own strategy and the model that would be the best strategy. Subjects whose judgments indicated use of the most sophisticated strategy were quite accurate in reporting their judgment rules. Subjects using the less accurate rules most commonly reported using strategies which could not have produced the obtained pattern of problem solutions. These findings suggest that self-report is a weak basis for conclusions about sources of error in covariation judgment.

Children's Judgments about Covariation between Events: A Series of Training Studies. Appendix A .

Statistical concepts represent one prime area for application of mathematical training. In particular, statistics are necessary for identifying predictability in an environment where relationships are frequently probabilistic (y is more likely when x is present) rather than deterministic (y always occurs when x is present). Problems such as these are common in identifying regularities in scientific phenomena, and in everyday contexts as well. In this respect, statistics provide a key link between basic mathematical concepts and central aspects of scientific and everyday problem solving. As an area for application of mathematical training, research on statistical reasoning may also be informative about children's and adult's abilities to apply their mathematical skills appropriately.

The focus of existing research in this area has been on probability judgments (e.g., Piaget & Inhelder, 1975; Fischbein, 1975; Yost, Siegel & Andrews, 1962). A statistical judgment common to reasoning about cause-effect relationships builds on probability assessments of this sort. An individual investigating the relationship between potential cause x and effect y would compare the likelihood of y occurring when x is present $P(y/x)$ with the likelihood that y occurs without x $P(y/\bar{x})$. The two events are independent if these conditional probabilities are equal; nonindependence is indicated by any difference. The comparison is made to identify contingency or covariation between events. Scientific procedure and statistical analyses testify to the key role of covariation analysis in professional practice. Although not sufficient for causal inference, covariation is a necessary condition between causes and events. Thus, covariation analysis may identify the set of possible causes of an event.

Many psychologists further assert that everyday causal judgment is similarly based on a covariation analysis (e.g., Michotte, 1963: Inhelder & Piaget, 1958; Kelley, 1967; Heider, 1958). That is, people search for likely explanations of everyday events by identifying event covariates. Thus, competence in covariation judgment may determine a person's adequacy in identifying real world cause-effect relationships.

In fact, a variety of investigators have found that adolescent and adult subjects show little competence in identifying event covariations (Niemark, 1975; Smedslund, 1963; Jenkins & Ward, 1965; Adi, Karplus, Lawson, & Pulos, 1978). While the evidence indicates that covariation judgments are often erroneous, those judgments may be rule-governed nonetheless. Several different rules have been proposed by past investigators as possible judgment strategies. These rules are discussed in terms of possible relationships between two events (A and B), each of which occurs in one of two states (1 and 2).

Least sophisticated of the proposed strategies is judgment according to the frequency with which the target events cooccur (A_1B_1, cell a in a traditionally labeled contingency table) failing to consider the other event-state pairings (A_1B_2, A_2B_1, A_2B_2) in defining the relationship. A subject using this strategy would identify a positive relationship between A_1 and B_1 if cell a frequency were the largest of the contingency table cells, a negative relationship if it were the smallest (cell a strategy). This strategy is identified by Inhelder and Piaget (1958) as common among younger adolescents. Smedslund (1963) and Nisbett and Ross (1980) suggest that the strategy is typical among adults as well. The strategy does consider some relevant information and may result in better-than-chance performance. However, the rule considers only a limited portion of the information that defines the relationship and would result in erroneous judgment of many relationships.

A second possible approach would compare the number of times target events A_1 and B_1 cooccur with the times A_1 occurs with B_2 (comparison of frequencies in contingency table cells a and b; strategy a versus b). This strategy is also identified by Inhelder and Piaget (1958) as a precursor of mature judgment. Again this strategy considers some of the relevant information and may result in accurate judgment of many event contingencies. However, failure to consider frequencies in cells c and d (event combinations A_2B_1 and A_2B_2) would be a particularly costly error when the direction of that frequency difference is the same as the difference between cells a and b.

A much improved approach would be the strategy defined by Inhelder and Piaget (1958) as characteristic of formal operational thinking. Specifically, covariation would be defined by comparing frequencies of events confirming (cells a and d) and disconfirming (cells b and c) the relationship. Thus, the rule would compare the sums of the diagonal cells in a contingency table (sum of diagonals strategy). Jenkins and Ward (1965), however, point out that this strategy has its limits as well. Specifically, the rule is an effective index only when the two states of at least one of the variables occur equally often. Otherwise, a correlation may be indicated when, in fact, independence is the case.

Instead, Jenkins and Ward (1965) suggest that covariation is more appropriately evaluated by comparing the probability of event A_1 given event B_1 $P(A_1/B_1)$ with the probability of A_1 given that B_2 has occurred $P(A_1/B_2)$. This is equivalent to a comparison of the frequency ratio in contingency table cells $\frac{a}{a+c}$ with that in cells $\frac{b}{b+d}$. By definition, independence is indicated by equivalence between these conditional probabilities; nonindependence is indicated by any difference (conditional probability strategy). This strategy should result in accurate judgment of any contingency problem.

Thus, four alternative strategies have been proposed to account for subjects' judgment patterns. Many of these rules were proposed on the basis of subjects' explanations of their judgments. For example, Smedslund's (1963) cell a strategy is based on the reports of over half of his sample that they judged the relation of symptom A and diagnosis F according to the number of AF pairings. Adi, Karplus, Lawson, and Pulos (1978) similarly categorized subjects according to their explanations. In this case, however, no subjects were classified as using a cell a strategy. Rather, subjects described themselves as using various combinations of two to four of the contingency table cells. Thus, two samples of subjects offer considerably different explanations of their judgment strategies. Two features of these studies make it hard to reconcile these differences. First, the two reports offer little information on the way the explanations were elicited. We might expect that different questions would result in different responses. Secondly, neither of the investigators report the level of agreement with which subject responses were categorized, so we know little about the reliability of the categorization schemes.

However, a more serious problem is relevant to any explanation-based strategy analysis. That is, such an approach is predicated on the assumption that subjects are able and willing to accurately describe their bases of judgment. In fact, a variety of research in psychology suggests that this assumption may not be justified. In developmental research in particular, young children's poor verbal skills may hinder their account of systematic judgment bases. Thus, verbal accounts frequently underestimate judgment competence in research with children (e.g., Brainerd, 1973; Bullock, Gelman, & Baillargeon, in press; Goldberg, 1966). Research with adults, on the other hand, indicates that subjects' explanations often overestimate judgment sophistication. Both expert and nonexpert judges (e.g., Goldberg, 1968; Nisbett & Wilson, 1977) describe themselves as using complex rules that bear little resemblence to the simpler patterns of their actual performance. Ericsson and Simon (1980) note that relative accuracy of verbal reports may depend on the conditions under which the information is gathered. These findings would suggest that explanation-based analyses of judgment strategies should be treated with caution.

An alternative approach would be to analyze judgment strategies on the basis of subject's actual performance patterns (Ward & Jenkins, 1965; Jenkins & Ward, 1965; Shaklee & Tucker, 1980). That is, four different rules have been proposed to account for subjects' judgments of event covariations. Since different rules produce different judgments, covariation problems could be identified which would differentiate between those rules. In fact, careful structuring of a problem set should allow us to identify the specific strategy a subject is using.

A set of such problems is illustrated in Table 1a. Problems are structured hierarchically such that cell _a_ problems are correctly solved by all strategies; strategy _a_ versus _b_ problems are correctly solved by _a_ versus _b_, sum of diagonals, and conditional probability strategies. Sum of diagonal problems will be accurately judged by sum of diagonal and conditional probability strategies. Conditional probability problems would be correctly solved by the conditional probability strategy alone. Solution accuracy is indexed by the direction of the judged relationship (i.e., A_1 more likely given B_1, B_2, or no difference). A subject's solution pattern on the set of problems indicates the strategy used. Problems on the first row of Table 1a illustrate judgments predicted by each of the proposed rules. All problems in the row indicate relationships in which A_1 is more likely given B_1 than given B_2. However, an individual using the cell a strategy would judge only the first problem as such a relationship (cell a is the largest of the cells). A person using the a versus b strategy would accurately judge the first two problems in the row, but would say that A_1 given B_1 is as likely as A_1 given B_2 in the third problem (2-2), and that A_1 was less likely given B_1 than B_2 in the last problem (2-12). The sum of diagonals rule would result in the correct judgment of the first three problems, but would say that A_1 was as likely to occur with B_1 as with B_2 on the last problem (2+10) - (12+0). A subject using the conditional probability rule should accurately judge all of the first row problems. Table 1b identifies the solution pattern congruent with each strategy type. The probability of matching these judgment patterns by chance alone is .11 for cell a, .04 for a versus b, .01 for sum of diagonal, and .005 for the conditional probability pattern.

In two experiments, Shaklee and Tucker (1980) employed this diagnostic approach to identify judgment rules of 10th grade and college students. Subjects judged relationships in three problems for each proposed strategy type. Each problem consisted of 24 instances in which event states were defined for two events. Problems were set in contexts of everyday events (e.g., cake rises or falls with or without "special ingredient," plants healthy or not healthy which do or do not receive plant food). Subjects' performance indicated general conformity to the strategy set. Congruence with the cell a strategy pattern was frequent among the high school subjects (17%) but rare in the college sample (1%). Response patterns matched that of the a versus b strategy for 18% of the college sample (use of this strategy was not tested among the high school subjects). Judgment patterns were congruent with the conditional probability strategy for 17% of the high school subjects and 33% of the college sample. In each experiment, the modal response pattern conformed to that of the sum of diagonals rule (35% of the college subjects, 41% of the high school subjects). Subsequent studies demonstrated that children use increasingly sophisticated rules with increasing age in the 4th grade to college age span (Shaklee & Mims, 1981), and that adults tend to use simpler rules as the decision environment becomes more complex (Shaklee & Mims, 1982).

In sum, the data from several studies indicate that a carefully structured problem set can be profitably used to indicate strategies underlying judgments of covariations between events. Subjects in these experiments demonstrated at least some sophistication about appropriate covariation judgment, however, the optimal judgment rule was used by a minority of subjects. Such judgments are particularly interesting since they build so directly on the basic mathematical understanding of ratios and fractions. That is, people making covariation judgments should be comparing two conditional probabilities, each of which is a ratio between two frequencies. Our evidence indicates that substantial use of such a strategy does not occur until the 10th grade, and then by only a minority of subjects. This evidence is congruent with other research indicating that problems in application of ratio concepts are common among adults as well as children (Karplus & Peterson, 1970; Kurtz & Karplus, 1979; Capon & Kuhn, 1979).

In addition, these findings conflict with the past interview-based strategy analyses. In particular, Smedslund's (1963) only commonly reported strategy, cell a, is rarely seen in the performance patterns of our subjects. In light of this conflict, a direct comparison of explanation and judgment-based strategy analyses would be profitable. By this approach, subjects would be asked to complete a diagnostic problem set, then explain their judgment bases. Comparison of classification by the two methods might show areas of systematic disagreement. In addition, interview responses offer new information in evaluating our judgment-based analysis. That is, subjects may describe themselves as using rules which may differ from any of our proposed rules, but which would produce a judgment pattern on the problem set congruent with that of one of our rules. Finally, we learn something about subjects' insight into their own reasoning. Such understanding of subjects' own impressions about their task solutions would be particularly important in any attempts to improve judgment competence. That is, training may be maximally effective when it is oriented toward the individual's own understanding of his or her rule use.

A second interest in this study is in subjects' evaluations of the adequacy of the rules they use. Those using less sophisticated rules may or may not be aware of rule limits. This study will measure judgments of rule adequacy by asking subjects to give confidence ratings as they make their judgments in the problem set. Subjects who are less confident of erroneous responses than of correct responses must be aware of their rule limitations. In addition, subjects will be asked to identify the best rule among our set of proposed strategies.

Subjects for this experiment will be male and female college students, since our past research suggests that this age group should provide substantial numbers of a versus b, sum of diagonals and conditional probability judges. Sex of subject will be considered as a factor in the design in light of common findings of sex differences in math skills among adolescents and adults (e.g., Maccoby & Jacklin, 1974).

Method

Subjects

Subjects in the experiment were students in an introductory psychology class who participated in the experiment as one option in fulfillment of a course requirement. Subjects ranged in age from 18 to 32 years, with a mean age of 19.42. Sixty-two female and 54 male students participated.

Problems

Subjects judged a set of 12 covariation problems, structured so that each of four judgment rules would produce a distinctive judgment pattern on a problem set. Table 1a lists the actual problems used. The 12 problems include three problems for each of the four strategy types. One noncontingent and two contingent relationships are included for each strategy problem type.

Twelve different problem contents were developed, each of which consisted of a set of observations picturing one of two states for two potentially related everyday events. Three problems pictured bakery products which either rose or fell in association with the presence or absence of yeast, baking powder, or a "special ingredient." In three other problems, plants were pictured as healthy or sick as a possible function of the presence or absence of plant food, bug spray, or a "special medicine." In three problems people or animals were pictured as sick or healthy as a possible function of the presence or absence of a shot, liquid medicine, or a pill. The remaining three problems pictured a possible association between space creatures appearing happy or sad in the presence or absence of one of three weather conditions (snow, fog, or rain).

For each problem, data instances are pictured in a 2 x 2 table. In each case, the manipulated factor (or environmental event) defined the table columns (e.g. plant food, no plant food in example below), and the outcomes defined the table rows (plants healthy, not healthy in the example below). Each problem is introduced with a paragraph describing a context in which several observations were made on two potentially related variables. Subjects were asked to look at the pictured information and to identify the relative likelihood of one of the events when the second event was either present or absent. An example problem follows:

> A plant grower had a bunch of sick plants. He gave some of them special plant food, but some plants didn't get special food. Some of the plants got better but some of them didn't. In the picture you will see how many times these things happened together. The picture indicates that plants which were given special food were:

+3	+2	+1	0	-1	-2	-3
much more likely	somewhat more likely	a bit more likely	just as likely	a bit less likely	somewhat less likely	much less likely

> to get better than plants that weren't given special food. On your answer sheet write the scale number that best completes the sentence.

In addition, after each covariation judgment subjects were asked to rate their confidence as follows:

How certain are you in the accuracy of the above response?

1 2 3 4 5 7 7 8 9 10

just guessing absolutely certain

The 12 problems were grouped into problem blocks, including one problem from each strategy type. Problems within each block were arranged in a single random sequence. The three problem blocks were sequenced in a single random order. Numbers in parentheses to the left of the problems in Table 1a indicate the position of each problem in the problem sequence.

Once the problem set was completed each subject was interviewed and asked the following questions about his or her judgment:

1a. You've just completed several problems about the relationship between events. Can you tell how you solved them?

1b. (Experimenter turns to the last problem in the set - a conditional probability problem.) Can you use this problem to show me how you solved it? (strategy explanation)

2. (The participant is shown models of the strategy types while they are described.) Can you indicate, from the models presented, the strategy you used to solve the problems? (model choice)

3. Overall, which do you feel is the "best" strategy? (best strategy)

Each subject was tested and interviewed individually.

Instructions

Initial instructions introduced the subject to the concept of covariation in the context of "things that go together". Naturally occurring examples were given of positive relationships (i.e., tall people are more likely to be heavy than short people), negative relationships (i.e , it is less likely to rain when it is sunny than when it is cloudy), and unrelated events (i.e., a green truck is just as likely to run out of gas as a red truck). Subjects were told that they would be given some problems about hypothetical events that may or may not tend to occur together. A sample problem involving the occurrence of snow as it did or did not relate to atmospheric temperature was used to explain the stimulus materials and the problem format. Each subject gave a solution to the sample problem and was invited to ask questions about the task. Subjects were allowed to progress through the problems at their own pace and were encouraged to use the scratch paper provided if they desired.

Results

Results can be grouped according to their relevance to two issues. First, subjects' performances can be characterized in terms of the accuracy of problem solutions. Confidence ratings on these problems indicate subjects' beliefs about their accuracy. Secondly, judgment strategies are identified according to subjects' solution patterns on the problem set and their responses to the interview questions.

Accuracy. Accuracy was assessed in terms of the direction of the judged relationship (i.e., A_1/B_1 more, less or equally likely than A_1/B_2). Data are analyzed in terms of the number of problems correct per problem type. Relevant means for this analysis are reported in Table 2. A sex by problem type analysis of variance shows a main effect of problem type ($F(3,342) = 164.36$, $p < .001$) with mean accuracy of 2.88 for cell *a*, 2.65 for *a* versus *b*, 1.47 for sum of diagonals, and 1.21 for conditional probability problems. A main effect of sex of subject was also significant ($F(1,114) = 6.67$, $p < .01$), with more problems correctly solved by males than by females. The sex by problem type interaction was also significant ($F(3,342) = 3.08$, $p = .03$), with the greatest sex differences in accuracy for the sum of diagonals and conditional probability problems (see Table 2).

A sex by problem type analysis of variance of confidence ratings showed that subjects had some insight into solution accuracy. This was reflected in a significant effect of problem type on confidence ratings, with confidence decreasing as problem difficulty increased ($F(3,342) = 25.60$, $p < .001$). Mean confidence ratings were 8.5 for cell a, 8.4 for a versus b, 7.8 for sum of diagonals, and 7.7 for conditional probability problems. Confidence judgments did not differ by sex either as a main effect or in interaction with problem type.

Strategy. Each subject's pattern of solution accuracy on problems of the four types was used to identify his or her judgment strategy. Performance patterns congruent with the four strategies are illustrated in Table 1b. A subject was said to have passed criterion on a given problem type if he or she was accurate on two or more of the three problems of that type. A conditional probability subject should pass criterion on all problem types, sum of diagonals judges should pass criterion on all problem types except the conditional probability problems. Judges using the a versus b rule should pass criterion on cell a and a versus b problems. Cell a subjects should pass cell a problems alone. Someone who passes no criteria would be labeled Strategy 0. Judgment patterns that do not match any of these predicted patterns are classified as "other." Classification by this method will be referred to as the judgment-based strategy.

Distribution of these judgment-based classifications is illustrated for each of the two sexes in Table 3. These results indicate that all subjects passed at least one criterion, indicating that they understood the stimuli and had at least a simple understanding of the judgment to be made. Most frequently occurring were judgment patterns congruent with a versus b and conditional probability rules (36.2% and 31.9% of the samples respectively). Cell *a* and sum of diagonals classifications were less common (5.2% and 15.5% respectively). Judgments of 13 subjects failed to match any of our proposed patterns and were classified as "other". Table 3 also shows males as generally using more sophisticated strategies than those used by females. The distributions of the two sexes were compared by assigning each subject a number corresponding to the number of problem type criteria passed (cell *a* = 1, conditional probability = 4). A *t* test comparing males and females on strategy classification shows the sex difference in strategy use to be reliable ($t(101) = 2.68$, $p < .01$).

A final judgment-based strategy analysis compares the confidence ratings of subjects in each of the strategy classifications. A subject strategy by problem type analysis of variance showed no significant difference as a function of subject judgment strategy ($F(3,99) = 1.54$, ns). However, subject strategy did interact with problem type ($F(9,297) = 2.68$, $p < .01$). In this interaction, subjects classified as *a* versus *b*, sum of diagonals, and conditional probability judges showed parallel decreasing confidence as problem difficulty increased. However, cell *a* judges were least confident on *a* versus *b* problems. As in the previous analysis, confidence ratings also showed a main effect of problem type ($F(3,297) = 28.68$, $p < .001$).

Independent categorizations of subjects' strategies were based on their responses to the interview questions. First, subjects were asked to state their strategies (question 1a) and to demonstrate that strategy on a sample problem (question 1b). These two responses were considered together and coded according to whether they conformed to one of our four strategies. Two alternative responses were also common. Several subjects described themselves as using a variant of the conditional probability strategy which compared ratios of cell frequencies $\frac{a}{c}$ with cell frequencies

$\frac{b}{d}$. This strategy would produce the same judgments as our conditional probability strategy and will be labelled cell ratios. A second common response was for a subject to say that he or she had just guessed. Responses that did not match any of these categories were labelled "other". All responses were independently categorized by two coders. These two raters agreed on 89% of their ratings. Table 4 illustrates these classifications of subjects' explanations.

Once subjects had stated their strategies, they were shown a model of each of the four proposed strategies and asked to identify the one which most closely resembled their problem solving approach. This classification is referred to as model choice. Frequency of choices of the various models is shown in Table 5. Responses not represented in the strategy examples were coded as "other". Of these unclassifiable subjects, six said that they used more than one rule, and the remaining subjects said that they used some strategy not listed in the models.

Finally, subjects were asked to indicate the best strategy among the four examples. This response will be labelled best strategy. Table 6 lists frequencies of subjects' choices of each of the strategies. The group categorized as "other" includes several subjects who thought that two or more categories were equally good, some subjects who thought the cell ratio strategy was best, and some subjects who preferred some strategy not listed in the examples.

As in the judgment-based strategy classification, a subject's strategy classification on each of these three measures was converted to a scale score corresponding to the level of his or her classification in the strategy hierarchy. Since cell ratio judges should produce the same judgments as conditional probability rule users, these two rules were grouped together in these analyses. Subjects who said that they guessed were given a score of 0. Comparisons between classification methods were made in terms of these scale scores. The unclassifiable subjects were not included in these analyses.

Correlations between the various strategy classifications indicate some congruence between methods. The correlation between judgment-based strategy classification and stated strategy is .58 ($p < .001$). Classification of subjects by the two methods is illustrated in Table 4. Comparisons between these classification systems indicate that differences between classifications by the two methods do not show a reliable direction ($t(94) < 1$, ns). A close inspection of Table 4 shows that performance-explanation congruence differed according to subjects' strategy classification. Subjects whose performance patterns showed use of a conditional probability rule were almost uniformly accurate in describing their strategies (97% of conditional probability subjects). Among the other groups combined (excluding "other") only 24% of the subjects described rules congruent with their performance patterns. A comparison of the two groups shows this difference to be reliable ($\chi^2 = 45.46$, df = 1, $p < .001$).

Comparison between judgment-based classification and subject's model choice also showed reliable congruence between the two methods ($r = .45$, $p < .001$). Table 5 shows classification of subjects by the two methods. Comparison between the classification methods shows that model choices were neither reliably more nor less sophisticated than their judgment-based strategy classification ($t(98) < 1$, ns). The correlation between the strategy explanation and model choice measure indicates some agreement between these two self-report measures ($r = .53$, $p < .001$) with the subject classifications neither better nor worse by the two methods ($t(99) < 1$, ns).

Finally, subjects' selection of best strategy was compared to their classifications by other methods. Model choice and best strategy used the same multiple choice method, and were thus deemed to make the best case for comparison (see Table 6 for classification by the two methods). Subjects' selections of best strategy were reliably more sophisticated than the strategy they identified as their own ($t(88) = 5.35$, $p < .001$), suggesting that subjects recognized a better way to solve the problems when one was provided. Their choices of best strategy were also more sophisticated than their judgment-based strategy classifications ($t(84) = 7.19$, $p < .001$).

Discussion

These results offer considerable evidence on relative congruence among self-report and performance-based methods of identifying strategies underlying covariation judgments. All comparisons suggest some agreement between methods, with correlations ranging from .45 - .58. Correlations at this level indicate that subjects have some insight into their judgment bases. However, closer inspection of Tables 4 and 5 indicate that some subjects show considerably better insight than others. In particular, conditional probability subjects (judgment-based classification) are impressively accurate, with 97% describing a conditional probability (or cell ratio) strategy in their strategy explanation, and 84% selecting that strategy in the model choice measure. In sharp contrast, all other subject groups show poor congruence between the performance-based and self-report measures, with 24% agreement between judgment and explanation measures, 25% agreement between judgment and model choice.

The strength of our judgment-based classification system is our ability to evaluate whether a stated rule would produce the obtained judgment pattern.

A close inspection of Table 4 illustrates this comparison. For example, no subject with a cell *a* judgment pattern described him or herself as using a cell *a* judgment rule. Our interpretation of this difference would be ambiguous if these subjects described rules which would produce a cell *a* judgment pattern on the problem set. However, this was not the case. Half of these subjects said they were guessing, an approach which would yield a cell *a* pattern only 11 percent of the time (i.e., the chance probability of producing the pattern). The remaining subjects with cell *a* performance patterns said they were using cell ratios, a strategy which would result in a conditional probability judgment pattern. Subjects showing *a* versus *b* patterns also showed poor insight into rule use, with 11 of 42 classifiable subjects describing themselves as using rules which should produce more errors than they actually showed, and 11 subjects describing strategies which should have produced more accurate records than actually obtained. Most of the subjects whose judgment performance indicated sum of diagonals strategy use described strategies that would produce conditional probability judgment patterns. Several subjects described themselves as comparing cells $\frac{a}{c}$ with $\frac{b}{d}$, a strategy which would mimic a conditional probability strategy on the problem set. However, it is interesting to note that only one of the subjects who said they were using cell ratios produced a judgment pattern congruent with their described rule. As noted earlier, self-report and judgment pattern were congruent for conditional probability judges. In these cases we are not simply noting relative agreement between performance and explanation. Our rule diagnostic problem set also allows us to show whether subjects' self-reported rules would have produced their actual performance patterns.

One possible interpretation of poor agreement between judgment and explanation might be that subjects shifted rule use at some point in the problem set. A subject may have judged the initial problems by one strategy, but changed strategy by the end of the problem set. This individual's judgments might yield a classification according to the initial strategy, but he or she would be accurate in describing use of a different strategy to solve the last problem. In fact, some of our subjects said that they used more than one rule in response to the model choice question. This possibility may explain a few judgment-explanation discrepancies, but our rule classification system makes it unlikely as a general account. That is, a subject had to accurately judge at least two of the three problems of each strategy type to have passed criterion on that type. The problems were blocked such that one problem of each strategy type appeared in each third of the problem sequence. A subject would have to shift strategy after the eighth problem of the set to have met the criteria for his or her initial problem solution strategy in the judgment-based classification. Shifts at other points should produce judgment records that do not conform to any of our strategy patterns. These subjects would be labeled "other" and not be included in our method comparisons. In fact, such unclassifiable subjects were infrequent in this sample (11.2%).

These results show that agreement between different self-report measures is limited as well. The correlation between subjects' strategy explanation and model choice was a modest (though significant) .53. Thus, the issue is not simply one of the validity of self-report of strategy use. Method of obtaining that self-report affects subjects' responses as well.

These comparisons suggest that self-report may be a weak data-base for research on covariation judgment. We note, however, that there may be conditions under which self-reports would be more accurate. Our subjects described their strategies after solving a series of problems. Ericsson and Simon (1980) argue that features of memory and attention might predict that reports would be erroneous under these conditions. In particular, subjects must retrieve the relevant information from long term memory in order to explain their judgment rule. Potential sources of error include problems in storing or retrieving the information from long term memory and incomplete reporting of the available information. Ericsson and Simon (1980) argue that such problems are minimized by gathering self-reports through a think aloud technique in which subjects verbalize their reasoning as they solve the problem.

Although alternative techniques may improve self-report accuracy, our method is most relevant for comparison with past research in this area. In particular, Smedslund (1963) and Adi and colleagues (1978) each asked subjects to explain their strategies after making several judgments about event covariations. Our evidence suggests that self-report of less-than-optimal strategies will be inaccurate under these circumstances.

Considering covariation judgment as a problem in applied mathematics, our findings also have implications for educational assessment. That is, self-report may be a poor method for diagnosing the sources of individual student's errors in applying ratio concepts. Our finding of strategy classification differences in self-report accuracy are somewhat ironic from an educational point of view. That is, the students best able to report their strategies would be those who need help the least. The success of a program to improve these judgments may well depend on the starting strategy of the individual involved. Our evidence indicates that student self-report is unlikely to yield an accurate diagnosis of sources of judgment error.

Our subjects do show some insight into the strengths and weaknesses of their chosen strategies. First, confidence ratings showed that subjects were less confident of their accuracy on problems where errors were high than on problems where error rates were low. Secondly, twice as many subjects selected the conditional probability rule as the best rule as were classified as using the rule in problem solutions (32 percent vs. 65 percent). One might wonder why subjects would persist in using a rule they knew was flawed. However, shifting rules requires that subjects be able to generate a better rule to use. This evidence indicates that subjects are better at recognizing good rules than at producing those rules on their own.

A final consistent finding worth noting is the sex difference in judgment accuracy and strategy use. This sex difference is not surprising in the light of much past research showing males better than females in mathematical reasoning beginning in junior high and continuing throughout adulthood (Maccoby & Jacklin, 1974). Since the conditional probability rule builds so directly on comparisons of two ratios, we might expect sex differences in this judgment as well. Our method offers the additional advantage of identifying specific strategies employed by subjects of each sex. Compared to males, females were especially unlikely to use the conditional probability rule (19.3 percent vs. 46.3 percent), preferring the simpler and less accurate _a_ versus _b_ rule (41.9 percent vs. 29.6 percent). This difference could have several possible sources. One likely source is simply that the two sexes came to the experiment with different training backgrounds. Other studies have found males and females to be substantially different in participation in math courses by the time they get to college (Fennema, 1977; Keeves, 1973; Hall & Shaklee, note 1, National Assessment of

Educational Progrees, 1979). Further work would be required to assess the role of differential math training in sex differences in covariation judgments.

In overview, our results indicate that subject's self-reports of covariation judgment rules show limited congruence with actual judgment patterns. Self-report was an especially poor method for identifying sources of inaccuracy in judgment patterns. Such effects of assessment method offer a ready explanation for poor agreement about strategy use in past studies of covariation judgment. These results suggest that self-report measures are weak bases for drawing conclusions about strategy use. These problems with self-report in covariation judgment accord well with other research showing poor correspondence between subjects' judgments and their explanations about those judgments.

Reference Note

1. Hall, L. and Shaklee, H. An analysis of sex differences in judgment about covariation between events. Master's Thesis, University of Iowa, 1982.

References

Adi, H., Karplus R., Lawson, S., & Pulos, R. ntellectual development beyond elementary school VI: Correlational reasoning. School Science and Mathematics, 1978, 78, 675-683.

Brainerd, C. Judgments and explanations as criteria for the presence of cognitive structures. Psychological Bulletin, 1973, 79, 172-179.

Bullock, M., Gelman, R., & Baillargeon, R. The development of causal reasoning. In W. Friedman (Ed.), The Developmental Psychology of Time. New York: Academic Press, in press.

Capon, N., & Kuhn, D. Logical reasoning in the supermarket: Adult female's use of a proportional reasoning strategy in an everyday context. Developmental Psychology, 1979, 15, 450-452.

Ericsson, K., & Simon, H. Verbal reports as data. Psychological Review, 1980, 87, 215-251.

Fennema, E. Influences of selected cognitive, affective, and educational variables on sex-related differences in mathematics learning and studying. In L. Fox, E. Fennema, & J. Sherman (Eds.), Women and Mathematics: Research Perspectives for Change. National Institute of Education: Washington, D. C., 1977.

Fischbein, E. The intuitive sources of probabilistic thinking in children. Boston, Mass.: R. Reidel Publishing Company, 1975.

Goldberg, L. Simple models or simple processes? American Psychologist, 1968, 23, 483-496.

Goldberg, S. Probability judgments by preschool children: Task conditions and performance. Child Development, 1966, 37, 157-167.

Heider, F. The psychology of interpersonal relations. New York: Wiley, 1958.

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

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

Karplus, R., & Peterson, R. Intellectual development beyond elementary school II: Ratio, a survey. School Science and Mathematics, 1970, 70, 813-820.

Keeves, J. Differences between the sexes in mathematics and science courses. International Review of Education, 1973, 19, 47-62.

Kelley, H. Attribution theory in social psychology. In D. Levine (Ed.), Nebraska Symposium on Motivation (Vol. 15). Lincoln, Nebraska: University of Nebraska Press, 1967.

Kurtz, P., & Karplus, R. Intellectual development beyond elementary school VII: Teaching for proportional reasoning. School Science and Mathematics, 1979, 79, 387-398.

Maccoby, E., & Jacklin, C. The Psychology of Sex Differences. Stanford, California: Stanford University Press, 1974.

Michotte, A. The perception of causality. London: Methuen, 1963.

National Assessment of Educational Progress. Mathematical knowledge and skills. (National Institute of Education, Report No. 09-MA-02), Denver, Colorado, 1979.

Niemark, E. Longitudinal development of formal operations thought. Genetic Psychology Monographs, 1975, 91, 171-225.

Nisbett, R., & Ross, L. Human Inference: Strategies and shortcomings of social judgment. Englewood Cliffs: Prentice-Hall, 1980.

Nisbett, R., & Wilson, T. Telling more than we can know: Verbal reports on mental processes. Psychological Review, 1977, 84, 231-259.

Piaget, J., & Inhelder, B. The origin of the idea of chance in children. New York: Norton, 1975.

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

Shaklee, H., & Mims, M. Sources of error in judging event covariations: Effects of memory demands Journal of Experimental Psychology: Learning, Memory & Cognition, 1982, 8, 208-224.

Shaklee, H., & Tucker, D. A rule analysis of judgments of covariation between events. Memory and Cognition, 1980, 8, 459-467.

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

Ward, W., & Jenkins, H. The display of information and the judgment of contingency. Canadian Journal of Psychology, 1965, 19, 231-241.

Yost, P., Sigel, A., & Andrews, T. Nonverbal probability judgments by young children. Child Development, 1962, 33, 769-780.

Footnotes

Partial support for this research was provided by NIE grant NIE-G-80-0091. Many thanks to Renee Smith for her help in collecting this data. Reprint requests should be sent to Harriet Shaklee, Department of Psychology, University of Iowa, Iowa City, Iowa 52242.

[1]We had some difficulty defining a noncontingent relationship for the sum-of-diagonals problems. The problem we included (middle problem, column 3, Table 1A) deviates slightly from independence ($P(A_1|B_1) - P(A_1|B_2) = -.06$) by the conditional-probability rule. As a result we scored responses as correct if subjects concluded that $A_1|B_1$ was either less likely or just as likely as $A_1|B_2$. The problem does discriminate appropriately between the other judgment rules. Cell-a and a-versus-b judges should say that $A_1|B_1$ is more likely than $A_1|B_2$, sum-of-diagonal judges should say the two outcomes are equally likely.

Table 1

A) Cell frequencies used for each problem type

Cell a Problems

(11)

	B_1	B_2
A_1	11	2
A_2	4	7

(3)

	B_1	B_2
A_1	6	6
A_2	6	6

(5)

	B_1	B_2
A_1	2	11
A_2	7	4

a versus b Problems

(6)

	B_1	B_2
A_1	7	3
A_2	2	12

(9)

	B_1	B_2
A_1	3	3
A_2	9	9

(4)

	B_1	B_2
A_1	4	11
A_2	8	1

Sum of Diagonal Problems

(2)

	B_1	B_2
A_1	2	2
A_2	2	18

(7)

	B_1	B_2
A_1	9	7
A_2	5	3

(10)

	B_1	B_2
A_1	8	8
A_2	8	0

Conditional Probability Problems

(8)

	B_1	B_2
A_1	2	12
A_2	0	10

(12)

	B_1	B_2
A_1	1	5
A_2	3	15

(1)

	B_1	B_2
A_1	12	2
A_2	10	0

B) Strategy use and resultant patterns of problem accuracy.

(+ = accurate, 0 = inaccurate)

		Problem Strategy Type			
		Cell a	a versus b	Sum of Diagonals	Conditional Probability
Subject Strategy Type	Conditional Probabilities	+	+	+	+
	Sum of Diagonals	+	+	+	0
	a versus b	+	+	0	0
	Cell a	+	0	0	0
	Strategy 0	0	0	0	0

Table 2

Mean Judgment Accuracy Per Problem Type

	cell _a_	_a_ versus _b_	sum of diagonals	conditional probability	all types
females	2.81	2.64	1.23	1.00	1.90
males	2.96	2.65	1.72	1.43	2.20
all	2.88	2.65	1.47	1.21	2.05

Table 3

Judgment-Based Strategy Classifications

(percentages)

	cell a	a versus b	sum of diagonals	conditional probability	other	N
males	3.7	29.6	11.1	46.3	9.3	54
females	6.4	41.9	19.3	19.3	12.9	62
all	5.2	36.2	15.5	31.9	11.2	116

Table 4

Frequencies of Strategy Classifications by Judgment-Based And Strategy Explanation Methods

Judgment Based	Strategy Explanation							
	guess	cell _a_	_a_ vs _b_	sum of diagonals	conditional probability	cell ratios	other	all
cell _a_	3	0	0	0	0	3	0	6
_a_ versus _b_	9	2	13	2	1	8	7	42
sum of diagonals	3	0	1	1	6	6	1	18
conditional probability	0	0	0	1	35	1	0	37
other	2	0	1	0	0	10	0	13
all	17	2	15	4	42	28	8	116

Table 5

Frequencies of Strategy Classifications by Judgment-Based And Model Choice Methods

	Model Choice						
Judgment Based	guess	cell a	a versus b	sum of diagonals	conditional probability	other	all
cell a	0	1	3	1	1	0	6
a versus b	6	4	14	7	10	1	42
sum of diagonals	2	2	2	1	9	2	18
conditional probability	2	0	0	3	31	1	37
other	1	0	1	1	6	4	13
all	11	7	20	13	57	8	116

Table 6

Frequencies of Strategy Classifications by Model Choice

and Best Strategy Methods

Best Strategy	Model Choice: guess	cell a	a versus b	sum of diagonals	conditional probability	other	all
cell a	0	1	0	1	0	0	2
a versus b	2	0	4	1	0	0	7
sum of diagonals	1	2	4	1	1	0	9
conditional probability	6	2	10	4	49	5	76
other	2	2	2	6	7	3	22
all	11	7	20	13	57	8	116