DOCUMENT RESUME

ED 315 270 SE 051 140

AUTHOR	Gal, Iddo; And Others
TITLE	Which Group Is Better? The Development of Statistical Reasoning in Elementary School Children.
SPONS AGENCY	Pennsylvania Univ., Philadelphia. Literacy Research Center.
PUB DATE	Apr 89
NOTE	9p.; Paper presented at the Meeting of the Society for Research in Child Development (Kansas City, MO, April 1989).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Elementary Education; *Elementary School Mathematics; Grade 3; Grade 6; *Mathematical Concepts; *Mathematical Logic; Mathematics Education; *Mathematics Skills; *Problem Solving; *Ratios (Mathematics); Statistics
IDENTIFIERS	*Statistical Reasoning

ABSTRACT

Two issues were investigated in this study: (1) how well do children do descriptive statistics (i.e., accuracy); and (2) how do they do it. A total of 31 children in third grade (with no formal instruction in statistics) and 31 children in sixth grade (who had learned how to calculate a mean) compared a set of data from two domains: (1) the outcomes of frog jumping contests; and (2) the scores on a school test. Children were interviewed individually and asked to make decisions and explain their decisions related to eight problems. Accuracy rates for different types of problems and various solution methods used by children were analyzed. Solution strategies were categorized into three groups: (1) statistical; (2) proto-statistical; and (3) other/task-specific strategies. While sixth-graders were more accurate than third-graders, many of the sixth-graders had difficulty reasoning proportionally. Very few students reasoned statistically about the data. Many students used proto-statistical strategies. The more successful solvers seemed to choose solution strategies, which took into account those particular characteristics of the data sets which were relevant to the solution of the given problem. Lists 9 references. (YP)

Which Group is Better?: The Development of Statistical Reasoning in Elementary School Children

Paper presented at the meeting of the Society for Research in Child Development, Kansas City, MO, April 1989.

By

Iddo Gal, Karen Rothschild & Daniel A. Wagner
University of Pennsylvania

This study, part of the Statistical Reasoning in Children project (STARC), was supported by a grant from the Research Foundation of the University of Pennsylvania. The authors thank Ludo Scheffer for his help with data collection and analysis.

Address correspondence to the first author at: Literacy Research Center, Graduate School of Education, University of Pennsylvania, Philadelphia, PA 19104-6216.

INTRODUCTION

There are two main reasons for our interest in statistical reasoning in children. The first one is that research has shown that understanding of statistical principles, and their appropriate usage, are related to the quality of decisions, judgments and inferences people make. However, most of this research was done with adults (cf. Kahneman, Slovic, and Tversky; 1982), and has focused on various judgmental errors people commonly make, in part by not taking into account statistical principles, and on conditions that affect the appearance of such errors (Nisbett, Krantz, Jepson & Kunda, 1983). Several studies, such as those by Pollatsek and his colleagues (e.g., Pollatsek, Well & Lima, 1981), have focused on difficulties adults have with statistical concepts that are normally acquired through formal instruction (e.g., weighted means), though without much discussion of how adults come to know or learn such concepts.

The second reason is that American children learn very little about statistics in school. Most are taught only how to mechanically read charts and graphs, and perhaps, by the 4th or 5th grade, the algorithm for calculating an average. At the same time, knowledge of statistics and the ability to reason statistically have been repeatedly emphasized in all recommendations for improvements of the ways mathematics are taught in American Schools. The most recent of these is the set of standards just released by the National Council of Teachers of Mathematics (NCTM, 1989). Despite this interest, we know relatively little about statistical reasoning in children.

Work with children has concentrated in two main areas: Studies of formal understanding of concepts related to probability and randomness (e.g., Piaget and Inhelder, 1975; Fischbein, 1975; Kuzmak & Gelman, 1986), and studies of understanding of school-based concepts, such as Strauss & Bichler's (1988) research on children's understanding of the properties of the arithmetic mean. In a useful review, Garfield and Ahlgren (1988) summarized most of this work as it relates to children and instruction in stochastic. Little is known, however, about how children put those concepts to use when they have to *reason* about sets of data.

To address these issues we sought answers to two key questions. First, do children engage in 'descriptive statistics'? Do they organize their observations and synthesize different features of information that they have? Can they make summary statements about a set of data despite inherent variability, and, most importantly, what strategies do they use to make comparisons between sets of data? Second, what characterizes the development of statistical reasoning in the absence of direct instruction? For example, what kinds of "naive" or "everyday" concepts do children bring with them to their formal studies of statistics at school?

Findings reported here pertain primarily to the first question above, and address two issues: *How well* do children do descriptive statistics (what we call 'accuracy'), and *how* they do it.

METHOD

In the present study subjects were 31 children in 3rd grade and 31 children in 6th grade from middle-class private schools in the Philadelphia area. The 3rd graders had received no formal instruction in statistics. The 6th graders had learned how to calculate a mean as part of their school mathematics studies. Children were asked to compare sets of data derived from two domains: outcomes of frog jumping contests, and scores on a school test.

Children in the 'frogs' condition were asked to pretend that they were judges at a frog jumping competition and had to judge the results of competitions between teams of jumping frogs. Jumps of each team were presented as locations on two "jumping tracks". Children were asked to decide whether either of the teams had, on the whole, jumped "a lot better, a little better, or whether the teams were the same". Children in the 'grades' condition were asked to pretend that they were teachers who were about to teach a new unit, and who had given several classes a test to see what their students already knew about the new topic. Test scores were presented on the teacher's "grade sheets". The actual values used for the distributions were the same in both the frogs and grades conditions, only the symbols were different. Children were interviewed individually for 30-40 minutes. Each session started with a training stage, which included practice questions to establish comprehension of the task and the materials. Then children were presented with 9 comparisons between groups, and in each asked to make a decision and explain their decision and how they arrived at it.

Distributions were constructed in each condition so as to enable discovery of various strategies that children use. Several factors were manipulated[1].

(a) Distance between the means: group means could have been equal, slightly different, or very different. The number of cues for differences between groups was varied, to see to what cues children are paying attention [Examples: problems #1-#3 (see Appendix), in which the two groups have different mode, range and mean].
(b) Size of the 2 groups compared: groups may had different number of datapoints [example: problem #6]. The key issue was whether such problems would cause children to refer to and compare the groups on a proportional basis, rather than by absolute numbers.
(c) Overall sample size: Small groups had 6-9 cases. Large groups 21-36 cases. The key issue addressed was the extent to which children use estimation strategies when they cannot easily count, add or employ other strategies due to the large number of datapoints in each group [example: problem #8].

---

[1] See Appendix for schematic drawings of data sets used in problems 1-8. Actual stimuli were colored and used images of either miniature green frogs, or test papers marked with grades.

RESULTS

Data are first presented on accuracy rates for different types of problems, and then discussed in terms of the various methods used by children to arrive at their decisions.

Analysis of accuracy rates: Children's decisions were collapsed into a three-point scale: group A is better, group B is better, or groups A and B are the same. We use the term 'accuracy' to refer to whether a child's decision on a problem matched the result expected from comparison of the *arithmetic means* of the groups compared.

It should be mentioned that there was practically no evidence for children simply "guessing" on any of the problems. During training, children were informed that they would be asked to explain their reasons for each decision, and they were almost always able to support their decisions. For example, they explained their answers by pointing to locations on the "jumping track", verbally describing various differences between the groups compared, or by showing results of calculations. Hence, the present results are not to be discussed in terms of chance levels, as they might be in certain tasks involving forced-choice responses.

Figure 1 (see appendix) shows accuracy rates for problems in which the means of the groups were very different. Virtually all children answered these problems correctly. Especially informative is problem 2, where all members of Group A performed less well than members of group B, except for a single "outlier" that outperformed all members of group B. Almost all children made some verbal reference to the outlier, but none were misled by it.

Accuracy rates for problems in which distributions overlapped significantly, and in which group means were close or equal, are presented in Figure 2. As can be seen, accuracy rates dropped. In particular, and contrary to our expectation, children were less accurate on problem 5 than on other problems in this group. In problem 5 the mean, mode and symmetry of the distributions were clearly the same. We anticipated that children would easily judge these groups as equal in performance, yet they apparently had difficulties.

Problems which required the comparison of groups of different sizes appear in Figure 3. As indicated earlier, we assume that such problems would require children to think about ratios and proportions. The accuracy rates in Figure 3 show that these problems proved to be the most difficult ones. It should be mentioned that in each and every one of these problems the interviewer emphasized that groups (i.e., teams or classes) had different sizes, and further specified the number of members in each group. This was done to ascertain that children were aware of this crucial piece of information. However, only about 1/3 of the 3rd-graders, and 2/3 of the 6th-graders, gave *any* indication that group-size information was taken into account in forming a decision. A child that did not notice the difference may have said, for example: "class A is a little better" (Q. "why?") "because they have *more* students with high grades". In contrast, a child that mentioned and also utilized the information about group differences may have said, for example: "This class has more students with lower

grades... ah, but they have more students overall, so in general I think they are about the same".

Interestingly, the difference between the frogs and grades tasks seemed to have little effect on accuracy rates of 3rd graders. In the sixth grade, however, accuracy rates were considerably higher in the grades domain.

Analysis of solution methods: Children used many solution strategies and provided a variety of explanations for their decisions. We have divided them into three categories, which at present we call *Statistical, Proto-statistical,* and *other/task-specific* methods.

Statistical strategies were used by children who made decisions by comparison of *summaries* of the data in each group. Summaries involved, for example, calculating or estimating the arithmetic mean of each group, or using more fuzzy notions of where the "bulk of the data" lay in each group. Such summaries involved integration or synthesis of *all* the different kinds of information available about a group (features like, e.g., range, dispersion, shape of distribution, central tendency), without over-attention to specific datapoints.

Children who used *proto-statistical* strategies were sensitive to some or all of the various features of the data that should be considered in summarizing a set of data, but either ignored other features, or were not able to synthesize all the information they had. Some students appeared to look at only part of the data. For example, 3rd-graders sometimes compared groups by focusing exclusively on their modes, and decided in favor of the group that had the "tallest" column, but without consideration of the actual value (i.e., location on the jumping track) of the modal column. Others attempted to "balance" high and low scores within a group, but subsequently were not able to coordinate the knowledge they gained in a way that allowed comparison of the two groups.

Other/task specific strategies included, for example, adding, in which students simply added jump lengths or grade points in a mechanical fashion. Many students blindly added even when groups were of unequal sizes, and hence made errors at predictable places. Students often labored at adding even when a visual inspection of the data (e.g., in problem 2) could lead to a straightforward decision. Qualitative explanations were also included in this category, for example, "the frogs in this team are less consistent, because they're spread out more than the other team. I think that the other team is better", or statements that a team with a smaller number of frogs (e.g., in problem 8) is better because they "try harder".

DISCUSSION

We have identified several factors that affect children's ability to correctly draw conclusions from data. The major ones are: (a) The number of features that children need to attend to and synthesize; and (b) Whether or not the situation requires the use of proportions, either to summarize within a single set of data, or to compare sets of data to each other. Most 3rd-graders did not seem to grasp the significance of the fact that in some problems one of the groups had a different number of datapoints, and that it should disqualify certain explanations. While 6th-graders were overall more accurate than 3rd-graders, many of the 6th-graders had difficulty reasoning proportionally.

With respect to reasoning strategies, children used many many different methods and explanations. Very few children reasoned statistically about the data. We did not have prior expectations with regard to 3rd graders, but were surprised that most 6th-graders, who had all learned about averages in school, did not apply this knowledge, and did not look for central tendency of distributions. Many students used strategies we termed "proto-statistical". They showed awareness of some of the factors that should be taken into account, but were not able to synthesize them and reason about them to come to correct conclusions. Finally, some students seemed to use strategies that were not "statistical" as we use the term, even though they were sometimes appropriate to use and could lead to correct conclusions.

The majority of children used more than one strategy, which was entirely appropriate because certain problems could be solved correctly by a variety of methods. However, the more successful solvers seemed to choose solution strategies which took into account those particular characteristics of the data sets which were relevant to the solution of the given problem. This should be contrasted with those that consistently used one type of explanation or method, and often made errors in predictable places.

Several questions emerge from this work. While we see age differences in performance, it is unclear how maturational changes, school and cultural effects interact to create the phenomena we observed. We are currently expanding our sample to include both older children, and children with different experiential backgrounds, to explore this interaction.

We are currently analyzing other data we have, about children's understanding of the word "average" as it is used in various contexts, and hope to be able to begin and answer other questions raised by our study. For example - the relationship between the learning of statistical *concepts* (such as mean or proportion) and the development of statistical *reasoning*.

From our perspective, it is important to further explore proto-statistical ways of reasoning about data, because they demonstrate how children can be aware of some of the parameters that go into a statistical analysis, but have not yet learned or developed to a point where they apply them appropriately. In educational terms, proto-statistical strategies would seem to be an important point of departure for pedagogical intervention.

REFERENCES

Fischbein, E. (1975). The intuitive sources of probabilistic thinking in children. Dordrecht. The Netherlands: D. Reidel.

Garfield, J., & Ahlgren, A. Difficulties in learning basic concepts in probability and statistics: implications for research, 1988, Journal of Research in Mathematics Education, 19(1), 44-63.

Kahneman, D., Slovic, P., & Tversky, A. Judgment under uncertainty: Heuristics and biases, in D. Kahneman, P. Slovic, & A. Tversky (Eds.) Judgment under uncertainty: Heuristics and biases, 1982, Cambridge, UK: Cambridge University Press.

Kuzmak, S. D., & Gelman, R. Young children's understanding of random phenomena, Child Development, 1983, 57, 559-566.

National Council of Teachers of Mathematics (NCTM). Curriculum and evaluation standards for school mathematics, 1989, Reston, VA: NCTM.

Nisbett, R.E., Krantz, D.H., Jepson, C., & Kunda, Z. The use of statistical heuristics in everyday inductive reasoning, 1983, Psychological Review, 90(4), 339-363.

Piaget, J., & Inhelder, B. The origin of the idea of chance in children, 1975, London: Routledge and Keegan Paul.

Pollatsek, A., Lima, S., & Well, A. D. Concept or computations: Students' understanding of the mean, 1981, Educational Studies in Mathematics, 12, 191-204.

Strauss, S., & Bichler, E. The development of children's concepts of the arithmetic average", 1988, Journal of Research in Mathematics Education, 19(1), 64-80.

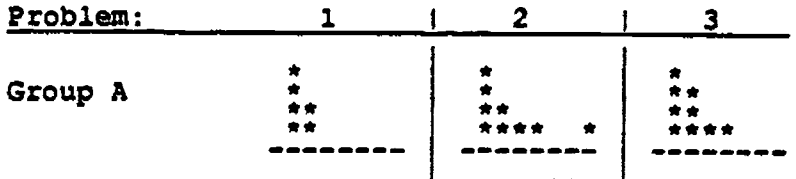
Handout for SRCD presentation - April 1989

Which Group is Better?: The Development of Statistical Reasoning in Elementary School Children

Iddo Gal, Karen Rothschild & Daniel A. Wagner

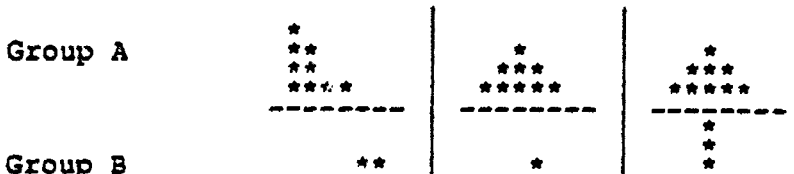
Literacy Research Center, Graduate School of Education
University of Pennsylvania, Philadelphia, PA 19104-6216.

Fig. 1: Percentage correct on problems in which means are very different



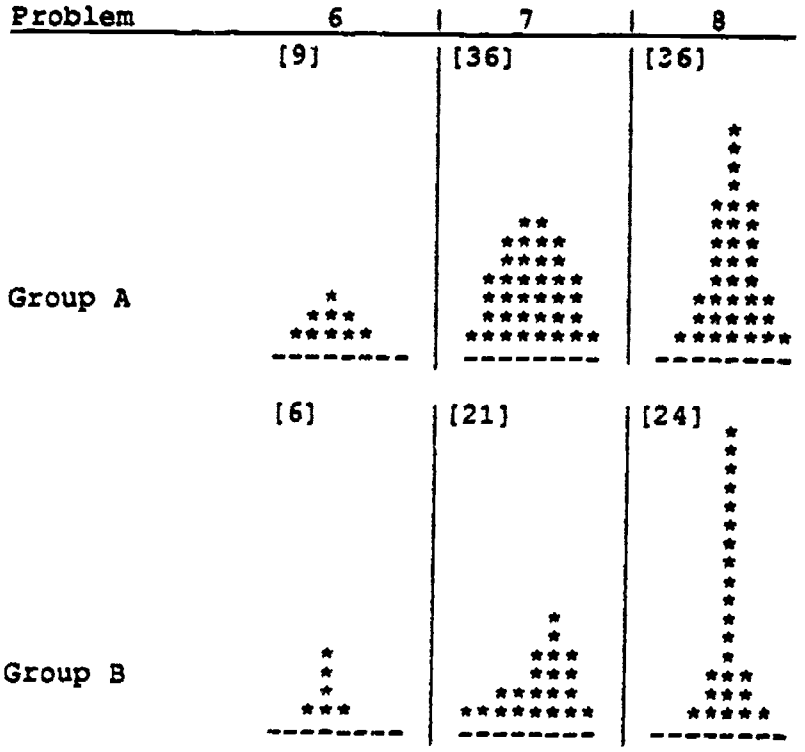
	Problem 1	Problem 2	Problem 3
Grade 3			
Frogs	100%	100	100
Grades	100	100	93
Grade 6			
Frogs	100	100	100
Grades	100	100	100

Fig. 2: Percentage correct on problems in which means are close or equal



	Problem 3	Problem 4	Problem 5
Grade 3			
Frogs	100%	75	31
Grades	93	93	40
Grade 6			
Frogs	100	67	57
Grades	100	60	70

Fig. 3: Percentage correct on problems in which groups have different sizes



	Problem 6	Problem 7	Problem 8
Grade 3			
Frogs	6%	31	38
Grades	13	0	15
Grade 6			
Frogs	24	38	29
Grades	50	22	40

Number of subjects (all problems):

Grade 3	Frogs	N=16
	Grades	N=17
Grade 6	Frogs	N=21
	Grades	N=10

Order of presentation: 1, 5, 6, 2, 4, 3, 7, 8