DOCUMENT RESUME

ED 120 022 SE 020 598

AUTHOR Osborne, Alan R., Ed.
TITLE Investigations in Mathematics Education, Vol. 8 No. 1.
INSTITUTION Ohio State Univ., Columbus. Center for Science and Mathematics Education.
PUB DATE 75
NOTE 71p.
AVAILABLE FROM The Ohio State University, Center for Science and Mathematics Education, 244 Arps Hall, Columbus, Ohio 43210 (Subscription, $6.00, $1.75 single copy)

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage
DESCRIPTORS *Abstracts; *Cognitive Development; Effective Teaching; Instruction; *Learning; Logic; *Mathematics Education; Measurement; Problem Solving; Research; *Research Reviews (Publications); Testing
IDENTIFIERS Proof (Mathematics)

ABSTRACT

Sixteen research reports related to mathematics education are abstracted and analyzed. The reports abstracted were selected from five educational journals, four psychological journals, a mathematics journal, and a book of readings. Eight of the articles are related to logical thinking, inference, proof, and problem solving. Three reports deal with measurement or conservation by students at various stages of cognitive development, while one deals with performance of young children on embedded figures tests. Two articles concern mathematics testing, one deals with learning hierarchies, and one concerns teacher effectiveness. Research related to mathematics education which was reported in RIE and CIJE between October and December 1974 is listed. (SD)

Vol. 8, No. 1 Winter 1975

INVESTIGATIONS IN MATHEMATICS EDUCATION

Expanded Abstracts
and
Critical Analyses
of
Recent Research

Center for Science and Mathematics Education
The Ohio State University
in cooperation with
the ERIC Science, Mathematics and
Environmental Education Clearinghouse

INVESTIGATIONS IN MATHEMATICS EDUCATION

Published quarterly by

The Center for Science and Mathematics Education
The Ohio State University
1945 North High Street
Columbus, Ohio 43210

With the cooperation of the ERIC Science, Mathematics, and Environmental Education Information Analysis Center

Subscription Price: $6.00 per year. Single Copy Price: $1.75
Add 25¢ for Canadian mailings and 50¢ for foreign mailings.

Dear I.M.E. Reader:

I am very pleased to announce a change in editorship of Investigations in Mathematics Education. Because of my growing involvement in coordinating ERIC products and a desire to become more actively involved in mathematics education research projects, I have asked to have the editorship of I.M.E. transferred. Beginning with Volume 8, the new editor of I.M.E. will be Alan Osborne, with Marilyn Suydam acting as associate editor. I'm sure you will agree with me that two finer people could not be found to assume this responsibility. Under their leadership I am sure that I.M.E. will continue to increase its role as a service to the mathematics education research community.

At this time, paid subscribers should have received all 4 issues of Volume 7. Volumes 8 (1975) and 9 (1976) will be produced on an accelerated schedule to bring us back to correspondence with the calendar year.

I will continue to contribute the ERIC research listings to each issue of I.M.E. and hope to continue to be involved in its future progress, including doing an occasional abstract myself. I would appreciate comments from readers on increasing the usefulness of this list. In addition the new editors welcome your comments regarding the total format and contents of I.M.E.

Jon L. Higgins

A note from the editor ...

I am pleased to assume the responsibility for editing Investigations in Mathematics Education. I feel that I.M.E. has served well two purposes in the past. First, it has given the community of scholars in mathematics education an opportunity to indulge in healthy self-criticism directed at improving the quality of research and scholarship in mathematics education. Second, I.M.E. has abstracted articles appearing in journals that are not regularly read by most mathematics educators. The intent has been to examine materials not widely known in the field but pertinent to mathematics education. The high priority of these two purposes will continue.

You can provide the editors a service relative to the second purpose. If you know of research reports that you feel deserve review and wider dissemination, please advise us of them. We have two primary criteria we respect in selecting research for abstracting: it must concern mathematics teaching or learning and it must be empirical in nature. The research may appear in foreign journals as well as those published in the United States and may be in languages other than English. Dissertations are not abstracted; I.M.E. favors publishing abstracts and commentary of research found in more readily accessible sources. We would appreciate your suggestions.

The Advisory Board has decided to produce a special issue of I.M.E. in the near future. The issue will be devoted to the NLSMA Reports. Our perception is that because of the sheer quantity of NLSMA material, many individuals have not acquired the familiarity with the reports that they deserve. We feel that a synopsis of NLSMA results and a critical examination of the studies will be helpful to many people.

I hope that you will continue to find I.M.E. helpful under my editorship. We will be publishing several issues during 1976 in order to match I.M.E.'s publication schedule with the calendar. Your help and support will be appreciated.

Alan R. Osborne
Editor

Winter 1975

MATHEMATICS EDUCATION RESEARCH STUDIES REPORTED IN RESEARCH IN EDUCATION
October - December, 1974

ED 092 219 Entwisle, Doris R.; Webster, Murray, Jr. Middle-Class and Lower-Class Children: Expectations in First Grade. 4p. MF and HC available from EDRS.

ED 092 371 Geeslin, William E. Comparison of Content Structure and Cognitive Structure in the Learning of Probability. 22p. MF and HC available from EDRS.

ED 092 382 Mack, Richard G.; And Others. Group Forces in Instruction: Relationship to Learning. A Student and Teacher Training Program in Small Group Learning Functions. 91p. Not available from EDRS.

ED 092 383 Barnett, Jeffrey C. Structural and Linguistic Variable Identification and Instruction in Verbal, Arithmetic Problem Solving. 33p. Not available from EDRS.

ED 092 384 Branca, Nicholas A. Learning Mathematical Structures. 22p. MF and HC available from EDRS.

ED 092 385 Gaslin, William L. A Comparison of Achievement and Attitudes of Students Using Conventional or Calculator Based Algorithms for Operations on Positive Rational Numbers in Ninth Grade General Mathematics. 20p. Not available from EDRS.

ED 092 402 Trimmer, Ronald G. A Review of the Research Relating Problem Solving and Mathematics Achievement to Psychological Variables and Relating These Variables to Methods Involving or Compatible with Self-Correcting Manipulative Mathematics Materials. 35p. MF and HC available from EDRS.

ED 092 407 Cox, Linda S. Analysis, Classification, and Frequency of Systematic Error Computational Patterns in the Addition, Subtraction, Multiplication, and Division Vertical Algorithms for Grades 2-6 and Special Education Classes. 130p. MF and HC available from EDRS.

ED 092 584 Axtell, Dayton. A Brief Study of Those Not Completing the Mathematical Part of the School and College Ability Test. 10p. MF and HC available from EDRS.

ED 092 599 Potts, George R. How Subjects Do Not Store and Retrieve Information About Ordered Relationships. 18p. MF and HC available from EDRS.

ED 093 511 Gurecki, Karen J.; Wurster, Stanley R. A Study of the Relationship of the Length of Continuous Attendance at a Single School to Reading and Arithmetic Achievement Test Scores. 47p. MF and HC available from EDRS.

ED 093 702 Hymel, Glenn M. An Investigation of John B. Carroll's Model of School Learning as a Theoretical Basis for the Organizational Structuring of Schools. Final Report. 363p. MF and HC available from EDRS.

ED 093 716 Wiles, Clyde A. Comparison of Three Instructional Sequences for the Addition and Subtraction Algorithms. Technical Report 273. 277p. MF and HC available from EDRS.

ED 093 717 Jurdak, Murad Eid. The Effects of Emphasizing Mathematical Structural Properties in Teaching and of Reflective Intelligence on Four Selected Criteria. Technical Report 275. 190p. MF and HC available from EDRS.

ED 093 718 Klausmeier, Herbert J.; Feldman, Katherine Vorwerk. The Effects of a Definition and a Varying Number of Examples and Nonexamples on Concept Attainment. 18p. MF and HC available from EDRS.

ED 093 720 Raschke, Jewel P.; And Others. Career Patterns of Secondary School Mathematics Teachers. 41p. MF and HC available from EDRS.

ED 093 721 Waller, T. Gary; Wright, Robert H. The Effect of Training on Accuracy of Angle Estimation. 34p. MF and HC available from EDRS.

ED 093 883 Edwards, Keith J.; DeVries, David L. The Effects of Teams-Games-Tournament and Two Instructional Variations on Classroom Process, Student Attitudes, and Student Achievement. Report Number 172. 33p. MF and HC available from EDRS.

ED 094 046 Blayton, Ethel J.; Ryals, Joe. The Effects of a Highly Concentrated Reading and Mathematics Program on the Achievement of Inner-School Primary Pupils: Nathan B. Forrest Elementary School, 1972-73. Research and Development Report, Vol. 7, No. 28, November 1973. 21p. MF and HC available from EDRS.

ED 094 841 Friesen, Chuck. An Analysis of General Education Mathematics Programs in Two-Year Colleges. A Research Report. 29p. MF and HC available from EDRS.

ED 095 008 Beardslee, Edward C.; Jerman, Max E. Structural, Linguistic and Topic Variables in Verbal and Computational Problems in Elementary Mathematics. 16p. Not available from EDRS.

MATHEMATICS EDUCATION RESEARCH STUDIES REPORTED IN JOURNALS
AS INDEXED BY CURRENT INDEX TO JOURNALS IN EDUCATION
October - December, 1974

EJ 098 336 Weiner, Susan L. "On the Development of More and Less." Journal of Experimental Child Psychology, v17 n2, pp271-287, Apr 74.

EJ 098 529 Johnson, Martin L. "The Effects of Instruction on Length Relations on the Classification, Seriation, and Transitivity Performances of First- and Second-Grade Children." Journal for Research in Mathematics Education, v5 n3, pp115-125, May 74.

EJ 098 531 Schoen, Harold L.; Kreye, Bettibel C. "Five Forms of Written Feedback to Homework in a Mathematics Course for Elementary Teachers." Journal for Research in Mathematics Education, v5 n3, pp140-145, May 74.

EJ 098 532 Austin, Joe Dan. "An Experimental Study of the Effects of Three Instructional Methods in Basic Probability and Statistics." Journal for Research in Mathematics Education, v5 n3, pp146-154, May 74.

EJ 098 533 Cromer, Fred Eugene. "Structural Models for Predicting the Difficulty of Multiplication Problems." Journal for Research in Mathematics Education, v5 n3 pp155-166, May 74.

EJ 098 534 Sowder, Larry. "The Influence of Verbalization of Discovered Numerical- or Sorting-Task Generalizations on Short-Term Retention in Connection with the Hendrix Hypothesis." Journal for Research in Mathematics Education, v5 n3, pp167-175, May 74.

EJ 098 540 Runnels, Patricia; Runnels, L. K. "A Kindergarten Mathematics Program." School Science and Mathematics, v74 n5, pp361-365, May-Jun 74.

EJ 098 547 Sowell, Evelyn. "Elementary Teachers Learn to Sequence Mathematics Instruction." School Science and Mathematics, v74 n5, pp403-406, May-Jun 74.

EJ 099 518 Van Osdol, B. M.; And Others. "The Effects of a Token Reinforcement System on Arithmetic Achievement and Short-Term Retention by Educable Mentally Retarded Students in a Public School Setting." Journal of Mental Deficiency Research, v17 pts 3&4, pp247-254, Sep-Dec 73.

EJ 101 575 Wegman, Edward J. "Computer Graphics in Undergraduate Statistics." International Journal of Mathematical Education in Science and Technology, v5 n1, pp15-23, Jan-Mar 74.

IMMEDIATE KNOWLEDGE OF RESULTS AND TEST PERFORMANCE. Beeson, Richard O. Journal of Educational Research, v66, pp224-226, January 1973.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Boyd Holtan, West Virginia University.

1. Purpose

To compare the effects of immediate knowledge of results and delayed knowledge of results on test performance.

2. Rationale

One of the basic precepts of programmed learning is that the immediate knowledge of results increases the quality of learning. The effect of immediate and delayed knowledge of results on achievement test scores was investigated in this study in an effort to find results paralleling those for programmed learning.

3. Research Design and Procedure

The sample consisted of three groups of mathematics students.

Group I	A University class in mathematics for elementary school teachers (n = 30).
Group II	A University class in remedial mathematics (n = 15).
Group III	A junior high school class in general mathematics (n = 30).

Each of the three groups was randomly separated into two subgroups.

The treatment consisted of multiple choice tests in which the results of items in either the first or last half of the tests were immediately available to the students through the use of punchboard examination format. One subgroup received immediate knowledge of results on either the first or second half of the items on the test. The other subgroup in each group received the opposite condition. The conditions were reversed between the subgroups on consecutive tests. Group I completed four 1-hour examinations and a final examination. Groups II and III completed three 1-hour examinations. The scores on each of the half tests for each subgroup were compared.

The means and standard deviations for KR (immediate knowledge of results) and DKR (delayed knowledge of results) scores for each group are shown in Table 1.

Table 1

Means and Standard Deviations for KR and DKR

		Group					
		I		II		III	
Exam	Variable	Mean	SD	Mean	SD	Mean	SD
1	KR	14.53	2.47	8.27	1.33	12.83	3.75
	DKR	14.00	2.89	7.67	1.78	11.67	3.82
2	KR	15.83	2.52	7.60	2.16	12.83	4.32
	DKR	16.10	3.36	7.00	1.77	11.40	4.44
3	KR	12.73	3.63	6.47	2.31	10.23	3.92
	DKR	12.80	3.17	6.07	3.21	10.63	4.14
4	KR	13.43	3.75				
	DKR	12.63	4.07				
Final	KR	28.97	6.08				
	DKR	27.27	6.38				

4. Findings

A two-way analysis of variance was used to test for significant differences between the means of the halves of the ten 1-hour examinations and the final examination. There were no statistically significant differences between student performance with KR and DKR on the ten 1-hour examinations. The difference on the Group I final examination, however, was significant at the .05 level in favor of KR.

Overall performance was slightly higher when students were given immediate knowledge of results.

5. Interpretations

The investigator suggests that these findings justify that mathematics teachers should make greater use of testing devices which provide knowledge of results.

Critical Commentary

Although the findings suggest the better performance with the immediate knowledge of results, this was statistically supported on only one of the eleven pairs, the final examination for Group I. The difference of the means for the final exam was very small, only 1.7 score points.

The samples used were intact classrooms of students at various levels. There is no further clarification or description of the samples so that generalizing results to other groups must be done with caution.

The means of the KR were not consistently better than the DKR within each of the three groups. Only in Group II, the college remedial mathematics class members, was the DK higher for all three tests. A conjecture may be that mathematical ability or past experience is interacting with the immediate knowledge of results. The better students knew more answers and needed less external reinforcement. The data did not indicate the possible total scores for each examination, so it is impossible to determine the proportion of each test which each group completed or if there was a test ceiling effect.

Boyd Holtan
West Virginia University

IDENTIFICATION OF EMBEDDED FIGURES BY PRIMARY GRADE CHILDREN. Bright, G. W. School Science and Mathematics, v75 n6, pp535-541, October 1975.

Expended Abstract and Analysis Prepared Especially for I.M.E. by Richard A. Lesh, Northwestern University.

1. Purpose

The study claims to report replication and extension data for three papers published by the Soviet psychologist Yakimanskaya (1958, 1959, 1962). Bright's statistical hypotheses were that among students in the primary grades:

(a) Overlapping and non-overlapping geometric figures are identified equally often in the context of complex drawings.

(b) The order of identification of figures is random.

2. Rationale

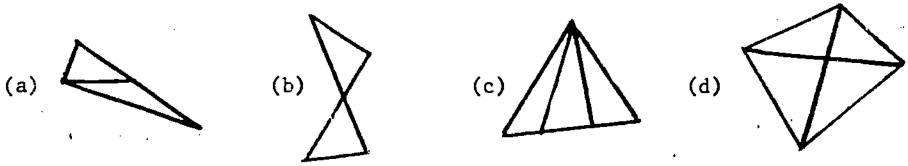
Yakimanskaya (1958) involved 350 pupils from the fourth through the eighth grades, and a variety of overlapping geometric figures formed by drawing diagonal lines inside various triangles, parallelograms, and rhombuses. The experiments showed that the amount of information a youngster reads out of (or into) a diagram depends directly upon the degree of sophistication of the processes of analysis, syntheses, and abstraction (Kalmykova, 1950). Subjects differed sharply in the method they used to identify geometric figures within a complex configuration, they differed in the quantity of simple figures they could identify within a complex configuration, and they differed concerning the type of figures that were identified - i.e., overlapping figures were not identified by youngsters who used primitive analysis and syntheses procedures.

Yakimanskaya (1959) observed that pupils who have difficulty reasoning abstractly often tend to rely heavily on diagrams (Zykova, 1955; Menchinskaya, 1955). Yakimanskaya (1959) systematically varied the verbal and visual components of a variety of mathematical problems and analyzed individual differences according to the ease, rapidity, and accuracy with which youngsters solved the problems. Some of the problems involved overlapping figures of the type used in Yakimanskaya (1959). Among the youngsters who were involved in the study (the number of subjects was not reported, but apparently $n > 50$), a significant number had difficulty analyzing a diagram but were able to analyze a problem's verbal condition with ease. Yakimanskaya concludes that instruction aimed at developing children's spatial abilities could improve their problem solving abilities.

Yakimanskaya (1962) reviews several "spatial abilities" studies (e.g., Gurvich, 1912; Galkina, 1961; Goworkova, 1962) including a program of research that Yakimanskaya had conducted in several Moscow schools. Taken together, the studies support the idea of developing spatial abilities through everyday problem solving activities and the idea that refined spatial abilities can improve youngster's problem solving abilities in a variety of situations. The study having most to do with Bright's study involved 30 subjects and concerned "overlapping figures" items similar to those in Yakimanskaya, 1958 and 1959. The study was a "teaching experiment" which aimed at increasing children's flexibility in dealing with geometric figures. Yakimanskaya reports that when youngsters (e.g., fourth graders) attempt to identify and classify geometric figures they often: (a) impose artificial restrictions on the figures (e.g., a square with diagonals is no longer considered to be a square, a square is not considered to be a rectangle, etc.), (b) ignore certain relevant attributes of the figures (e.g., because they center on the lengths of the sides of a rhombus and ignore other attributes, a rhombus may be considered to be a square). Youngsters are also not very active in looking for figures within complex configurations (e.g., fourth graders might recognize only a small number of line segments on a ruler, or only a small number of squares on a checkerboard). The study concluded that elementary school children can be taught to overcome many of these spatial abilities difficulties through instructional exercises involving overlapping figures.

3. Research Design and Procedure

Fifteen subjects (5 seven, 4 eight, 5 nine, and 1 ten year old; 12 females) randomly selected from a summer school art enrichment class, were interviewed individually concerning the following four geometric figures.



As preparation for the interview, each subject was asked to describe, identify, and then draw a triangle. Then, each was asked to draw a different triangle. Finally, the subject was asked to identify (and trace with a finger) each of the triangles in figures a, b, c, and d. The order of tracings was recorded. If no overlapping figures were identified, the experimenter traced one of the overlapping triangles in figure d and then asked the subject to try to find more triangles in figures c and d.

4. Results

(a) Only one subject failed to identify all of the non-overlapping triangles. On the other hand, only one subject spontaneously identified an overlapping triangle without a hint from the experimenter. Therefore, hypothesis 1 was rejected.

(b) After a hint from the experimenter, 11 subjects were able to identify one more overlapping triangle in figure d, and six of these subjects identified another overlapping triangle in figure c. However, none of the subjects identified more than one overlapping triangle within figure c or d. No age trends were apparent.

(c) From data concerning the order in which non-overlapping figures were identified, the following identification strategies were noted: figure a) upper first, lower first; figure b) upper first, lower first; figure c) left to right, right to left, other; figure d) clockwise, counter-clockwise, other. A χ^2 test was used to reject the null hypothesis ($p < .01$) for figures c odd.

4. Interpretations

The experimenter recognized the obvious weaknesses of the study, e.g., small and somewhat atypical sample, sex bias, limited methods of quantifying data. Nonetheless, he concludes: (a) that primary school children can be expected to have difficulties identifying overlapping figures within complex configurations, (b) that subjects seem to use identifiable search strategies.

Critical Commentary

To call Bright's study a replication and extension of Yakimanskaya's research (1958, 1959, 1962) is presumptuous to say the least. Bright's study does deal with primary grade children whereas Yakamanskaya's research focused on the upper grades; but (a) no attempt was made to investigate variables like the degree of sophistication of the processes of analysis, synthesis, or abstraction, (b) the "complex figures" in Bright's study were far simpler than those in Yakimanskaya's research, and fewer different kinds of overlapping figures were involved, (c) the search strategies that Yakimanskaya identified were quite different from the simple "left-right," "clockwise-counter clockwise" identification strategies in Bright's study, (d) it is not clear that the identification strategies in Bright's study correspond to any procedure that the children actually used. For example, a child who was classified as using a "left-right" strategy might actually have been using a "bigger-smaller" strategy, a random strategy, or some other strategy. And (e) it is not clear how a χ^2 test can be used to test hypotheses b (strategy-vs-no strategy) when the alternative categories were (i) the clockwise strategy (ii) the counter-clockwise strategy, (iii) other strategies, or (1) the "right-left" strategy, (ii) the "left-right" strategy.

It is a well known fact that primary grade children have difficulty identifying overlapping figures. This fact was clear even from Yakimanskaya's research with fourth through sixth graders. Furthermore, a variety of theoretical perspectives have given explanations for why the phenomenon occurs.

Consequently, before further data gathering is done in this area, a thorough review of the literature seems in order. Relevant information is available from Gestalt theory, Piagetian theory, Perception theory, and from other theoretical perspectives (Von Hiele, Bruner).

It seems clear that the directions Bright used affected the results that were obtained. For example, it seems odd that none of the children recognized that figures b and d were triangles. Other directions would certainly have elicited this simple identification.

Yakimanskaya's research emphasizes that instruction can radically influence children's spatial abilities, and his most interesting results were obtained from children who were beginning to refine some of these abilities. So, well designed replication and extension studies are certainly in order - especially if an effort is made to establish links between "spatial abilities" and other areas of concept formation.

Richard A. Lesh
Northwestern University

EJ 080 320 TM 500 951

STABILITY OF TEACHER EFFECTIVENESS. Brophy, Jere E. American Educational Research Journal, v10, pp245-252, Summer 1973.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Terrence G. Coburn, Oakland Schools, Pontiac, Michigan.

1. Purpose

To study the stability of teacher effectiveness and gain some insights into the degree of consistency of typical teachers over three years.

a) How stable is the effectiveness of typical teachers?

b) Are some teachers more stable in their effectiveness than others?

2. Rationale

It is assumed that achievement gain is an acceptable and important criterion of teacher effectiveness at the primary grade level. Correlations involving within-year gains in achievement scores across various curriculum areas and across sex as well as between contiguous years within curriculum areas are taken as indicators of stability. Previously reported stability coefficients (Rosenshine, 1970) were .34 ($p < .001$), .36 ($p < .05$), .20 (ms), and .09 (ns).

A common sense criterion for identifying individual teacher consistency involves the observation of linear consistency in a teacher's set of several yearly mean-adjusted gain scores. Linear trends in the form of improvement or linear decline also represent a form of consistency, but not stability in the sense of linear constancy.

3. Research Design and Procedure

A city school system's records containing grade level equivalent scores for a regular fall administration of Metropolitan Achievement Test batteries were used to compute residual gain scores for second- and third-grade students. The study included all teachers in these two grades who were teaching at the same grade level for all three years of the study, 1967-70. Data were available from 15 Title I schools and 35 non-Title I schools. These data were analyzed separately as different MAT batteries were used in the two types of schools. MAT Primary I (1958) was given each fall in the Title I schools to second graders. Primary II (1958) was administered to Title I third graders and non-Title I second graders. The Elementary (1959) battery was used for Title I fourth graders and non-Title I third and fourth graders. Scores from the fall of grade 2 were the pre-scores for grade 2 and post-scores were taken from the fall of grade 3--which in turn furnished the pre-scores for grade 3. Post-scores for grade 3 were taken from the fourth grade test.

Except for grade 2 in Title I schools, residual gain scores were computed for five subtests: work knowledge (WK), word discrimination (WD), reading (R), arithmetic computation (AC), and arithmetic concepts and problem solving (AR). School records contained only the composite total score for arithmetic on the Primary II and the Primary I battery has only one arithmetic subtest. Subtest pre-scores were used as covariates. Second grade teachers in Title I schools have only one set of data for arithmetic. The composite total arithmetic score was used as a covariate for both AC and AR in grade 3 of Title I schools and grade 2 in non-Title I schools. Residual gain scores were computed for each student ($g = y - (a + bx)$) within sex and within each school year. Data for teachers were compiled by computing means for each of their three respective classes.

Sample sizes were: Title I schools, grade 2, N = 34, grade 3, N = 26; non-Title I schools, grade 2, N = 54, grade 3, N = 51. These N's vary due to incomplete data for certain analyses. Arbitrary cutoff points for class size were used.

Correlational analyses were used to investigate consistency across subtests within year and stability across the two sexes and across the three years within subtest. A second step in a planned series of investigations on teacher effectiveness involved identifying consistent teachers for further study. Similar residual means within a subtest across the three years was taken as a sign of linear constancy.

4. Findings

(a) Consistency across subtests within years. All but 8 of the 108 possible correlations were significant. Forty-eight were significant at the .001 level. Most correlations were moderate to high. Language arts regularly correlated higher with AR.

(b) Consistency within years across sex of student. These correlations were all close to 1.00. An informal analysis of 68 teacher's data revealed only 4 with sizable and consistent sex differences. Two of these did better with boys and two with girls (see Brophy, 1972).

(c) Stability within subtests across time. Stability coefficients were computed for mean residual gains across years 1-2, 2-3, and 1-3. Correlations between contiguous years were generally higher (the medians were: .25 (ns), .42 ($p < .01$), .39 ($p < .05$), and .40 ($p < .01$)) than those obtained by Rosenshine (1970). Coefficients for grade 2, Title I schools, were considerably lower than those for the other grades and schools. Coefficients resulting from within year analysis were generally higher than those from between years.

(d) Identification of consistent teachers. Author judgment was used in observing the degree of linear constancy in the three means within each subtest for each teacher. Twenty-eight percent showed linear constancy, 13% linear improvement, 11% linear decline, and 49% non-linearity.

5. Interpretations

(a) Teachers are more consistent in producing gains within the curriculum areas related to the subtests than across them.

(b) Individual teachers do not tend to be differentially effective with boys versus girls.

(c) Lower stability coefficients for grade 2 of Title I schools may be due to a combination of the age and degree cognitive development of the children (see Hess, 1970).

> An apparent class effect is shown by the generally higher coefficients for the within-year correlations. It is possible than general motivation or classroom atmosphere exert some effect on the achievement of the class as a whole in a given year.

In summary, this study indicates that some grade 2 and grade 3 teachers are more stable in their overall effectiveness. These teachers are about equally effective with boys as with girls. Teacher effectiveness shows great individual differences and only moderate stability from year to year. Observational studies of consistent teachers are needed in order to establish a more rational and valid basis for constructing process observation devices. The multitude sources of instability argue the use of student gain or general achievement for teacher accountability criteria.

Critical Commentary

Teacher accountability is a ve controversial issue. This study helps emphasize the inappropriatene of using student achievement data as a major criterion of teacher effectiveness. One significant contribution of this study is the finding concerning the lack of differential effectiveness regarding the sex of the student.

There are many sources of error and occasion for compromise in a post hoc study involving grade level achievement scores. The researcher mentions several uncontrolled variables including class size, absenteeism, student aptitude, and ability grouping. The report was incomplete in describing the details of some of the analyses, e.g., the computation of residual gain scores and the incomplete description of the sample after arbitrary cutoff points were used. Studies which help to establish ways to identify good teaching are desirable. This study is a step in that direction. Unfortunately, the sources of possible error and the lack of pertinent detail reduces the amount of confidence that one has in the findings.

Terrence G. Coburn
Oakland Schools
Pontiac, Michigan

EJ 109 955 SE 512 613
MEASUREMENT CONCEPTS OF FIRST- AND SECOND-GRADE STUDENTS. Carpenter, T. P. Journal for Research in Mathematics Education, v6 n1, pp3-13, January 1975.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Doyal Nelson, The University of Alberta, Canada.

1. Purpose

To assess (a) the degree to which children rely on perceptual information in measurement operations, and (b) whether measurement comparisons involving equal quantities are as difficult as those involving unequal quantities.

2. Rationale

Carpenter refers to the same basic literature on measurement conservation which indicates that conservation errors are a result mainly of misuse of perceptual information in making comparisons. Previous research of Carpenter and others, however, has indicated that if children are given an opportunity to respond to numerical information in such situations they will respond to this information as readily as they will respond to perceptual information.

By comparing children's performance in measurement conservation situations where all cues are visual with performance in similar situations where cues are either visual or numerical, Carpenter attempts in this study to determine whether conservation errors are a function of perceptual dependence or a result of the child centering on the last available cue.

The visual explanation for conservation errors is that children are unable to see that an increase in one dimension (e.g., height) is compensated for by a decrease in another dimension (e.g., diameter). In this study Carpenter wants to determine whether this inability on the part of children would also apply where units are concerned in measurement. That is, when measuring two equal quantities of liquid with two different units, it will take fewer of the larger unit and more of the smaller unit. Would there be any effect if, on the one hand, units were visibly different and on the other, not visibly different?

Most conservation studies begin with equal quantities, followed by some transformation, after which children are required to judge whether the quantities are still equal or not. The literature involving the use of unequal quantities is confusing. By controlling relations between quantities at the outset of conservation and measurement tasks, Carpenter hopes to resolve some of the existing confusion. To do this he compares children's performance in situations where:

1. Equal quantities are transformed to appear unequal: Equivalence (E);

2. Unequal quantities are transformed so that the dominant dimension in the situation is not changed: Non-Equivalence I(N_1);

3. Unequal quantities are transformed so that the direction of the inequality appears to be reversed: Non Equivalence II(N_2).

3. Research Design and Procedure

To accomplish the purposes outlined for this study, Carpenter designed 13 conservation and measurement items all of which involved pouring liquids or measuring liquids from one container to another.

There were five basic sets of items each involving either two or all three of the relations just outlined:

Set 1 - Classical conservation items in which liquid was poured from one of two containers into a taller, narrower one.

Set 2 - Similar to the first except that the two quantities of liquid were measured into opaque containers using different-sized units of measure.

Set 3 - Identical to the second set but in this case the units were different but not perceptibly so.

Set 4 - Liquids were measured from opaque containers into different shaped containers using a single unit of measure.

Set 5 - Same as the fourth set but liquids were measured from different shaped containers with a single unit into identical opaque containers.

The first three sets of problems contained three items: an E item, an N_1 item and an N_2 item. The fourth and fifth set each contained an E item and an N_2 item.

The subjects for the study were 75 first- and second-graders selected from a single elementary school. The items were divided into two parts and subjects randomly assigned to each part. There were 61 subjects in part A and 68 subjects in part B. Subjects in part A did items in Sets 1, 2, and 3 (9 items), while subjects in part B did items in Sets 2, 3, 4, and 5 but only with E and N_2 relations (8 items).

The null hypotheses tested in part A were as follows:

H_1: There is no significant difference between performance on E and N_2 problems for any of the problem types.

H_2: There is no significant difference between performance on E and N_1 measurement problems in which the larger unit of measure is not identifiable visually.

H_3: There is no significant difference between performance on E and N_1 items for conservation problems or for measurement problems in which the larger unit is identifiable visually.

H_4: There is no significant difference between performance on conservation problems and corresponding measurement problems.

H_5: There is no significant difference between performance on measurement problems in which it is possible to identify visually the larger unit and those in which it is not.

Hypotheses H_1 to H_5 were also tested in part B.

For part B the following were also tested:

H_6: There is no significant difference in performance between measurement problems in which correct visual cues are followed by distracting numerical cues and corresponding problems in which correct number cues are followed by distracting visual cues.

H_7: There is no significant difference in performance between measurement problems in which the correct measurement cues appear before distracting visual cues and those in which they appear after the distracting visual cues.

Items were administered to each child individually in a small room apart from the classroom. Procedures and protocols for items were kept as consistent as possible from subject to subject.

The following controls were established:

1. Order of items was randomized for each subject.

2. Some children were asked on a random basis whether the quantities were more or the same; others were asked whether they were the same or more. Also, in non-equivalence problems some children measured the smaller quantity first and others measured the larger quantity first.

4. Findings

The item means for part A and part B are given in the tables below:

Table 1

Item Means for Part A

Problem Type	Relation		
	Equivalence	Nonequiv. I	Nonequiv. II
Conservation	.41	.56	.39
Measurement with distinguishably different units	.43	.57	.41
Measurement with indistinguishably different units	.33	.34	.33

Table 2

Item Means for Part B

Problem Type	Equivalence	Nonequivalence
Measurement with distinguishably different units	.32	.38
Measurement with indistinguishably different units	.16	.19
Measurement into different shaped containers	.69	.71
Measurement from different shaped containers	.94	.85

A summary of the results reported for part A is given first.

Nearly all of the errors in part A, according to the experimenter, resulted from children responding strictly on the basis of number of units. Eight subjects could measure successfully when the larger unit was distinguishable from the smaller, but seven of them were unsuccessful when the size of the unit could not be so dintinguished. All eight subjects gave compensation as one reason for their response when the response was correct.

An analysis of variance procedure was used to test the various hypotheses and a summary of the results reported follows:

1. There were no significant differences between E and N_2 relations in any set of items or between any of the relations in which size of units could not be distinguished.

2. There were no significant differences between N_1 and the other two relations (E and N_2) for the conservation items and for those measurement items in which the units could be distinguished.

3. There were no significant differences in the two measurement problems for E and N_2 relations or between the measurement problems in which the units could be distinguished or in the conservation items for any of the relations.

For part B, Carpenter indicates that most of the errors resulted from subjects responding strictly to number of units. Only two subjects in items which involved an indistinguishably larger unit could use measurement information to identify the larger unit. A compensation rationale was used by only two subjects of the five who measured with distinguishably different units.

The analysis of variance results were reported in a table and summarized as follows:

> There were no significant differences for E and N_2 relations, but there were significant differences between all four measurement problems. The items in which liquids were measured with the same unit into different-shaped containers were easier than corresponding conservation items in part A.

5. Interpretations

Carpenter maintains that a significant number of children in the study centered on a single dominant dimension but that the order in which they received the cues was a major factor in determining which cue they attended to. He further maintains that children are not dominated by perceptual qualities of an event and conservation errors are not simply a result of being distracted by the perceptual aspects of conservation and measurement tasks. Children, he says, can as easily be distracted by the numerical aspects of the situation as by the visual and problems in which the correct cues are numerical are easier than those in which the correct cues are visual.

Failure to find significant differences between items involving visual transformations leads Carpenter to hypothesize that whereas the type of comparison (number, weight, volume) will affect the difficulty of the conservation problem, the type of transformation (numerical or visual) will not.

The differences in results for part A and part B on items in which the units were perceptually different and on those in which they were not leads Carpenter to suggest caution in interpreting how children will use compensation in explaining conservation and measurement phenomena. He suggests that interaction with other tasks may effect the role of compensation. Since that aspect of the study which dealt with conservation and measurement situations involving both equal and unequal quantities showed no striking differences, Carpenter suggests that for assessment or instruction one could as well begin with unequal quantities as with equal ones.

Carpenter states that Grade 1 and 2 children can give some meaning to the measurement process. However, he points out that many children who could interpret measurement results when there was no conflicting cue would abandon the results of measurement in the face of conflicting visual information.

Only about 1 in 4 subjects appreciated the significance of a single unit of measure in making comparisons. He suggests that children should be given experience in measuring equal quantities with different units. He warns, however, that such experience may lead children to the conclusion that the quantities being measured were, after all, not equal. The act of measurement cannot be relied upon to make children abandon their reliance on visual comparisons.

Finally, Carpenter claims that notions of measurement occur earlier than is suggested by previous studies. He points out that this may be because previous research did not provide any measurement cues for children to respond to.

Critical Commentary

Carpenter has addressed some important aspects of the development of measurement concepts on the part of young children. In general, the interpretations of results are insightful and convincing.

There are, however, some discrepant results in the two parts of the study which have not been adequately explained. For example, the item means in part A for E and N_2 relations and with indistinguishably different units are both .33. The corresponding means in part B are .16 and .19 respectively. Why should subjects in part B find these items so much more difficult than did subjects in part A? Some reference is made in the discussion to the possibility that interaction might be involved. From the information given, there appear to be equally likely competing explanations for the discrepancies. The only way to resolve the problem would be to replicate the study.

A major criticism of the report of this study is that not enough information is provided to permit replication. Protocols are described only vaguely. Indeed, one cannot be completely sure whether the experimenter performed all the transformations for the children or whether the transformations were carried out by the children themselves. From the general descriptions of the sets of items it is not at all clear how the specific items in the sets were adjusted to accommodate the three relations. Much of the confusion that exists in mathematics education research might be eliminated if reports of investigations would include very detailed descriptions of protocols and procedures used in collecting data.

There is a point that should be made about the conclusions. Some of them appear to be somewhat stronger than the data warrant. For example, the statement that "young children are not dominated by the physical qualities of an event" is not entirely supported in this study. Indeed the author later says, "measurement will not convince many children of the inaccuracy of their visual comparisons." Also in part B, when units were perceptibly different, children made fewer errors. The direction of differences also applied to part A.

Doyal Nelson
The University of Alberta
Canada

ED 111 681 SE 019 680

THE PERFORMANCE OF FIRST- AND SECOND-GRADE CHILDREN ON LIQUID CONSERVATION AND MEASUREMENT PROBLEMS EMPLOYING EQUIVALENCE AND ORDER RELATIONS. Carpenter, Thomas P. In Steffe, L. P. (ed.) Research on Mathematical Thinking of Young Children: Six Empirical Studies. Reston, Virginia: NCTM, 1975.

Expanded Abstract and Analysis Prepared Especially for J.M.E. by Edward C. Rathmell, Iowa Area Education Agency VII.

1. Purpose

The major stated purposes of this study involving liquid quantities were (1) "to assess the degree to which young children possess the logical structures to assimilate and apply information from measurement processes," and (2)"to identify some of the factors involved in the development of measurement and conservation." In addition, the effects of equivalence and order relations on conservation and measurement tasks and the relation between responses to visual and numerical cues were investigated.

2. Rationale

Piaget and others have documented the effect that visual cues have on the young child's responses to measurement and conservation tasks. Carpenter found evidence to support the idea that children also rely on numerical cues. It was hypothesized that children respond to the last cue available regardless of whether it is visual or numerical.

While no clear-cut differences in difficulty have been found between equivalence and nonequivalence tasks, judgments of equality appear to be more difficult and a stable concept of nonequivalence may precede a stable concept of equivalence. This is in direct contradiction to the logical development in which the concept of nonequivalence presupposes the concept of equivalence.

3. Research Design and Procedure

The study involved three different types of order and equivalence relations and five different types of conservation and measurement problems with visual or numerical distractions. The three relations were (1) Equivalence, where equal quantities were transformed to appear unequal, (2) Nonequivalence I, where unequal quantities were transformed so the distracting cue indicated equality, and (3) Nonequivalence II, where unequal quantities were transformed to make the inequality appear to be reversed. The five different problem types were (1) Conservation of continuous quantity, (2) Measurement with visibly different units, (3) Measurement with indistinguishably different units, (4) Measurement from different-shaped containers with the same unit, and (5) Measurement into different-shaped containers with the same unit.

The sample consisted of 75 first graders and 54 second graders from a rural community in Wisconsin. The Ss were randomly assigned to one of two groups, 61 to Part A and 68 to Part B. The Ss within each group were asked to solve the same problems; however, the order of the problems and certain procedures within the problems were randomly varied. The problems were administered in two sittings. Responses were judged correct or incorrect without regard for explanation, although reasons for the responses and the types of errors were also recorded.

The Ss in Part A responded to problems involving each of the three relations for the Conservation problems and for the Measurement problems where different units were used. A 3 x 3 repeated measures design was used to analyze Part A. The Ss in Part B responded to Equivalence and Nonequivalence II relations for all the Measurement problems. A 2 x 4 repeated measures design was used to analyze Part B. A variety of hypotheses were tested using a multivariate analysis of variance with single degrees of freedom-planned contrasts.

4. Findings

Part A: There were no significant differences between Equivalence and Nonequivalence II relations for any of the problem types. The mean score for Nonequivalence I relations was significantly greater than the mean scores for Equivalence or Nonequivalence II relations, except in the case of Measurement problems with indistinguishably different units. No significant differences were found between the Measurement problems or between the Conservation and Measurement problems.

Part B: There were no significant differences found between Equivalence and Nonequivalence II relations. Significant differences were found between each of the four types of Measurement problems. Problems where one unit was used were significantly easier than problems where different units were used.

In general, there were no differences due to sex, order of the items, or protocol variations.

5. Interpretations

The results clearly show that misleading numerical cues cause students to make the same errors in conservation tasks that visual cues do. Furthermore, problems in which correct numerical cues are followed by misleading visual cues are significantly easier than either problems in which (1) both cues are visual or (2) correct visual cues are followed by incorrect numerical cues. While not the only factor, it appears that children are most likely to attend to the last cue available.

The type of relations between quantities, equivalence or nonequivalence, apparently does not affect performance. Nonequivalence I problems were easier, but they do not require genuine conservation. Final states can be used to make correct comparisons.

By the end of first grade, virtually all children have some meaning for measurement operations. They recognize that two quantities with the same number of units are equivalent. Similarly, if or · quantity measures more units than the other, it is greater. However, m children at this state are still unable to apply these measurement proc :es to conservation tasks.

"It appears that centering on a single dominant dimension is the major reason for most conservation and measurement failures and the development of conservation and measurement concepts can be described in terms of increasing ability to decenter."

Critical Commentary

This study was well conceived and designed. It provides important evidence that indicates the extent to which young children rely on numerical information and are able to apply measurement processes.

Some primary teachers place a heavy emphasis on counting and other number skills too early. On the other hand, some interpreters of Piagetian research suggest that number work out to be postponed. This study indicates that young children are able to use number ideas in many situations. We need to learn when and how to exploit this knowledge. Studies like this one will eventually help us find an appropriate curricular balance between the two extremes mentioned above.

Little progress was made in determining the relationship between the child's development and a logical development of equivalence and nonequivalence concepts. This relationship might well be more easily determined by longitudinal studies closely following the development of individuals or by training studies.

Overall the report was well done; however, Table 1 was not included. Also, it makes sense to the abstractor that H3 in Table 3 should be corrected to read DE + CE = D1 + C1 and H5 in Table 7 should be corrected to read D = I.

Edward C. Rathmell
Iowa Area Education Agency VII

EJ 090 366 TM 501 065

COMPARATIVE VALIDITIES AND TESTING TIMES FOR COMPOSITES OF STRUCTURE-OF-INTELLECT TESTS AND PREVIOUS MATHEMATICS GRADES WITH COMPOSITES OF TRADITIONAL STANDARDIZED TEST MEASURES AND PREVIOUS MATHEMATICS GRADES IN PREDICTION OF SUCCESS IN HIGH SCHOOL MODERN ALGEBRA. Holly, K. A.; Michael, W. B. Educational and Psychological Measurement, v33, pp915-919, 1973.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by John Gregory, University of Florida.

1. Purpose

The objective of the study was to provide more detailed and specific information derived from a previous study[1] regarding comparative validities and testing time requirements of selected composites of measures (standardized test scores and GPA in 8th grade mathematics vs. structure-of-intellect tests and GPA in 8th grade mathematics) in predicting success in high school algebra.

2. Rationale

Although no related studies except the authors' original one are cited, the rationale seems to stem from a desire to discover more efficient alternatives to traditional commercial standardized tests used in predicting success in modern algebra.

3. Research Design and Procedure

Since this is a re-analysis of data from a previously reported investigation, description of the population, sampling and testing procedures are sketchy. It is reported that 177 students taking algebra in a middle-class suburban secondary school provided the scores. Test administration procedures are not described but the performance measures were derived from the Cooperative Mathematics Tests, Algebra (CMT); nine scales of the California Achievement Test; and three scales of the California Test of Mental Maturity. Grade point average (?) for eighth-grade mathematics and for algebra plus fifteen measures of Guilford's structure-of-intellect factor abilities (these SI tests are not described) were also utilized. The GPA for algebra and performance on the CMT constituted the criterion variables. Prediction of each of these two criterion variables was achieved through step-wise multiple regression analyses using composites of predictor variables. These composites are of two types:

[1] Holly, K. A. and Michael, W. B. The relationship of structure-of-intellect factor abilities to performance in high school modern algebra. Educational and Psychological Measurement, v32, pp447-450, 1972.

Type A: selected combinations of commercial tests alone and selected combinations of commercial tests and 8th grade mathematics grades;

Type B: selected combinations of SI tests alone and selected combinations of SI tests and 8th grade mathematics grades.

The basis for composite selection is not provided.

4. Findings

For both criteria of success in algebra, the Type B composites (those involving SI factors) produced higher multiple R's (ranging from .56 to .60) than those of Type A (involving commercial standardized tests) for each of thirty analyses (21 for GPA, 9 for CMT).

Data collection via SI tests which predict the criterion variables took less time than standardized test administrations.

5. Interpretations

The investigators conclude that the use of SI tests provide promising alternatives to commercial standardized tests in predicting success in modern algebra.

Critical Commentary

Although the article was not directed to researchers in mathematics education in particular, the following comments and questions may suggest a closer relationship to mathematics education research efforts.

1. Prediction of success in algebra has limited value in high school mathematics department. But the discovery of non-static predictive variables could lead to the identification of teaching practices which enhance algebra learning. This possibility leads to three questions relative to the study under review.

a. Can the SI factor abilities be enhanced through special training? If so, what educational practices seem to improve these abilities?

b. Can investigations be constructed to determine if the SI factor abilities are directly responsible for performance in algebra or are they strictly correlational?

c. Can the SI factor abilities be further specified so as to determine their function relative to algebra success?

2. GPA or any measures based upon grades may serve as an indicator of algebra ability, but other measures seem to offer more reliable data. Although this study utilized a standardized test in addition to GPA as a criterion variable, the following questions come to mind.

a. Do SI factor abilities predict success for specific algebraic skills measured by standardized tests or other less global tests?

b. Are SI factor abilities related to the learning of other mathematics?

c. Would the same predictive qualities of the SI factor abilities result if a larger sample, receiving fewer tests, were utilized?

d. Are SI factor abilities interactive with specific mathematics instructional treatments?

John Gregory
University of Florida

ED 111 681 SE 019 680

LEARNING OF CLASSIFICATION AND SERIATION BY YOUNG CHILDREN. Johnson, Martin L. In Steffe, L. P. (ed.), Research on Mathematical Thinking of Young Children: Six Empirical Studies. Reston, Virginia: NCTM, 1975.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Douglas A. Grouws, University of Missouri-Columbia.

1. Purpose

The main purpose of this study was to determine the influence of training on the ability of first and second grade children to (1) classify and seriate objects on the basis of length, and (2) conserve length and use the transitive properties of length relations. Also investigated was the relationship between the ability to use the transitive property of the "same length as" ["longer than" and "shorter than"] relation and the child's ability to classify [seriate] on the basis of the relation.

2. Rationale

Some previous research has shown that classification performance can be improved through training although the results of other studies have not been favorable. The author points out that in previous studies classification has been approached as a general categorizing process and not as the forming of equivalence classes, and thus any relationship between mathematical properties of an equivalence relation and classification based on that relation has not been explicated. Little training research aimed at facilitating seriation ability has been done and favorable results in such studies have not isolated the relative influence of age increases and instructional activities.

At a theoretical level, classification and seriation can be analyzed psychologically, as Piaget and his associates have done through postulated models of cognition, and mathematically where these processes can be interpreted as being logically dependent on equivalence and order relations. Relationships between components of these two frameworks and the ability to classify and seriate are of interest and a focus of this study.

3. Research Design and Procedure

Thirty-nine first-grade children and forty-two second grade children selected from two schools formed the sample for this study.

Two instructional units were developed. Unit I consisted of six lessons designed to acquaint students with the "same length as," "longer than" and "shorter than" relations and to make proper comparisons based on these relations. Unit II consisted of ten twenty-minute lessons designed to give experience in classifying using the "same length as" relation and in seriating using the two order relations.

A criterion test administered at the conclusion of instruction on Unit I was used to insure that all children in the study understood the relations and terms to be used in subsequent testing. Children meeting criterion were randomly placed into either an experimental or control group. The experimental group received instruction on Unit II while the control group received no further instruction. All instruction was given to children in small groups by the researcher.

Four additional tests were developed: a 6 item scale measuring conservation of length, a 6 item scale measuring transitivity of length relations, a 12 item scale measuring the ability to seriate, and a three item scale measuring the ability to classify. The conservation and transitivity tests were used as both pretests and posttests whereas the seriation and classification tests were used as posttest only.

Data were analyzed by performing a 2 x 2 x 2 (School X Grade x Treatment) ANOVA on seriation test scores and a 2 x 2 x 2 ANCOVA on conservation and transitivity test scores using pretest scores as covariates. The classification data were analyzed item by item for treatment effects using contingency tables and Chi-square tests of independence. To determine relationships between transitivity, seriation, and classification, contingency tables were also constructed and analyzed.

4. <u>Findings</u>

Students who received training significantly outperformed students in the control group on the seriation test, but there was no significant difference in their performance on the conservation of length test or on the transitivity of length relations test.

Second grade children significantly outperformed first grade children on the seriation and conservation tests but not on the transitivity test.

There was a significant school effect on both the conservation and transitivity tests.

The item by item analysis of the classification test revealed no significant relationships ($p < .05$) between treatments and test performance.

The author states that performance on classification items one and two (but not item three) was "slightly related" ($p < .10$) to the ability to use the transitive property of the "same length as" relation. No relationship was found between seriation ability and the ability to use transitive properties of appropriate order relations.

5. Interpretations

The results of this study support the hypothesis that the ability to seriate "linear" objects increases with age and can be improved by training. Whether this increased performance represents the learning of an algorithm or more children becoming operational in the Piagetian sense could not be determined.

Training on seriation had no effect on transitivity ability; in fact, no significant relationship could be detected. If the children were operational this evidence contradicts Piaget's hypothesis that seriation behavior implies transitivity. If, however, the children were not operational then no meaningful conclusion about the hypothesis can be drawn.

No relationship between classification ability and transitivity was found. The author implies that this was probably due to transitivity not being necessary for the classification items used in the testing.

Training did not improve performance on the conservation test. This result was unexpected and contrary to the results of several previous studies.

Critical Commentary

The author should be commended for relating his study to previous research and placing it in a theoretical context.

A number of the results of the study were not those anticipated and consequently interpretations were difficult to make. The author seemed, however, much more inclined to fault his own work than to cast theoretical doubts. This was not always warranted. Resolution of many of the issues raised will require additional research as the author indicates.

This line of research may be more fruitful in the future as more sensitive instruments are developed. One wonders about the reliability (not in the internal consistency sense) of such tests, given their brevity and the age of the children to which they are administered. The value of retention tests should also be given serious consideration in research of this type.

Douglas A. Grouws
University of Missouri

WHEN IS NEGATION EASIER THAN AFFIRMATION? Johnson-Laird, P. N. and Tridgell, J. M. Quarterly Journal of Experimental Psychology, v24, pp87-91, 1972.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by H. Laverne Thomas, State University of New York at Oneonta.

1. Purpose

This study investigated the relative difficulty of drawing the correct conclusion from the premises for a disjunctive syllogism (P or Q, $\sim$Q, therefore, P) where the statements Q, $\sim$Q have one of three forms: (1) Q is an affirmative, $\sim$Q is its logical denial, e.g., "the ball is red" and "the ball is not red;" (2) Q is an affirmative as in (1) but $\sim$Q is formed using an antonym, e.g., "the ball is green;" (3) Q is a negative and $\sim$Q is the corresponding affirmative, e.g., "John is not rich" and "John is rich." The three forms are described respectively as "appropriate negative," "affirmative," and "inappropriate negative."

The research hypothesis was that the three types of problems would have increasing order of difficulty with type (1) the easiest.

2. Rationale

Previous research on the understanding of negative and affirmative sentences has indicated that negative sentences are the more difficult to understand. The investigators propose that in "daily life" negatives do not pose the difficulties seen in research studies. The thesis is advanced that in normal situations there is some plausible conception such as "John likes Mary" that the negative is called on to deny. In experiments, the lack of such contextual support could add to the difficulty of analyzing negative sentences. The study reported is related to studies showing support for this thesis.

3. Research Design and Procedure

The independent variables were problem type (1, 2, 3), lexical contents, and test (two presentations of each problem type). Dependent variables were measured time to response and number of errors. The three types of problems were randomly ordered to construct six forms for the test (three items each). A second test was constructed by mirroring the three item types in three new problems, resulting in six test-test combinations. From the 24 subjects, four were randomly assigned to each of the six test-test combinations. These subjects were college undergraduates who had not worked with this type of task and who had not studied formal logic.

The six test-test forms were identical with respect to the order of appearance of the names objects and their traits (lexical contents). Three boys' and three girls' names represented objects and three pairs of traits were employed. Test administration was individual and in a single session. The subjects were told that the test involved reasoning and that they were to respond as "quickly as was compatible with drawing a correct conclusion." Each problem was read aloud by the experimenter and the time to response then measured. A single practice problem of another logical variety was employed.

Mean response times in seconds were reported for each type of problem on each test and overall. Total error figures by problem type were reported also. The sign test was employed to analyze the order of difficulty in terms of time to response. Analysis of variance was carried out on the mean times to response with respect to problem type, test presentation and lexical contents (objects and traits employed).

Further, individual response patterns were scored in such a manner as to indicate the degree to which they did or did not conform to the hypothesized order of difficulty. A sign test was employed to analyze these scores.

4. Findings

The sign test on overall mean response times showed a highly significant trend in favor of the hypothesis ($P = 0.001$). The analysis of variance confirmed the apparent significant differences between mean response times for the different problem types ($P < 0.001$), but there were no significant differences attributable to lexical contents or test presentation. The sign test performed on the response pattern scores showed a "reliable trend" in favor of the hypothesis ($P = 0.003$).

Some related remarks were that the main error consisted in stating the negative of the correct conclusion and that on the second presentation the "inappropriate negative" took more time but there were fewer errors. Many subjects complained that statements such as "Either John is intelligent or he is not rich" are in some way invalid or ungrammatical. Some subjects complained (irrelevantly) that it wasn't clear whether or not the disjunctions were mutually exclusive.

5. Interpretations

The results of this study were contrasted with other studies and it was noted that in such studies it has often been the task to match the semantic content of a statement with a picture or other situation. Here, the emphasis is on a one-to-one correspondence in meaning between the sentence and the picture. In this study, however, the key strategy was the identification of an inconsistency between the minor premise and one part of the major premise. Thus the more explicitly the inconsistency appears, the easier the task.

The preceding explains the greater difficulty of the "affirmative" problems wherein contrary but affirmative statements are employed but not the even greater difficulty of the "inappropriate negative" problems. It is speculated that subjects lose track of attributes and the overall argument in dealing with the double negative involved, focusing on Q and $\sim(\sim Q)$. Thus, the minor premise is thought of only in terms of simply denying a part of the major premise rather than ruling it out in favor of the alternative. Finally, the investigators remark that "it should not surprise us that the proper function of affirmatives is to make assertions, and of negatives to make denials."

Critical Commentary

The research reported here is of obvious interest to mathematics educators and again illustrates the close ties between researchers in psychology and in mathematics education. Perhaps the most closely related research among mathematics educators is that of O'Brien and Shapiro (AERJ, vol. 5, pp. 531-542, 1968) and, more recently, Eisenberg and McGinty (JRME, vol. 5, pp 225-237, 1974).

In reading the report, several questions of interest occurred to the abstractor:

(1) Were the problems available to the subjects in written form?

(2) Were the forms P or Q, $\sim$P and $\sim$P or Q, P employed?

(3) What were the error totals by item type for each test?

(4) What was the distribution of response patterns?

(5) What were the specific elements and data for the analysis of variance?

Since the variations on the disjunctive syllogism employed had such a strong effect, it would appear reasonable to investigate other variations in a similar way. Also, inclusion of premises such as P or Q, P which do not have a necessary consequent could yield further knowledge of learner difficulties with logical analysis and the role of negation.

The investigators suggest a difficulty with the affirmative problems that should be echoed here. Is the logical negative of "John is rich" to be "John is poor?" Can this cause difficulties related to subjects' remarks about mutual exclusivity? Could this hint as to the availability of a third alternative have entered into the difficulties with the inappropriate negative problems?

The bias of the abstractor is that useful information in studies such as this can be obtained by systematically questioning subjects as to their thinking in completing the tasks. Such information adds a valuable dimension by giving additional explanatory power and insight for the generation of new hypotheses.

H. Laverne Thomas
State University of New York at Oneonta

EJ 068 476 SE 507 560
A FORMATIVE DEVELOPMENT OF AN ELEMENTARY SCHOOL UNIT ON PROOF. King, Irvin L. Journal for Research in Mathematics Education, v4 n1, pp57-63, January 1973.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Robert D. Bechtel, Purdue University, Calumet Campus.

1. Purpose

To develop a unit of instruction on proof for use with capable sixth-grade students, and in so doing, to examine the feasibility of the Cambridge Conference on School Mathematics (CCSM) proposal that proof be taught in the elementary school.

The manner in which the study was conducted exhibited other purposes:

1) to demonstrate the effectiveness of a particular iterative curriculum development model; and

2) to demonstrate the effectiveness of using "mastery-learning" procedures in teaching the unit to a small group.

2. Rationale

This study is tied by the investigator to the CCSM Report. The recommendation cited is that the study of mathematical proof should begin in the elementary school. "Although many psychologists assert that the early adolescent possesses the cognitive structures necessary for formal reasoning, there is a scarcity of empirical evidence concerning the ability of elementary school children to learn mathematical proofs." This study was conducted in response to a need for experimentation in this area.

3. Research Design and Procedure

Six theorems of the kind recommended in the CCSM Report were selected for the unit.

Theorem 1. If $N \mid A$ and $N \mid B$, then $N \mid (A + B)$.

Theorem 2. If $N \mid A$ and $N \mid B$, then $N \mid (A - B)$.

Theorem 3. If $N \mid A$ and $N \mid B$ and $N \mid C$, then $N \mid (A + B + C)$.

Theorem 4. If $N \nmid A$ and $N \mid B$, then $N \nmid (A + B)$.

Theorem 5. If $N \nmid A$ and $N \mid B$, then $N \nmid (A - B)$.

Theorem 6. There is no largest prime number.

A four-stage curriculum development model (Romberg and DeVault, 1967) was used in developing the unit. This study carried out the first two phases, analysis and pilot examination. Content and instructional analyses led to the first pilot examination. This was followed by another analysis and a second pilot trial. Formative evaluation procedures (Scriven, 1967) were used in the two pilot phases to assess the effectiveness of components of the unit. Six sixth-grade students were the subjects of the first pilot phase. The second pilot study was conducted with two fifth-grade and five sixth-grade students.

After the second pilot study had been completed a Nonequivalent Control Group Design (Campbell & Stanley, 1968) was used in testing the unit. An experimental group of 10 students (mean IQ score 117) was selected from an intact classroom, and a control group (mean IQ score 121) was selected by matching procedures from the other two classes. Sex, Henmon-Nelson IQ scores, previous math grades, and teacher evaluations were the variables considered. A 25-item test containing both prerequisite skills and the proofs of the six theorems was administered to both groups prior to instruction. During the experiment, the control group attended the regular math class, where they studied multiplication and division of decimal fractions. A certified elementary teacher, whose academic preparation included six semester hours of mathematics beyond calculus, taught the unit to the experimental group. She employed "mastery learning" procedures. After 17 days of instruction, the test was readministered to both groups.

4. Findings

The results of the posttest showed that the experimental group learned 96% of the prerequisite skills and 97% of the proofs (pretest scores were 29% and 0%, respectively, for these categories), whereas the control group knew 32% of the prerequisite skills and none of the proofs (pretest scores were 30% and 0% respectively). An analysis of variance confirmed the obvious.

In addition to the posttest, an attempt was made to determine the extent to which the students understood the meanings of the theorems and proofs. At the conclusion of the experiment each student in the experimental group was asked to perform the following tasks in an interview situation: (a) give numerical examples to illustrate the meaning of the theorems (all students were able to do this); (b) explain and defend each step in the proofs (nine of 10 students were able to do so for Theorems 1, 2, and 3, seven were able to do so for Theorems 4 and 5, and eight were able to do so for Theorem 6); (c) apply the theorems to a given set of divisibility facts (all students were able to do this).

5. Interpretations

The investigator concluded from the study that:

1. It is possible to teach capable sixth-grade students to reproduce and explain the proofs of the theorems contained in the unit.

2. Since the feasibility of the unit on proof has been established, the effectiveness of the basic approach has been demonstrated.

3. Mastery learning can be very effective in a small-group situation.

The investigator proposed the following recommendations for further study:

1. The formative development of the materials in the unit should be continued in an attempt to determine if the unit will be effective with other groups of average and above-average sixth-grade students.

2. It might be profitable to continue experimentation with other proof materials at other grade levels. For example, proving the divisibility rules for 3 and 9 might be attempted at Grade 5.

3. The curriculum-development model should be validated with other subject matter.

4. The mastery-learning procedure employed in this study was very effective. This suggests that there are many researchable questions in this area relative to curriculum materials (development and utilization) and student variables (achievement, aptitude, attitude, retention, learning styles, and so on).

Critical Commentary

The investigator's remark: "The analysis phase and the pilot examination phase of the curriculum development model [the complete model for this study] tended to be more exploratory than experimental in nature, the primary concern being to find materials and procedures that worked in the classroom," showes the tenuous nature of any conclusions that can be drawn from the study. Developing curriculum materials is serious business, and the efforts described in this study - writing, testing, rewriting, and retesting prior to disseminating the materials - are to be commended and fostered.

However, one should seriously question whether this "experiment" shows the feasibility of teaching proof in the elementary classroom. First of all, does teaching proof mean to teach the proof of a theorem in an almost programmed style, and then have the student reproduce the given proof and the proof with minor variations in proving corollaries? If a contradiction is involved in the argument then showing the proof of a theorem seems foreign to the natural expectation of generating several proofs for the same theorem. Hence, there is a philosophical question of the compatibility of teaching proof in conjunction with a carefully sequenced instructional unit.

In retrospect the choice of the control group in the study could have been improved. It would seem that both the control and experimental groups could be selected from students who possess the prerequisite skills. The divisibility symbolism in the study seemed initially to be quite foreign to all the students. Students could not "read" the pretest items with understanding. One means of alleviating this difficulty would be to use mathematical language that is familiar to the students when writing the pre-test. For example, write "the number A is a factor of the number N" instead of "$A \mid N$." New symbolism which needed to be introduced could be taught to all students prior to administering the pre-test. A student who can read a theorem might be able to prove the theorem. A student who cannot read a theorem has no chance of proving the theorem.

As was stated by the investigator, "the study was exploratory." Further experimental studies on the feasibility of teaching proof in the elementary schools are appropriate. As stated in the Cambridge Report, and quoted in the study, "More information through experiment is needed."

Robert D. Bechtel
Purdue University
Calumet Campus

EJ 109 956 SE 512 614

DEVELOPMENTAL ASPECTS OF CHILDREN'S ABILITY TO UNDERSTAND MATHEMATICAL PROOF. Lester, F. K. Journal for Research in Mathematics Education, v6 n1, pp14-25, January 1975.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Nicholas A. Branca, Pennsylvania State University.

1. Purpose

To examine the development of the ability to write a correct mathematical proof by investigating certain developmental aspects of problem-solving abilities in a specific mathematical system.

2. Rationale

By considering the making of a mathematical proof as a complex problem-solving task, an analysis of the psychological processes involved can be undertaken by investigating several quantitative variables that have been identified as measures of problem-solving performance. This study considers the relation between problem-solving ability and chronological age as a primary factor and by focusing on basic research intends to help build a theory of mathematical problem-solving.

3. Research Design and Procedure

The sample for this study consisted of 80 subjects ranging in age from 6 to 18 years. Four groups, designated A_1, A_2, A_3 and A_4 were formed by selecting 20 students from middle class schools at each of the following respective grade levels; 1-3, 4-6, 7-9, and 10-12.

The experimental task was based on an altered version of a mathematical system devised by Suppes that was concerned with the generation of finite strings of I's and O's according to five inference rules and one axiom. The existence of the symbol I is the axiom of the system, Rule 1 allows two O's to be placed at the right end of any nonempty string; Rule 2 allows two I's to be placed at the right end of any nonempty string; Rule 3 allows the removal of an O from the right end of a string if and only if it is immediately preceded by I-O, which in turn must be preceded by a nonempty string; Rule 4 allows the removal of an I from the right end of a string if and only if it is immediately preceded by O-I, which in turn must be preceded by a nonempty string; Rule 5 allows the removal of an I from the left end of of a string if there are at least two letters in the string.

Mathematically the rules can be stated as follows where S denotes a nonempty string of I's and O's:

Rule 1: S → S-O-O
Rule 2: S → S-I-I
Rule 3: S-I-O-O → S-I-O
Rule 4: S-O-I-I → S-O-I
Rule 5: I-S → S

A theorem of the system is any nonempty finite string of I's and O's that can be obtained from the axiom by a finite application of rules 1-5. In this experiment subjects were given instruction on the axiom and rules of the system and asked to supply proofs of theorems limited to strings of from one to five letters. The tasks were presented to the subjects on a computer terminal which kept track of seven criterion variables identified as being appropriate measures of problem-solving performance. These variables were: number of theorems proved (abbreviated TP), number of theorems attempted (TA), number of incorrect applications of rules of inference (IAR), number of trials beyond minimal length per theorem proved (TB), trial difficulty per theorem attempted (TD), presolution time per theorem proved (PST), and total time per theorem attempted (TT).

Two one-way multivariate analyses of variance were performed to test the null hypotheses that there are no significant differences in mean performance among groups A_1, A_2, A_3, and A_4 on each of the seven criterion variables.

4. Findings

A significant difference among the four groups was found on four measures of problem-solving ability. The following table summarizes these significant findings.

Summary of Significant Results from Multiple Comparisons of Groups Using the Tukey Procedure

Variable	Significant Results	Level of Significance
TP	$A_1 < A_4$	.01
	$A_1 < A_3$	.01
IAR	$A_4 < A_1$	.01
	$A_3 < A_1$	.05
	$A_2 < A_1$	.01
TD	$A_4 < A_1$	.01
	$A_3 < A_1$	.01
TT	$A_4 < A_2$	.01
	$A_4 < A_1$	.01
	$A_3 < A_2$	.01
	$A_3 < A_1$	.01

5. Interpretations

The author presents the following three possible limitations to his study for consideration:

1. proofs in the experimental mathematical system are restricted to the generation of finite strings of I's and O's, and may not be representative of the processes that are involved in more complex systems;

2. the tasks were not difficult enough to make evident the differences in problem-solving ability of subjects of different ages or they do not involve logical reasoning beyond the concrete operational stage;

3. only quantitative measures of problem-solving behavior were used, and these particular criterion variables were not the only ones, or even the best ones, that should have been considered.

With these in mind it is stated that subjects of junior high school age are as capable of solving problems in the experimental mathematical system as are subjects of senior high school age. Also, subjects in the upper elementary grades (4-6) are able to solve problems in this system just as successfully as the older subjects, except that they require more time. The author calls for future research with an emphasis on an examination of problem-solving strategies and analysis of the correlation between performance and the subject's Piagetian stage of development.

Critical Commentary

This study is well-designed, focusing upon important basic questions of problem-solving abilities. The author is careful to point out possible limitations of the study that must be dealt with in determining its significance. A major question of the study deals with the relationship of the subject's actions in constructing proofs and their understanding of mathematical proof. The title of the article indicates that the study deals with developmental aspects of children's ability to understand mathematical proof yet the nature of what the author means by "understand" and "proof" is not clear. The construction of finite chains of I's and O's according to the five rules can be accomplished in a manner analagous to building blocks without much thought of inference. On the other hand, however, the rules can be processed abstractly leading to many hypotheses regarding what constructions are or are not possible. Unfortunately, the study did not get at how the subjects thought about the rules but only on how they used them. Perhaps developmental difference would have been indicated more if this aspect were also investigated. The author indicates that the prospects for valuable research in mathematical problem-solving using a system such as used in the study are many and varied. A discussion of what these prospects are and where they may lead, however, is unfortunately missing.

Nicholas A. Branca
Pennsylvania State University

EJ 090 179 SE 510 250
THE TEACHING OF HEURISTIC PROBLEM SOLVING STRATEGIES IN ELEMENTARY CALCULUS. Lucas, John F. Journal for Research in Mathematics Education, v5 n1, pp36-46, January 1974.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Shmuel Avital, Institute of Technology, Technion, Haifa, Israel.[1]

1. Purpose

This paper summarizes an exploratory investigation of the influence of instruction in heuristic strategies on problem solving performance of a small group of first-year university students.

2. Rationale

The study was designed, with minor modifications, within the framework for heuristic behavior in the solution of word problems in mathematics, as developed by Kilpatrick in his doctoral dissertation.[2] In its modified form the author explores more than 30 dependent variables categorized under the following five aspects of performance in the solution of word problems in mathematics: (i) usage of heuristic strategies (20 variables); (ii) modes of difficulty (3 variables); (iii) types of error (7 variables); (iv) time measures (3 variables); (v) score measures (4 variables). Many of the variables were highly interrelated.

3. Research Design and Procedure

The subjects in this study were 30 first-year university students enrolled in two calculus classes, one experimental (17 subjects) and one control (13 subjects). The only randomization used was the assignment of classes to treatment of the major experimental variable. The treatment consisted of two experimental variables:

(i) Exposure (or non-exposure) to instruction in heuristic strategies in the solution of problems during an eight-week course in freshmen calculus taught by the investigator.

(ii) Application (or non-application) of a pretest in the form of diagnostic observation of usage of heuristics in the solution of seven problems "not specific to any branch of mathematics." The first was considered to be the major variable under investigation. The pretest was applied to 8 experimental and 6 control subjects.

The subjects treated to instruction in heuristic strategies were rewarded during the instruction period for usage of heuristics in the solution of assigned problems. All 30 subjects were administered a two-hour oral posttest of problems different in content but similar to those used in the pretest.

Pre- and post-testing were carried out in the form of individual two-hour interviews in which the subjects were required to "think aloud" as they worked. All interviews were tape-recorded. The records and written work were analyzed according to the pre-established categories.

The researcher dichotomized all data except for the time measures. A χ^2 analysis was used for the dichotomized variables and ANOVA for the time measures.

4. Findings

Results are reported only for the major experimental variable: exposure or non-exposure to instruction in usage of heuristics. Significant differences ($p < 0.05$) in favor of the experimental group were found for (i) four heuristic strategies (usage of notation, usage of methods or of results of related problems, and separation or summarization of data); (ii) two indicators of difficulty (frequency of rereading and hesitations); (iii) three types of dichotomized scores (approach, plan, and total).

Marginally significant differences ($0.05 < p \leq 0.1$) were found for two heuristic variables (reasoning by analysis and avoidance of drawing non-representative diagrams), for score by results, and for time spent looking back. No difference was found for any type or error committed by the solvers.

Neither means nor variances are given in the paper for any of the variables. The author does not quote any results either about significance of the minor experimental variable (application of the pre-test), or about the differences in performance on the calculus tests in the regular course.

5. Interpretations

In his interpretation and conclusions, the author draws attention to the limitation of the study. He nevertheless suggests that the results support the possibility of a long-range transfer effect of those heuristic strategies for which significant differences were found in the study. He also emphasizes the possibility of incorporation into regular university teaching instruction in heuristic strategies and approaches to problem solving without undue expense to normal course time or course content.

Critical Commentary

This is an interesting study which tries to meet a very important need. However, some data and results not reported in the paper make it impossible for the reader to arrive at any meaningful conclusions.

Problem solving is a generic term. It might encompass anything from finding the price of 13 items if each of them costs 14¢ to discovering whether there is an infinite number of pairs of primes of the form p and p + 2. The only description given in the paper about the nature of the problems used in the pre- or post-test was "mathematical problems not specific to any branch of mathematics." What does this mean?

The lack of reporting about the significance of the minor experimental variable, participation or non-participation in the pre-test, makes the interpretation of the results meaningless. A (hypothetical) significant interaction between the two experimental variables would give a completely different slant to the results. Also some data about the performance of the twogroups in the calculus tests would have been of great importance to a university instructor who might intend to replicate the study.

Lack of reporting of the means and variances of the scores makes it impossible to interpret the significance of the differences. A difference of 1 or 2 points might sometimes be statistically significant while being educationally completely unimportant. One can assume that all these results are given in the original doctoral dissertation, of which this paper is supposed to be a summary, but a journal article has to be self-contained.

A major fault in the design is the lack of use of independent observers who should have evaluated the similarities and differences in the treatment of the experimental and control groups during the regular instruction in calculus. Whatever one may think of the experimenter effect, the researcher who formulates hypotheses has some stakes in their support.

An outcome of major importance, in this reviewer's opinion, is the fact mentioned in the last statement of the summary: instruction in and through problem solving can be carried out not necessarily at the expense of course time or course content.

We need more studies of this nature, with larger samples, longer duration, stricter control, and more comprehensive reporting.

Shmuel Avital
Institute of Technology, Technion
Haifa, Israel

[1]This abstract was written while the abstractor was visiting with the Department of Mathematics, University of Oregon, Eugene, Oregon.

[2]Kilpatrick, J. Analyzing the solution of word problems in mathematics: an exploratory study. Ann Arbor, Michigan: University Microfilms No. 68-6442, 1967.

EJ 087 867 EM 503 621

ON USING STUDENT-GENERATED SEQUENCES IN THE DEVELOPMENT OF A LEARNING HIERARCHY. McKeen, Ronald L. and Eisenberg, Theodore A. Improving Human Performance: A Research Quarterly, v2, pp97-106, 1973.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Harold Schoen, University of Iowa.

1. Purpose

To determine whether students organize their behaviors in a logical manner pursuant to the solution of a specific problem or learning task.

2. Rationale

Gagné and many others have considered knowledge relevant to a given task as a set of subordinate capabilities or behaviors. This viewpoint has led to the use of learning hierarchies in curriculum construction. The traditional method of constructing a learning hierarchy is by an a priori task analysis of the desired instructional outcome. However, this approach has been criticized for being too narrow and stifling to students' imagination and creativity. In addition, instructor-generated learning sequences have been shown to vary from student-generated sequences of the same content.

3. Research Design and Procedure

A skeletal hierarchy of less than five cells (subtasks) was constructed by an a priori task analysis depicting a typical path to the following mathematical behavior:

> Given a sequence of real numbers whose general term can be expressed as a rational expression of order ≤ 2, construct a proof demonstrating the existence or non-existence of a limit using the N-ε method.

The behaviors were then translated into representative problem sets. Twelve college students, 11 elementary education majors and one English major who were similar in mathematical background and considered to be "talkative" by the investigators, were assigned to groups of 3 or 4 each.

As each group worked collectively toward solutions to the stated problems in the skeletal hierarchy, the investigators observed the dovetailing of student behaviors and the chronological order in which the behaviors were exhibited. Repeating this procedure with each group, the skeletal hierarchy was expanded and refined.

Once the hierarchy was developed via student input, a second sample of 16 elementary education majors was chosen to validate the hierarchy. Two groups of 8 students each were formed. These groups, the high group and the low group, were significantly different with respect to mathematics background and grade point average. The investigators presented the content of the hierarchy to the two groups using an expository teaching method and the final experimental hierarchy as a pedagogical format. After a 3-week instructional period, evaluation booklets consisting of a random selection of representative tasks from each cell in the hierarchy were distributed to the students. The results provided the data for testing the 72 hypotheses of task dependency in the hierarchy. Task dependency refers to whether or not acquisition of a behavior in a subordinate cell is really a necessary condition for acquisition of the behavior in its immediate superordinate cell. Each superordinate cell in the hierarchy and its immediate subordinate cells constitute a hypothesis of task dependency.

4. Findings

The tasks in the evaluation booklet were classified into one of four groups. Group One consisted of those tasks for which at least 90% of the Ss acquired 90% of the behaviors. If between 80% and 90% of Ss acquired at least 90% of the specified behaviors, the task was placed in Group Two. Group Three and Group Four consisted of those tasks in which between 70% and 80% of Ss and less than 70% of Ss, respectively, acquired at least 90% of the specified behaviors. The findings are summarized in the Table that follows.

Table

Percent of Tasks by Groups*

Task Group	Student Group		
	High	Low	Total
One	84	23	74
Two	7	21	17
Three	5	37	7
Four	4	33	2

*estimated from a bar graph

For the high group, 85% of the dependency hypotheses were accepted as valid at the .90 acceptance level. Only 33% of the hypotheses were accepted as valid in the low group. Overall, 46% of the dependency hypotheses were accepted as valid. The data did not support the hypothesis; that is, the student-generated hierarchy was not validated.

5. Interpretations

It has been demonstrated in this study that student input can be obtained for purposes of producing a learning hierarchy, and the constructed hierarchy will have a high percentage of valid hypotheses for a particular group of students - the high group in this study. Evidence is also provided in this study to indicate that student-generated learning hierarchies are structurally different for students of differing aptitudes. Since student-generated hierarchies are also often structurally different from those that are produced by the traditional task analysis procedure, the commitment of the extra time and effort by curriculum developers to this sort of hierarchy generation will result in a better final product.

Critical Commentary

According to McKeen and Eisenberg, "a validated learning hierarchy is (1) [a pyramid of tasks] in which the subordinate behaviors have been demonstrated to be necessary and sufficient conditions for performance of the superordinate behavior, and that (2) the sequencing of behaviors is a logical one." It is clear that the second criterion in the definition can be met without the first one. The real difficulty comes in meeting the first. The researchers in this study have made an important contribution by pointing out (1) that the a priori task analytic approach to developing learning hierarchies may not be the most useful one, and (2) students of differing aptitudes are likely to generate structurally different hierarchies. Nonetheless, there are still many difficulties with the theory of learning hierarchies and its application to curriculum development. For example, McKeen and Eisenberg tacitly assume that there is a "best" sequence for achieving a terminal behavior which is in some way innate to the student, or is within the subject matter and discoverable with little or no guidance by the student. But is it not also true that within certain limits the sequence (or hierarchy) of learning is a function of the sequence chosen for instruction? A floundering student with little guidance will not necessarily stumble upon a valid hierarchy that will be useful for organizing curriculum and/or instruction.

The most appropriate method of generating a learning hierarchy must depend upon its intended use, a fact not mentioned by McKeen and Eisenberg. For example, hierarchies are sometimes used to place remedial students in an instructional sequence by starting them at the lowest level at which they fail to meet the performance criterion. Neither the task analysis method nor the student-generation approach is likely to provide a hierarchy that will be valid for this purpose since neither considers the effects of previous instruction and forgetting. Forgetting probably does not occur in "reverse" hierarchical fashion.

Finally, there appears to be an error in the graph presented in the Findings section of the McKeen-Eisenberg article. The percent of tasks in the low group totals 114% across task groups by my estimation. There is also a discrepancy between the percent of tasks in Group One in the low student group on the graph (23%) and that stated in the body of the paper (37%).

Harold Schoen
University of Iowa

EJ 080 045 PS 502 694

CONSERVATION OF MASS, WEIGHT, AND VOLUME AS EVIDENCED BY ADOLESCENT GIRLS IN EIGHTH GRADE. Nadel, C. and Schoeppe, A. Journal of Genetic Psychology, v132, pp309-313, June 1973.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by F. Richard Kidder, Longwood College, Farmville, Virginia.

1. Purpose

This study is a replication of a part of an experiment by Elkind on conservation of quantity by adolescents. Elkind's results differed from those of Piaget, who found that children aged 11-12 had attained conservation of mass, weight, and volume. In contrast, only 29% of Elkind's sample evidenced conservation of volume, although most had attained conceptions of conservation of mass and weight.

2. Rationale

One of the continuing needs in research is the verification through replication of the findings of significant experiments. Attempts to verify are abundantly evident in the many replications of Piaget's experiments. Elkind's replication of Piaget's quantity experiments produced conflicting results. Nadel and Schoeppe sought to clarify the conflicting findings of Piaget and Elkind as to the attainment of volume conservation by adolescents through replication of a part of Elkind's experiment.

3. Research Design and Procedure

Piaget's "sausage" tests were administered to a home economics class at the Jonas E. Salk Junior High School in Levittown, New York. The girls ranged in age from 13 years, 1 month to 13 years, 11 months. The tests for conservation of mass, weight, and volume were given to groups of fourteen subjects each.

In the test for conservation of mass, the experimenter (regular classroom teacher) showed the subjects two identical balls of clay and demonstrated that they weighed the same. After the subjects had agreed that the balls weighed the same, the experimenter rolled one of the balls into a hot dog shape and asked, "Do they both contain the same amount of clay now?" Answers were recorded on index cards.

The same procedure was used to test for conservation of weight and volume. After reforming the hot dog shape into a ball and reweighing the two balls of caly, the experimenter again rolled one of the balls into a hot dog shape and asked, "Do they each weigh the same?" Then, the clay ball and the hot dog were exhibited and the subjects were asked, "Do the two pieces of clay each take up the same amount of room in space?"

The percentage of subjects answering in the affirmative to the experimenter's questions were used to test the experimental hypothesis: the sample has attained conservation of mass and weight but has not attained an abstract conception of conservation of volume.

4. Findings

Of the twenty-eight subjects tested, 89% conserved mass, 100% conserved weight, but only 29% conserved volume. These results closely parallel Elkind's findings of 90%, 92%, and 29% of subjects conserving mass, weight, and volume respectively. The three subjects who did not conserve mass also did not conserve volume.

5. Interpretations

Nadel and Schoeppe feel that the findings confirm Elkind's results and lend further support to the hypothesis that young adolescents have not attained volume conservation although they have abstract conceptions of mass and weight. They also feel that the results are supportive of Piaget's theory that volume conservation is attained at about the same age at which the adolescent is developing formal mental operations and that Elkind's hypothesis, that the emergence of formal operational thinking may be an interfering factor with the acquisition of volume conservation, has merit.

Critical Commentary

This study sheds little new light on the child's conception of mass, weight, or volume. Elkind's original experiment is more definitive and illuminating. Elkind tested each subject individually, permitting an explanation of why the subject answered as he did and insuring that each subject understood the questions he was being asked.

There appears to be no empirical basis for Nadel and Schoeppe's support of Elkind's hypothesis that the emergence of formal operational thinking can explain why so many young people who spontaneously discover mass and weight fail to discover conservation of volume. The question as to why only 29% of the sample failed to conserve volume was not examined. The "why" is of importance to mathematics educators, however, and needs further exploration. In the "sausage" type experiments, volume is "how much space does an object occupy?" In elementary schools, is volume not normally thought of as capacity, that is, "how much will an object hold?" Do these have the same meaning to a child? Could this previously learned concept of volume be a factor in the young adolescent's incorrect answer to the conservation of volume question?

F. Richard Kidder
Longwood College
Farmville, Virginia

ED 111 681 SE 019 680

LEARNING OF EQUIVALENCE AND ORDER RELATIONS BY DISADVANTAGED FIVE- AND SIX-YEAR OLD CHILDREN. Owens, Douglas T. In Steffe, L. P. (ed.), Research on Mathematical Thinking of Young Children: Six Empirical Studies. Reston, Virginia: NCTM, 1975.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Karen Fuson, Northwestern University.

1. Purpose

The purpose of this study was to examine some effects of training on conservation and on transitivity of the relation of matching objects. The effects tested were conservation, transitivity, symmetry, asymmetry, and reversibility of the relation of matching objects and upon the same characteristics of the relation of length.

2. Rationale

Piaget (1952) and Northman and Gruen (1970) argue that transitivity is involved in conservation. However, Smedslund (1964) obtained results indicating that conservation of length precedes transitivity of length. This study aimed to clarify the relationship between the attainment of conservation and of transitivity. Another purpose of the study was to examine the effect of training of transitivity of the matching relation upon the other grouping operations of this relation: symmetry, asymmetry, and reversibility. A final aspect concerned the generalizability of the effects of training on the matching relation to properties of another relation, that of comparison of length.

3. Research Design and Procedure

The subjects were 23 kindergarten (ages 5-1 to 5-10) and 24 first-grade (ages 6-1 to 6-10) children. They were randomly selected from the population of children attending a lower socio-economic, predominantly black, school in Atlanta, Georgia.

Four training units were designed. The first two aimed at developing the ability to establish matching and length relations. The third and fourth units involved training of conservation and transitivity of the matching relations. The training lessons were given to groups of 4-6 children at a time. Each lesson lasted 20-30 minutes. There were 10 lessons in Unit I, 7 in Unit II, 4 in Unit III, and 5 in Unit IV.

A Solomon four group design was used. Children of each age were randomly assigned to a treatment group, and part of each of the two treatment groups was randomly selected to be given the pretest. All children were given the posttest. Age and treatmert were the independent variables. There were two levels of each of the two independent variables: five and six years of age, and Full (use of all four training units) or Partial (use of the first two training units) Treatment.

Twelve tests were constructed to serve as both pre- and posttests. Half of the test concerned the matching of five to six objects, and half concerned the comparison of the length of two sticks. The six tests for each relation checked the ability to establish and conserve the relation and to use four properties of the relation: transitivity, symmetry, asymmetry, and reversibility. A thirteenth test involved a Transitivity Problem which required the application of transitivity for its solution.

The lessons in Units I and II were given to all children in the study. Then the tests on the establishment, conservation, and transitivity of both matching and length relations were given to part of each treatment group (Pretest). The Full Treatment group was then given Units III and IV.

Near the end of the Full Treatment it became apparent that the posttest did not use a language pattern used by many of the children. Thus approximations to desired terminology (e.g., "the same" for "as many as") were accepted in the posttests. In addition, if children were answering contradictory statements with identical answers, the questions were repeated using an alternate terminology.

4. Findings

Univariate analyses of variance yielded several significant F ratios in the Treatment versus Age analyses. The Full Treatment had a significant effect only on transitivity of the matching relation (Full: 58% correct, Partial: 36%) and on asymmetry of the matching relation (Full: 73% correct, Partial: 52%). Age had a significant effect on the establishment of the matching relation (Six: 87% correct, Five: 59%) and on the following properties of the matching relation: conservation (Six: 62%, Five: 36%), symmetry (Six: 81%, Five: 48%), asymmetry (Six: 74%, Five: 50%), and reversibility (Six: 72%, Five: 50%). No significant difference was found on transitivity of the matching relation (Six: 51%, Five: 43%).

The Full Treatment had no significant effect on any aspect of the length relation. Group means for the five and the six year olds were significantly different only for the ability to establish the length relation (Six: 93% correct, Five: 81%) and not for conservation, transitivity, symmetry, asymmetry, or reversibility of the length relation. The group means for the performance on all of these properties of the length relation fell between the scores for five-year-olds and those for six-year-olds on the matching relation.

Chi-square tests on a 2 by 2 table composed of the number of children who reached and did not reach the criterion for conservation and that for transitivity of a given relation (matching, length) were nonsignificant, indicating that these abilities were independent. For matching, 10 children conserved and did not use transitivity, while 8 used transitivity and did not conserve. For length, the corresponding figures are 15 and 9.

Nonsignificant chi-square statistics for a 2 by 2 table of children reaching and not reaching the criteria for conservation of matching and of length indicated that these, too, were independent attainments. Ten children conserved length but not matching, and five conserved matching and not length. Ability to use transitivity for matching and length was also independent. Ten children used transitivity with length but not matching, and nine used it for matching and not for length.

The chi-square data on two 2 by 2 tables comparing criterion level performance on the Transitivity Problem (an application of transitivity different from that used in the training) with that on conservation and transitivity of the matching relation indicated that successful performance was strongly related to ability to conserve the matching relation and approached (χ = 5.45) being significantly related to ability to use transitivity in the matching relation.

Three separate multivariate analyses of variance were performed in which two factors and their two-way interaction were considered: Age X Treatment, Treatment X Pretest, and Age X Pretest. None of the F ratios for any factor or two-way interaction were significant at the .05 level.

In the univariate analyses of variance only two interactions were significant. These were both Pretest X Treatment interactions. Full Treatment-Pretest means were considerably and significantly lower than Full Treatment-No Pretest means for transitivity (Pretest: 42%, No Pretest: 85%) and for reversibility (Pretest: 54%, No Pretest: 91%) of the matching relation. Partial Treatment means were not significantly different; they fell between the Pretest and No Pretest values for Full Treatment.

5. Interpretations

The Full Treatment involved four lessons on conservation followed by five on transitivity, but performance of the subjects improved only on transitivity. The author concludes that the conservation training may have been inappropriate or too short. He further suggested that the conservation training may have facilitated training on transitivity, or that the one week delay before testing on conservation (testing followed the completion of all lessons) may have influenced the conservation results. Both the resistance to conservation training and the success in transitivity training is at variance with other studies (see Lovell, 1975, for a discussion). However, the studies which report that conservation precedes transitivity do not involve training on transitivity. This may account in part for the different result obtained here. In addition, Skypek (1966) reported that for lower socio-economic children, the development of conservation was erratic. The relationship between conservation and transitivity may be similarly erratic here, and a replication with middle class children might have different results.

The training on transitivity of the matching relation neither generalized to other properties of the matching relation (except to asymmetry, which the author reports was actually trained in the lessons on transitivity), nor to transitivity of the length relation. Thus the grouping operations seem to be acquired independently, as does a single operation (transitivity) with respect to different real world referents (sets and length).

A possible interpretation of the Pretest X Treatment interaction is that the pretest interfered with the treatment. But because the training did not involve symmetry or reversibility, the author concludes that the pretests had essentially no effect on the performance on the posttests.

Critical Commentary

The study suffers from some design faults. The place of the variable of age in the Solomon four group design is not specified. The numbers given the pretest are not indicated. If age were used in addition to Treatment in this design, the numbers in each cell would be quite small. If it were not used, the pretest data is uninterpretable. A design using a control group given no treatment would have solved the pretest problem and made the place of the variable of age clearer. The sample is quite small for the partitioning it had to undergo. With small cells, one may get fluke results.

Insufficient data is provided by the author in several instances. The design data described above is one. No information about the content of the training is provided. The composition of the small groups of children who received the training and whether any attempt was made to control the quality of this training from small group to small group is not discussed.

The tests as reported do not seem to be strictly comparable. Sometimes "as many as" was the key phrase and in others, "more than" was used. It is not clear whether different versions were used with different children, or simply different questions with different properties of the relations. The results on Piaget-type testing made quite clear that minor differences in the questions can create different results. More care on this vital point should be taken if results from different tests are to be comparable.

The effect of the pretest on the Full Treatment group is very interesting. For the group not given the pretest, the training on transitivity seemed to generalize to symmetry and to reversibility, but for the group given the pretest, the training suppressed posttest means on symmetry and reversibility to a level below that of children given no training on transitivity. One possible explanation is that, on the rather rigidly administered pretest, children learned that they did not know these answers, and then learned only what "they were directly taught," while children not given the pretest were not sensitized to the different properties and generalized more readily.

Owens does not present data on how many children were able to perform adequately on the posttests when questions were rephrased and more flexible answers were permitted. This would have provided helpful evidence concerning the magnitude of the underestimation of the ability of lower socio-economic children when assessed by middle class language patterns.

Several results of this study seem particularly interesting and worth replicating with a larger population of both lower and middle class children. The results of lack of generalizability of the transitive property to other grouping operations seems fairly clear, and the tests for all of these other properties could be omitted, giving time for a larger population. The interesting and controversial points include: the strange effect of the pretest (examination of this effect might require a separate study), conservation and transitivity of the matching relation are independent, means for the establishment of and conservation of the length relation are not different for five and for six year olds, and a training for transitivity of the matching relation was successful while training for conservation was not.

Karen Fuson
Northwestern University

References

Lovell, Kenneth. Summary and implications. In Steffe, L. P. (ed.), Research on Mathematical Thinking of Young Children: Six Empirical Studies. Reston, Virginia: NCTM, 1975.

Northman, J. E. and Guren, B. E. Relationship between identity and equivalence conservation. Developmental Psychology, v2, p311, 1970.

Piaget, Jean. The Child's Conception of Number. London: Routledge and Kegan Paul, 1952.

Skypek, D. H. The relationship of socio-economic status to the development of conservation of number. Unpublished doctoral dissertation, University of Wisconsin, 1966.

Smedslund, J. Concrete reasoning: a study of intellectual development. Monographs of the Society for Research in Child Development, (Serial No. 93), v29, 1964.

MATHEMATICAL PROBLEM SOLVING. Scandura, J. M. American Mathematical Monthly, v81 n3, pp273-280, March 1974.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by John A. Dossey, Illinois State University.

1. Purpose

The four studies reviewed in this article dealt with various aspects of Scandura's model of mathematical behavior as a collection of rule governed activities (Scandura, 1971a). In particular, the studies examined the relationship between a S's ability to perform higher order rule activities and the S's ability to succeed in problem solving situations. In addition, the relative efficacy of a curricular program for elementary school teachers employing a higher order rule approach was examined.

2. Rationale

Scandura's model for mathematical behavior accounts for problem solving in terms of a S's ability to act on previously learned rules applicable to a problem situation where previous rules will not apply individually. A careful and detailed explication of the nature of these higher order rules is given in another article by Scandura (1971b). If problem solving is characterized according to model, one might question whether or not problem solving behaviors could be improved by instruction on higher order rules and whether curricular materials designed around higher order rules might be more effective ways of improving students problem solving abilities. It is to these questions the set of studies in the present article are directed.

3. Research Design, Procedures, and Findings

Study 1. This study concerned the ability of elementary school Ss to form the composition of two trading rules they were presented with in order to derive a new rule to solve a problem solving situation involving trading rules. The higher order rule required for solving the problem was that of composition of relations.

Following instruction on simple trading rules, Ss with any prior knowledge of composition of trading rules were eliminated from the sample. The remaining Ss were divided into a control and an experimental group. The control group received no further instruction and the Ss in the experimental group were provided instruction on the use of the higher order rule involving the composition of relations.

After this period of instruction, a transfer test involving the composition of trading rules was given to all of the Ss. All of the Ss in the experimental group were successful on these tasks while none of the students in the control group were.

Study 2. This study concerned a higher order generalization rule involving function rule relationships and employed a sample of Ss from a secondary school remedial mathematics class.

Specifically, the study involved two classes of rules:

Class I: Restricted and General Rules

1. Restricted rules consisted of three input-output pairs where one of the input numbers was 1 and a pair of the input numbers were successive natural numbers.

2. General rules consisted of function rules of the form $n \rightarrow an$ or $n \rightarrow an + d$ where a and d are natural numbers.

Class II: Higher Order Rules

1. A Division Rule shows how to form general rules of the form $n \rightarrow an$ from certain sets of restricted rules.

2. A Finite Difference Rule shows how to form a general rule of the form $n \rightarrow an + d$ from a wider set of restricted rules.

The Ss were taught how to interpret given restricted rules and general rules. They were then tested with a transfer task concerning the derivation of a general rule to correspond to a given restricted rule.

The Ss were then assigned to groups on the basis of their performance on this transfer test. Ss who could not derive any general rules were assigned to a group labeled NONE. Those that could solve for general rules of the form $n \rightarrow an$, but not of the form $n \rightarrow an + d$ were placed in the group labeled AN. All other Ss were dropped from the study. The Ss in NONE and AN were further divided into treatment groups. Those in NONE were assigned to one of three treatments: those in group C were given no new instruction, those in group D were given instruction on the higher order division rule, and those in F were given instruction on the higher order finite difference rule. The Ss in AN were divided into D and F treatment groups in a like fashion. (Note that the D treatment group serves as a control group for the F group within the AN Ss.) Ten Ss were assigned to each of these groups from those available in the NONE and AN groups.

The results of the posttest on transfer items involving function rule problems showed:

Table I

Number of Students who Reached Criterion

Grps		Interpretation		Pretest			Posttest	
		Rest. Rules	Gen. Rules	$n \rightarrow an$ Probs.	$n \rightarrow an + d$ Probs.	H.O. Rule Treat.	$n \rightarrow an$ Prob.	$n \rightarrow an + d$ Prob.
NONE	C	10	10	0	0	--	0	0
NONE	D	10	10	0	0	10	10	0
NONE	F	10	10	0	0	10	10	10
AN	D	10	10	10	0	10	10	0
AN	F	10	10	10	0	10	10	10

The results of this study were in exact agreement with Scandura's hypothesis that Ss would be able to perform on a particular problem solving task if they has received instruction on the corresponding higher order rule.

Study 3. This study examined whether predictions on Ss' problem solving abilities and knowledge of higher order rules could be made via a testing route rather than through a test and instruction process as in the other studies. To test this, the preceding study was repeated with an unidentified group of Ss. Following the pretest, the predictions were made and the posttest was administered. The results showed that 86% of the predictions concerning performance on the $n \rightarrow an$ rules were correct and 92% of the predictions concerning the $n \rightarrow an + d$ rules were correct.

Study 4. This study concerned the use of a rule oriented approach to the teaching of a basic content course in mathematics to two groups of elementary school teachers. The discrete rules group received training on over 300 specific task related rules in the course. The higher order rules group received training on a reduced list of only 170 of the specific rule and 5 relevant higher order rules.

The follow-up study showed that the students in the higher order rules group did as well as the Ss in the posttest over the 300 rules and did statistically better than the discrete rules group on new transfer problem solving tasks. Hence the higher order rules group achieved more while being given less.

4. Interpretations

The results of the studies tend to substantiate the higher order rule characterization of problem solving behaviors. To date, the studies have been limited to very controlled situations. The author hopes that they can be extended to more complex problem solving situations which involve conjecturing, proving, and developing counterexamples. In addition, the author notes the possibility of further studies in the area of the development of curricular materials in terms of higher order rules.

Critical Commentary

The research report, due to its broad coverage and its publication in the Monthly, lacked many of the details necessary to make a critical analysis of the research design or experimental procedure followed in each study. There remain many unanswered questions in the abstractor's mind concerning the nature of the experimental settings, the instructions given the experimental groups, the time of instruction, the time period between the instructional periods and the testing periods, the nature and length of the tests, and the relation of the test items to the instructional items. With the exception of these questions, the direction these studies take seems to be a profitable one for future studies into the role of rule governed behaviors in more complex aspects of problem solving and curricular design.

John A. Dossey
Illinois State University

Scandura, J. M. A theory of mathematical knowledge: can rules account for creative behavior? Journal for Research in Mathematics Education, v2 n3, pp183-196, May 1971. (a)

Scandura, J. M. Deterministic theorizing in structural learning: three levels of empiricism. Journal of Structural Learning, v3, pp21-53, 1971. (b)

A CONFLICT BETWEEN SELECTING AND EVALUATING INFORMATION IN AN INFERENTIAL TASK. Wason, P. C.; Johnson-Laird, P. N. British Journal of Psychology, v61, pp509-515, 1970.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Theodore A. Eisenberg, Northern Michigan University.

1. Purpose

To study the interaction between selecting a response to an inferential task and evaluating the selection.

2. Rationale

This study is a refinement of a previous experiment by Wason (1969). We must understand the previous experiment before stating the rationale of the present study.

In the previous study, subjects were presented with four cards showing a red triangle, a blue triangle, a red circle, and a blue circle. On the back of each card was a triangle (red or blue) or a circle (red or blue). The subjects were presented with the following rule about the four cards: *Every card which has a red triangle on one side has a blue circle on the other side.* The subjects were then asked which cards needed to be turned over to determine whether or not the rule was correct. They found this task extremely difficult: none got it right initially. (The correct response is "the red triangle and the red circle," because if these two appear on the same card, the rule is false--otherwise it is true.)

The previous experiment contained a confusing element. Half of the information was referred to as being on one side of the card and the other half on the other side of the card. It was argued that the subjects may have interpreted "being on the other side" as the side that faced *downward*.

The present study presented all information on the same side of the card. A second aspect of the present study was to introduce an evaluation process into the procedure. After the subject made his selection he was asked to evaluate his response with respect to the rule. This evaluation process was expected to change the selection process, when the subject was confronted with his incorrect selection.

3. Procedure

Thirty-six paid volunteers, who were undergraduate students of University College, London, were assigned to one of three groups (n = 12).

For Group A the rule was: *Every card which has a circle on it has a border round it.* The stimuli were presented as follows: (see Figure 1): p = a circle with a mask around the edge of the card; $\bar{p}$ = no circle with a mask around the edge of the card; q = a curly border around the edge of the card with a mask over the center of the card; $\bar{q}$ = no border around the edge of the card with a mask over the center of the card. (The correct answer to determine if the rule is true or false is p and $\bar{q}$.)

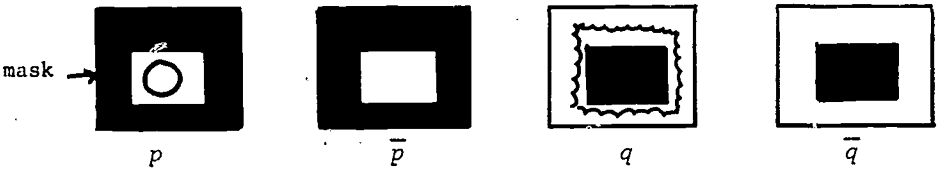


Figure 1

For Group B the rule was: *Every card which has a circle on it has two borders round it.* A triangle was substituted for p, two borders for q, and one border for q. Appropriate cards were constructed.

For Group C the rule was: *Every card which has a circle on one side has two borders on the other side*, and four similar cards to those in Group B were presented.

It was hypothesized that the frequency of correct responses would be Group A > Group B > Group C.

The subjects were asked which cards had to be exposed (or turned over) to determine if the rule was correct. After the subject made his choice, the experimenter uncovered the correct cards, and then informally discussed the subject's response, asking him if he was happy with his selection.

4. Findings

Only two of the 36 subjects (both from Group A) were correct initially, and an additional 21 corrected their selection after discussion. Presenting information on the same side of the card did not present an advantage over presenting information on both sides of the card. The hypothesis that subjects may have had a tendency to interpret "the other side of the card" as being the side which faced "downward" was incorrect.

The evaluation process which allowed the subject, after discussion, to change his selection proved effective for 21 subjects. Yet, 13 subjects failed to gain insight into the selection process even when relevant information was exposed and the consequences of their selection discussed.

5. Interpretations

In some cases the subjects refused to change incorrect answers even after realizing that their procedure was in error. The authors believe that this is because the subjects had to "... resolve two seemingly irreconcilable thought processes." The selection process dominated the evaluation process because it occurred first, and involved a decision to act in a particular way; the evaluation process was merely a passive judgment.

In other cases, even at the end of the experiment, subjects lacked insight into the selection procedure. The authors claim that this was because the subjects had internalized their selection procedure to the point where it was immune to modification.

Critical Commentary

This study was well designed and is similar to many others conducted in psycholinguistics. One of the main goals of psycholinguistic research is to gain insight into information processing. Mathematics educators are also interested in information processing, and more generally concept formation. This particular study has shown how selection and evaluation processes either interact or pass each other by. It would be interesting to see whether similar conclusions would hold for different populations and for different tasks. Overall, this study asked an important question, and presented some meaningful answers.

Theodore A. Eisenberg
Northern Michigan University

Wason, P. C. Regression in reasoning. British Journal of Psychology, v60, pp471-480, 1969.