DOCUMENT RESUME

ED 134 475 SE 022 050

AUTHOR Osborne, Alan R., Ed.
TITLE Investigations in Mathematics Education, Vol. 9 No. 4.
INSTITUTION Ohio State Univ., Columbus. Center for Science and Mathematics Education.
PUB DATE 76
NOTE 72p.
AVAILABLE FROM Information Reference Center (ERIC/IRC), The Ohio State University, 1200 Chambers Rd., 3rd Floor, Columbus, Ohio 43212 (Subscription $6.00, $1.75 single copy)

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS *Abstracts; Cognitive Development; Effective Teaching; Elementary Secondary Education; *Instruction; *Learning; *Mathematics Education; Problem Solving; Research; *Research Reviews (Publications); State of the Art Reviews
IDENTIFIERS Number Operations

ABSTRACT

Seventeen research reports related to mathematics education are abstracted and critically analyzed. Seven of the reports deal with aspects of learning theories, four with topics in mathematics instruction (subtraction, problem solving, division by zero, and high school geometry), two with teacher effectiveness and competency, and four with testing. In addition, an introductory editorial on the state of research in mathematics education is included in this volume. Finally, research related to mathematics education which was reported in RESOURCES IN EDUCATION (RIE) and CURRENT INDEX TO JOURNAL IN EDUCATION (CIJE) between July and September 1976 is listed. (DT)

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

Volume 9, Number 4 - Autumn 1976

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

A note from the editor....

I have served as the editor of I.M.E. through the production of two volumes or eight issues. These eight issues contain critical abstracts of 126 research reports as well as one special issue concerned exclusively with the NLSMA Reports. A total of 112 studies from 24 different journals and 14 from non-journal sources were abstracted. The Journal for Research in Mathematics Education (JRME) was the most heavily represented journal; a typical issue of I.M.E. contained abstracts of six JRME articles.

I.M.E. exists to provide two primary services for research in mathematics education. One service is to abstract research reports pertaining to mathematics education but appearing in a variety of sources not all of which are read commonly by mathematics educators and researchers in the field. The other important service is directed to improving the quality of research and research reporting. This service is accomplished by the abstractors including critical commentary of the research they have described. Thus, the critical commentary section concluding each abstract is considered particularly important and warrants careful reading. Most abstractors are quite careful in leveling criticism at the work of their peers and offer thoughtful considered analysis of the reported research.

During the course of editing these eight issues of I.M.E. I have become aware of the fact that the set of critical commentaries constitutes a measure of the state of the art for research in mathematics education. It is significant to note that the commentaries represent the differing views of research held by a diverse set of individuals who are active researchers or consumers of research.

The critical commentary section provides the abstractor an opportunity to be critical; many respond by pointing out deficiencies or difficulties and do not discuss the strengths of the research reports. Thus, the total set of critical commentaries yields a sense of what is wrong with our research rather than its power. It does, however, provide a sense of how researchers in mathematics education regard research.

Recently I examined all of the critical commentaries found in I.M.E. during the two volume publication period in order to acquire a more accurate "fix" on the profession's perception of itself. I thought you might find the results interesting. I began in a relatively disorganized fashion by simply reading all of the critical commentaries and registering overall impressions. Two of these impressions warrant comment.

First, the journals whose primary readership is teachers appear to have editorial policies that are not adequate to deal with research reporting. The critical commentaries of research articles appearing in this type of journal frequently have statements that say, in effect, "Enough data must be communicated, reported, and analyzed to make the conclusions of the research understandable to teachers." (One notes that some superior abstractors conclude the statement with "...understandable _even_ to teachers."!) Of particular concern to abstractors were the articles reporting on teaching methods that do not describe thoroughly enough the teaching procedures used to allow the teacher to replicate or employ the procedure in the classroom. Abstractors of articles appearing in _School Science and Mathematics_, the _Arithmetic Teacher_, the _Mathematics Teacher_, and the _American Mathematical Monthly_ all raised the issue of what is the appropriate amount of information about the research to relate for an audience of non-researchers. Most abstractors communicated at least implicitly an imperative for reporting research for or to the teacher and would subscribe to the point-of-view that journals such as those listed above ought to continue reporting timely research for teachers. But the editorial policies appear to lead to articles that are stylistically pointed to the researcher but suffer from deficiencies that would make the article unacceptable to research journals. Rather than applying an editorial policy that encourages a clear, informal writing style that communicates, many journals apparently settle for simply deleting some of the technical reporting of the experiment. Simply glossing over or ignoring research deficiencies is not a responsible approach to serving the needs of teachers who want to use research results in their classrooms. Perhaps this contributes to the low regard that some teachers have for research. At any rate, journals and research reporters who have primarily a target audience of teachers need to realize that for research to be useful in instructional practice it must be well desi[illegible], adequately reported, and well written.

A second general conclusion of my informal reading of all of the critical commentary sections is that mathematics education research has what I will label as an accumulation problem. That is, the research does not "add up." There are a considerable number of studies that appear to be at best tangentially related to previously accomplished and reported research. This is very evident within the comments of the abstractors. Checking the list of authors of articles reinforces this point of view; many of the authors whose articles are abstracted have only one article during the period of time of publication (approximately three years) from which articles abstracted in these two volumes were selected. Few individual researchers or groups of researchers have focused upon a single research area or topic sufficiently to have generated a set of outcomes or results that is additive. Although continuous research on a single topic by a single researcher is not a necessary condition for research "adding up", the rarity of such sets of individually produced studies is symptomatic of the problem.

The initial, informal perusal of the critical commentaries led me to examine them in a more organized fashion for the journal that is represented most completely in I.M.E.; namely, JRME. Since JRME is the premiere journal for researchers in mathematics education in North America, this seemed particularly appropriate. Close to forty articles from JRME have been abstracted in these two volumes of I.M.E. I examined each critical commentary section of abstracts for JRME articles in terms of the following nine categories that are frequently used in judging the quality of research.

1. Definition of problem
2. Design
3. Control of critical variables
4. Sampling
5. Measurement instruments and/or techniques
6. Data analysis
7. Interpretations and/or generalizations
8. Reporting
9. Significance of the problem

If in my subjective judgment a criticism was registered by the abstractor for one of these categories, then I made a tally mark for that category. If for a given category two criticisms were lodged for the study, then only one tally was made for that category. Only criticisms were registered; no offsetting quality points were noted if the abstractor remarked a particular strength of a research report. I did not attempt to differentiate between minor criticisms and what an abstractor would consider a major flaw severely affecting or torpedoing the research.

I found the following:

1. Forty-six percent of the abstractors quarrelled with the definition of the research problem. Typically the abstractor would claim the research was not studying what the researcher claimed. For example, an abstractor might claim that a purported study of problem solving really was a study of memory and routine algorithmic application of particular heuristics.

2. Thirty percent of the studies had reported design problems. Usually this boiled down to a problem of how the control group was handled.

3. Problems in the control of the critical variables were noted for 43 percent of the studies. Although sometimes derivative from the problem definition (category 1), many times the criticism was lodged in terms of an incomplete reporting of the treatments that would not allow replication.

4. Sampling difficulties were noted for 19 percent of the studies.

5. Instrumentation was a difficulty noted for 24 percent of the studies. Typically the comments focused upon incomplete reporting of instrument characteristics of reliability and/or validity. Failure to report interview protocols thoroughly was another major criticism.

6. Thirty-two percent of the abstractors would have employed different statistical procedures.

7. Abstractors reported stronger interpretations or conclusions than the data warranted or the overlooking of important results in 32 percent of the studies. The latter criticism was made but rarely.

8. The reporting of the experiment was questioned in 51 percent of the research reports. Incomplete description of a portion of the experiment was the major category factor noted by most abstractors.

9. The significance of the problem was questioned in 27 percent of the studies. That is, close to a third of the studies were deemed to be a waste of time for the researcher to have conducted and should not have been reported in the journal since that consumes limited page space. A couple of the criticisms identified the problem area as significant but noted that this particular study was neither relevant nor productive for the area. Three studies were criticized as having some limited relevance for basic research but no relevance for curriculum or instruction in the classroom.

It might be interesting for the reader to realize that I have from time to time received cover letters with abstracts in which the abstractor will note that he or she has not raised the question of significance of the problem in the abstract but that I should feel free not to use the abstract since the mere fact of publication of the study had given enough visibility to a study of an insignificant problem. I have also not reported the number of turn-downs of abstract requests that arrive with the accompanying statement that the study is of insufficient importance to be abstracted. Some abstractors (researchers) appear to be loath to raise the basic question of significance of the problem in the public arena of I.M.E.

I find the significance category of criticism the most disconcerting. The questions concerning the inconsequential, non-significant character of research were being raised about studies reported in JRME, a journal specific to mathematics education and refereed by mathematics educators. The 27 percent criticism rate concerning the significance of the problem represents a different order of criticism than found in the other categories. Most of the other criticisms identify correctable features in the research. For instance, tinkering with the design, improving the instrumentation, or selecting a better sample represent factors that can be reacted to, provided for, and improved on in replicating or extending a study. The criticism of lack of significance of the study is a message to forget the endeavour, to return to "Go," and to start over. It represents wasted effort in the judgment of the abstractor. The criticism was applied to studies that were well designed as well as to those that were poorly designed. Since some effort is made to match the expertise and interest of the abstractor with the type of study he or she is requested to abstract, this is a particularly damning commentary on the state of the art in research in mathematics education. The criticism has seldom reflected basic philosophical differences between the abstractor and the researcher. It is evidence that the field needs more opportunities to collect sets of researchers together to argue, discuss and debate what the priorities of research efforts in mathematics education should be. It is evidence that journals such as JRME need more issue-oriented articles that question the goals and payoffs associated with particular types of research. It is a message to researchers to orient their thinking toward problems of importance for mathematics learning and teaching in the schools and the classrooms.

The field of research in mathematics education is relatively adolescent. Clearly the two volumes of I.M.E. considered in this analysis indicate important strengths in the field. Equally apparent is the fact that we have some distance to go if research is to have an accumulative, additive impact on instructional and curricular practices. The editorial policies of journals and the behavior of professionals in the field of mathematics education research need to address the problems of significance and importance of research if the field is to attain productive maturity.

Alan Osborne

Autumn 1976

STABILITY OF TEACHER EFFECTIVENESS: A REPLICATION. Acland, Henry. Journal of Educational Research, v69, pp289-292, April 1976.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Charles E. Lamb, The University of Texas at Austin.

1. Purpose

The purpose of the study was to assess fifth-grade teachers in order to "...establish the temporal stability of their relative effectiveness on the average level of students' educational achievement in two consecutive years."

2. Rationale

Rosenshine's (1970) review of studies dealing with the stability of teacher effectiveness produced no definitive statement concerning results. The three studies which he reviewed had used somewhat different methods and designs, thus making it hard to synthesize findings. Since then, Brophy (1973) has conducted a study which reported stability of effectiveness data for another group of elementary-school teachers. The present study was conducted in order to replicate the previous studies. The replication was not an exact one, but used "...the same basic method as earlier studies...". The present study also sought to provide a new plan for representing the importance of the stability of teacher effectiveness.

3. Research Design and Procedure

The study involved 89 fifth-grade teachers from a large urban school district. All teachers in the sample were from regular, self-contained classrooms. Children in these classrooms were given the intermediate battery of the Metropolitan Achievement Test (1959) in the fall and spring of two consecutive years. Residual gain scores were computed for each year and used as a measure of teacher effectiveness. The number of teachers used was the maximum number available, given the constraint that each had at least 11 students with recorded test data on a given testing occasion. Therefore, the selection of teachers was neither systematic nor random.

The basic data for the study came from the nine subtests of the Metropolitan Achievement Test. Class mean scores were derived in order to compute residual gain scores. There are several different approaches to this derivation. The sets of students could be matched or unmatched (matched sets are made up of students who were tested in both fall and spring, while unmatched sets contain students who were tested at one time or the other, but not necessarily both). Secondly, the choice of metric allows three possibilities: raw scores, standardized scores, or grade equivalent scores. Thirdly, the change from raw scores to another metric could take place before or after class means are calculated. After consideration of these possibilities (by use of computed correlations among the possibilities), the researcher chose to use standardized mean scores for matched students with the transformation taking place prior to the averaging of scores. Following computation of the means, the researcher also calculated standard deviations and reliability coefficients for the nine subtests.

The class means were used to calculate residual gains for each of the two years of the study. Therefore, each teacher had an indication of class average gains for consecutive years of teaching. The gain scores for 1970-71 were correlated with the gain scores for 1971-72 for each of the subtests.

In order to give a better picture of the practical importance of teacher consistency, the researcher calculated a teacher "effect" by multiplying the interannual correlations between gain scores and the average of the standard deviations for year 1 and year 2. The effect produced represents the expected difference in points for class mean scores attributable to the stability of teachers one standard deviation apart on the measure of teacher stability. It would also be possible to look at effects by considering teachers farther apart on the consistency scale.

Finally, the variance in individual students' scores was decomposed to determine the amount caused by stable teacher effectiveness.

4. Findings

The median correlation for the nine subtests was found to be .398. This result comes very close to Brophy's (1973) overall median of .39 for correlations between successive years. Previous studies had reported correlations ranging between -.12 and .78. The present results indicate that the stability of teacher effects in the fifth grade is similar to that reported in earlier research reports.

The teacher "effect" can be used to accentuate the impact of teacher stability and effectiveness when comparisons are made between "best" and "worst" teachers. Further evaluation of the importance of the consistency component of teacher effectiveness by decomposing the variance of individual students' scores indicated that only 3 to 7 percent of the variance in scores could be assigned to this difference among teachers.

5. Interpretations

The correlations determined in the study were interpreted, in a manner similar to other studies, as evidence of a stable teacher effect.

Limitations to the analysis were reported as follows:

(a) The gains measured impact on class mean scores, but did not take into account performance measures such as the spread of scores.

(b) The gains overlook the possibility of absolute teacher effects; that is, the possibility that all teachers raise the mean score of their class during both years by the same amount. This would produce correlations of zero. Thus, teachers could have impact, but no relative difference in effectiveness.

(c) The Metropolitan Achievement Test may not represent the cognitive goals of teachers.

(d) Residual gain scores may not take into account other factors affecting teachers' relative effectiveness.

Methods of quantifying important teacher behaviors may be inadequate at this time. The researcher suggests future research be designed to as to provide ways of identifying and isolating behaviors which will be predictable and have a stable influence on student learning.

Critical Commentary

1. The author seems to contradict himself early in the report. He states that earlier studies used varied methods and thus made it difficult to synthesize findings. He then says that the present study uses "... the same basic methods as earlier studies..." More discussion at this point would have provided clarification for the reader.

2. The presentation, in general, could prove confusing to the reader. The report is divided into sections labeled Method, Results, and Discussion. However, it seems that the description of the study jumps around, with certain points pertinent to Method being discussed in the Results section and so on.

3. The topic of teacher effectiveness is certainly an important one, especially in these days of emphasis on teacher accountability. Studies in this area are important for mathematics education as well as other academic areas.

4. The author carefully indicated some of the limitations of the study. Of great importance is the question of the suitability of the Metropolitan Achievement Test as a measure of teachers' goals. For example, it seems reasonable that instructional emphasis had been altered (in mathematics or other areas) from 1959 to the date of the study.

5. As well as the concern expressed in point 4, it might be wise to consider research using affective as well as cognitive goals.

Charles E. Lamb
The University of Texas at Austin

THE DEVELOPMENT OF CHILDREN'S UNDERSTANDING OF PROPORTION. Chapman, R. H. Child Development, v46 n1, pp141-148, March 1975.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by David F. Robitaille, University of British Columbia.

1. Purpose

The major purpose of the study was to determine whether or not concepts of proportionality develop before the formal operations stage.

2. Rationale

Piaget and Inhelder found that concrete operational children were misled by quantitative cues when given tasks requiring probabilistic reasoning. Given two collections of objects differing in one attribute (e.g., color), such children typically chose the container with the greater number of members of a designated class rather than the one with the greater ratio. Conflicting results have been reported by others, but the author notes that in two such studies all of the tasks presented to the subjects could be resolved by attending only to the relative numbers of objects regardless of the ratios.

3. Research Design and Procedure

Ten boys and 10 girls from each of grades 1, 3, and 5 as well as 10 male and 10 female college students were administered a 23-item test on understanding of proportions. Test items were of three kinds: six one-container (1C) items where a subject was shown a number of brown and yellow candles and asked to predict the result of a random draw, 14 two-container (2C) items where the subject decided from which of two containers to draw in order to obtain a candy of a designated color, and three conservation-of-proportion items. After each 1C and 2C item was presented, the subject drew a candy at random. Subjects were awarded one point if the result of their draw corresponded with their prediction; otherwise, the experimenter received a point. Subjects were asked to explain their responses to three of the 2C items and to the three conservation items. These protocols were recorded and scored.

4. Findings

No significant effects for sex, age, or their interaction were found for the 1C items and the six corresponding 2C items. Performance on the remaining 2C items was found to be significantly affected by sex and by grade, with boys outperforming girls. Children's scores on the conservation items were low, with only grade 5 boys (40% correct) exceeding the chance level of 33%. With one exception, children's errors were attributable to their choosing the container with the greater number of candies of the target color. College students' scores on all items were very high.

Chi square analyses of the verbal response data resulted in significant differences attributable to the effects of grade level and sex.

5. Interpretations

The author states that the results support Piaget and Inhelder's view regarding the development of understanding of proportions. Authors of previous studies are criticized for failing to determine how children actually solved the items. It is suggested that further research is needed to determine the specific nature of the sex differences reported in the study.

Critical Commentary

The paper raises many concerns. First of all, the author has not clarified the relationship between understanding of proportions and probabilistic thinking. Does a student's failure to respond correctly to a probabilistic reasoning task ("Which jar would you pick from in order to get a brown candy?") necessarily indicate a lack of attainment of formal operations in understanding of proportions? A student could conceivably understand proportions and yet not make use of that knowledge in responding to the tasks. Secondly, the reward system that was used may have affected the children's performances. If a child chooses the correct (i.e., the more likely) color or container on a given task but then draws a candy of the "wrong" color, might this not affect his choices on subsequent tasks? Thirdly, in spite of the author's criticism of other studies, 16 of his 23 items (about 70%) could be correctly solved on the basis of the numbers of candies involved with no reference to proportions. In only two of the 2C items, the author states that the fifth graders' 75% correct performance "was far inferior" to college students' 100% correct performance. He then argues that this result shows that fifth graders have "not yet attained formal operations" despite the fact that first and third graders' mean performance levels on these items were both 40%. It would appear that no criterion performance level had been decided upon in advance, and the author's argument is not very compelling.

The concerns listed here seriously detract from the credibility of the author's results and their subsequent interpretation. Caution should be exercised in interpreting these results as supporting Piaget and Inhelder's findings.

David F. Robitaille
University of British Columbia

COMPUTATIONAL ERRORS MADE BY TEACHERS OF ARITHMETIC: 1930, 1973.
Eisenberg, Theodore A. Elementary School Journal, v76 n4, pp229-237, January 1976.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Leland F. Webb, California State College, Bakersfield.

1. Purpose

To compare computational errors on a 25-item diagnostic computational test made by teachers of arithmetic in 1930 and in 1973.

2. Rationale

Many significant changes have occurred in mathematics education and in the mathematics curriculum during the past four decades. The 1930s possibly could be considered infancy days in mathematics education, but the 1930s predate the emphasis on specific subject matter and the large number of reforms that have occurred in the school mathematics curriculum. One "time-honored" goal that has survived these changes and reforms is that the elementary-school teacher should be able to compute. This ability is considered to be a requisite for understanding the real number system. Studies by Glennon and by Leonard have considered questions about mathematical understanding and errors at various educational levels.

Certification requirements have also changed substantially. Requirements in 1930 for certification in Ohio elementary schools were "graduation from a 'first grade' high school (or equivalent) and graduation from a two-year normal school" (p. 230). In 1973, the requirement for a similar teaching position were a bachelor's degree from a four-year, accredited institution, with academic and professional courses specified. Comparative studies are necessary to assess the effects on our educational programs.

3. Research Design and Procedure

The subjects for this study were 22 elementary-school teachers who had originally participated in a study by Guiler in 1930, and a similar group, matched by grade level, of 22 teachers enrolled in a 1973 graduate-level mathematics course at The Ohio State University. The number of teachers in each grade were: 2, grade 1; 5, grade 2; 3, grade 3; 4, grade 4; 1, grade 5; 2, grade 6; 4, grade 8; 1, grade 9. In cases where more than the required number of teachers at the appropriate grade level were in the 1973 course, a random sample was selected.

Because of the detailed reporting of 25 of 50 test questions in the 1930 study, comparisons were possible between the two groups on those 25 questions. The test used was the Guiler-Christofferson Diagnostic Survey Test. All 50 items were administered in the 1930 study, but in 1973 only the 25 used for comparison purposes were administered. The test covered five areas of computation: (1) whole numbers, (2) fractions, (3) decimals, (4) practical measurements, and (5) percentage.

The results of the tests were compared in the following manner:

1. Mean scores and types of errors were analyzed for each examination area.

2. An item analysis to compare types of errors was reported.

3. The response patterns of the top 27 percent, the bottom 27 percent, and the middle group were compared across each sample.

One-tailed *t*-tests were used to compare performances of the two groups on the total test score and for each of the five subscale scores. The populations for the comparisons were the total group and the following subgroups: (1) the top 27 percent, (2) the bottom 27 percent, and (3) the middle group. A total of 24 *t*-tests were calculated (6 for each group by four groups). In all cases, the level of significance was .01.

4. Findings

The *t*-tests indicated that 10 of 24 differences were significant:

1.	Total group, total test score	$t_{.01,42} = 1.83$
2.	Middle group, total test score	$t_{.01,18} = 3.07$
3.	Total group, whole numbers subscale score	$t_{.01,42} = 2.80$
4.	Total group, measurement subscale score	$t_{.01,42} = 2.66$
5.	Top 27 percent, whole numbers subscale score	$t_{.01,10} = 2.36$
6.	Top 27 percent, percent subscale score	$t_{.01,10} = 2.24$
7.	Bottom 27 percent whole numbers subscale score	$t_{.01,10} = 2.45$
8.	Bottom 27 percent, measurement subscale score	$t_{.01,10} = 2.24$
9.	Middle group, fraction subscale score	$t_{.01,18} = 3.51$
10.	Middle group, measurement subscale score	$t_{.01,18} = 2.77$

Overall, the teachers in the 1973 sample performed significantly better on the examination. The comparison in performance of the top 27 percent of each group did not differ significantly. The same results were found for the comparison between the bottom 27 percent of each group. The middle group of 1973 teachers was significantly more accurate in computation than their 1930 counterparts.

With respect to the subscale scores of the examination, the 1973 total group was significantly higher in the whole numbers category and the measurement category. For the top 27 percent, differences were found for the whole numbers category and the percent category. For the bottom 27 percent, differences were found for the whole numbers category and the measurement category. For the middle group, differences were found for the fractions category and the measurement category. All differences favored the 1973 group.

The top group of both the 1930 teachers and the 1973 teachers displayed significantly more skill than their respective bottom groups. This was true in all cases except for measurement in the 1973 group.

5. Interpretations

Teachers still have a tremendous amount of trouble with both percent and decimals. Progress has been made, particularly in the areas of whole numbers, fractions, and measurement. Teachers are more accurate in computation in 1973 than in 1930, but still at least 59 percent of the teachers have difficulty answering a simple percentage problem.

The two groups used in the study are not comparable because there is no 1973 counterpart in academic background to the teacher of 1930. The 1930 teachers did significantly better than 1930 college freshmen. But in 1973 all teachers in the sample were college graduates, some with master's degrees. Hence, the differences observed between the 1930 and 1973 teachers becomes more suspect. The worth of the "mathematical revolution" educating elementary school teachers in computational skills becomes questionable.

Comparative studies, while necessary, can be discouraging considering the slight improvement in mathematical skills between the 1973 groups and the 1930 group. There is still room for improvement.

Critical Commentary

This study was interesting; usually insufficient records are kept of studies conducted 30 to 40 years ago to allow the type of comparisons made in this study. In that regard, the study has considerable merit. As the author indicates, the findings would suggest that mathematics educators continue to question the revolution which the curriculum in mathematics seems to be undergoing.

There are several questions which need to be raised that tend to negate some of the findings. The questions deal with the decisions made about the procedures used in the study:

a. The most critical problem in the study is the apparent error in the calculation of the critical values in the *t*-tests. Of the 10 significant *t*-tests, 5 appear to be incorrect. Of the significant *t*-tests listed earlier in this abstract, 1, 5, 6, 7, and 8 do not reach the critical value required for a one-tailed *t*-test. The critical values required are $t_{.01,10}=2.764$,

$t_{.01,18} = 2.55$ and $t_{.01,42} = 2.42$. Since these five tests are not significant at the .01 level, the results of the study as they are interpreted by the author become suspect. A reinterpretation of the findings in the light of this information reinforces the fact that the real difference between the 1930 and 1973 groups is in the middle group of teachers.

b. The use of a one-tailed *t*-test is questionable. The author gives no indication as to why a one-tailed test is selected, in lieu of a two-tailed test.

c. No hypotheses are formally stated in the study. Yet, 24 statistical tests using directional hypotheses are conducted. It would have been appropriate to state the directional hypotheses in the study and to justify their use.

d. Reference is made to a significant difference in skill between the top and bottom groups of both the 1930 and 1973 groups. No statistical table is presented to support this statement.

e. Guiler administered a 50-item test, but reported data on only 25 items. Why? No indication or reason is provided as to what happened to the other 25 items. Eisenberg administered only the analyzed items to the 1973 group. What effect does this have, if any, on the test results? A reliability coefficient could have been easily calculated and reported, but this was not done.

Eisenberg does report that the sample was not identical, and it is obvious that he was restricted in conducting comparisons to only what Guiler had done. Studies of this sort are important, but one must reconsider the findings in light of the comments made above, particularly because 50 percent of the reported statistically significant findings were, in fact, not significant.

Leland F. Webb
California State College, Bakersfield

BIAS IN PREDICTION: A TEST OF THREE MODELS WITH ELEMENTARY SCHOOL CHILDREN. Frazer, W. G.; Miller, T. L.; Epstein, L. Journal of Educational Psychology, v67 n4, pp490-494, August 1975.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Elizabeth Fennema, University of Wisconsin-Madison.

1. Purpose

To examine the fairness (lack of bias) of three alternative prediction methodologies: the traditional single equation regression model and the Cleary and Thorndike two-equation models.

2. Rationale

In recent years standardized tests, particularly intelligence and achievement tests, have been accused increasingly of being biased. Many authors have contended that current normative-based tests (primarily standardized on and directed to white middle-class populations' values and experiences) are essentially unfair and unrepresentative for subjects of culturally different backgrounds. Thus, the bias, conceptualized as using tests which were standardized on subjects whose background experiences were of a different nature than the tested sample, as well as the bias in test items drawn from such samples, precludes an equal opportunity for success on standardized assessment instruments. However, another point of view is that a test can be said to be fair or biased to the extent it provides equity in predictive information. Equity thus becomes equality in the precision of prediction of academic success for different subgroups. In this view test bias becomes primarily a technical relationship between the instrument and the criterion.

According to Cleary, bias occurs if members of a subgroup obtain predicted scores which are systematically higher or lower than those received on the criterion. Cleary believes that the common regression equation is generally unfair since the mean of each subgroup would not typically be equal to the mean actually obtained on the criterion. Cleary's solution is to use a separate regression equation for each subgroup.

Thorndike considers a selection procedure to be fair only if it admits individuals of each subgroup in such a way that the number admitted is proportional to the number who would succeed if all applicants were admitted.

Each of the procedures appears to result in somewhat different biases if subgroups of a population differ from the obtained population mean. The Cleary procedure penalizes the higher scoring subgroup, while the traditional procedure would result in a selection of a disproportionately higher number from the subgroup with the lower mean. The Thorndike procedure results in the selection of a number from each subgroup which is proportional to the number of that subgroup who are potentially successful on the criterion.

3. Research Design and Procedure

In order to test the three procedures, a sample of 101 female and 95 male fifth-grade children was drawn from a large metropolitan school district. The children's grade point average (Y), reading achievement scores (X_1), and arithmetic achievement scores (X_2) were obtained from the school records. The females were defined as one subgroup and the males as the other.

Predicted scores ($\hat{Y}$) were calculated for each individual by regressing the grade point averages on the other two variables using both the traditional regression procedure and the Cleary two-equation procedure. The multiple correlation between the predictors and the criterion was .81. Selection under each of the three models was examined using four different selection ratios. The definition of success on the criterion was set so that the results would maximize the differences between the three models. If a proportion, p, of the total group was to be selected, the criterion scores in the top 100p percent of the total group were defined as successful. For example, under the first selection ratio examined, 12.8% of the total group was to be selected. Individuals having scores on the criterion which were in the top 12.8% were defined as successful.

4. Results

The males' grade point averages as well as reading and arithmetic achievement scores were lower than the comparable female scores. The Cleary procedure resulted in the selection of the fewest males. The traditional procedure selected more males and fewer females than would actually succeed. The traditional procedure produced bias in favor of the subgroup with lower mean criterion scores (males), while the Cleary procedure produced the opposite. The Thorndike model did not result in selection bias in either direction.

5. Interpretations

The three models differ substantially in the selection of individuals based on a quota system. When the predictive value of any instruments is unity, there is no bias. However, if correlation between criterion and prediction is less than unity (.81 in this case), then the models differ in terms of the individuals selected and the accuracy of the selection.

Critical Commentary

The issue addressed in this study is a highly significant one in today's world of limited university and professional school enrollments. Criteria for admission to such institutions includes the use of various examination scores as predictors of success. If either the test or prediction equation used includes bias, then some subgroup will be denied equity in admission procedures.

The authors of this paper clearly explicate the rationale for fairness in item selection. However, their rationale for selection of a prediction procedure, while accurate, appears limited in scope. For a more complete discussion of this problem see: Reed, C. W. "Statistical Issues Raised by Title IX Requirements on Admission Procedures." Educational Testing Service, Princeton, New Jersey, 1976.

Elizabeth Fennema
University of Wisconsin-Madison

24

WHAT DO MATHEMATICS TEACHERS THINK ABOUT THE HIGH SCHOOL GEOMETRY CONTROVERSY? Gearhart, George. Mathematics Teacher, v68 n6, pp486-493, October 1975.

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

1. Purpose

To examine the attitudes of secondary school mathematics teachers toward several aspects of the contemporary high school geometry course.

2. Rationale

Teachers are ultimately responsible for the implementation of curriculum changes. Hence, their opinions, attitudes, and preparedness should be known and taken into account before major changes are made in the mathematics curriculum.

3. Research Design and Procedure

A 57-item questionnaire was developed to determine teachers' reactions to the "standard geometry course," which was operationally defined to be "the usual one-year geometry course, based on Euclid's development and using a text influenced by such curriculum groups as SMSG (or an earlier text)." Items were written to reflect recent criticisms and proposals for this course as found in various journals. The items were in the form of given statements to which teachers responded using a five-point scale.

The questionnaire was sent to a random sample of 999 secondary school mathematics teachers from across the United States. Usable responses were received from 605 teachers. Thirty non-respondents were surveyed by telephone calls. These calls revealed that most of the non-respondents had either moved from the given school or did not feel qualified to respond to the questionnaire.

4. Findings

A majority of the teachers (86%) thought the course was valuable to students. They also felt (73%) that the usual one-year course was about the right amount of time to devote to the subject.

Most teachers (52%) were of the opinion that the course should not be made less formal and rigorous. The time spent on writing proofs seemed about right to most teachers (59%) and most thought that writing proofs was not too difficult for the average college-preparatory student (89%).

Teachers tended to support a number of changes in the course: make the approach more concrete (68%); include coordinate methods (85%);

include symbolic logic (67%); and include vector methods (54%). On the other hand, most teachers agreed (57%) that there is no extra time in the standard course for additional topics.

Most teachers, while favoring the inclusion of particular topics, did not favor basing the course on these topics: basing it on coordinate methods (44% against, 35% undecided); basing it on vectors (66% against, 29% undecided); and basing it on geometric transformations (57% against, 32% undecided). Less than one-half of the teachers (40%) favored a unified course integrating geometry and algebra.

The following portion of teachers reported that they would need additional workshops or courses before they could teach the following topics: coordinate methods in geometry (8%); vector methods in geometry (28%); geometric transformations (37%); symbolic logic (20%); elementary topological concepts (54%); and non-Euclidean geometries (41%).

5. Interpretations

Teachers view geometry as an important part of the secondary school mathematics curriculum. Major changes in the mathematical development of the course are not advocated, although the inclusion of topics such as logic, coordinates, vectors, and transformations appeal to many teachers.

The survey suggests that teachers with training in a topic (or who have taught the topic) are more interested (than other teachers) in its inclusion and are more likely to believe that average students can learn the material. More-experienced teachers also indicate that more students like geometry and learn the necessary concepts and skills. Teachers' backgrounds in mathematics are positively correlated with their feeling of preparation to teach new material and their interest in doing so.

There is a definite need to provide teachers with information about new approaches to geometry.

Critical Commentary

Few people would argue against taking into account teachers' opinions and experiences when making curricular decisions. The best way to do this, however, is a difficult problem. Some of the difficulties with interpreting survey data associated with curricular questions are reflected in the following questions.

1. Did the teachers surveyed interpret the standard geometry course as it was defined? Were the teachers inclined to react to the geometry course taught in their school whether or not it fit the definition?

2. Is it appropriate to react to the inclusion of particular content in a course without consideration of the broad goals of the course? Is it likely that the teachers in the sample

would be in agreement with regard to the goals for the secondary school geometry course?

3. Do the data reflect the general attitude of society to "go along with the status quo" or do they represent careful consideration of the pros and cons of various decisions?

4. Is it possible to reconcile responses that appear to be very contradictory? Do many teachers really believe that the course should be made more concrete and at the same time feel that symbolic logic should be added to the course content? To what extent can more topics be added while it is suggested that there is no additional time for new topics?

This study does generate interesting hypotheses about important ideas and issues. Many of these could be profitably followed up using individual interview techniques. This would provide an opportunity to clarify questions as needed, determine the rationale for answers, and resolve contradictory responses. To avoid some of the problems previously mentioned, future surveys might make use of items that require the respondent to choose between alternatives.

Douglas A. Grouws
University of Missouri-Columbia

ACHIEVEMENT TEST SCORE DECLINE: DO WE NEED TO WORRY? Harnischfeger, Annegret; Wiley, David E. Chicago: CEMREL, Inc. December 1975.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by F. Joe Crosswhite, The Ohio State University.

1. Purpose

This monograph, prepared with assistance from the Ford Foundation, examines the reported decline in student achievement test scores. It examines the questions, "Are the reported declines real or merely artifacts of the tests?" and "If they are real, why do they occur?"

2. Rationale

Recent reports on test-score declines have spurred groups of experts from measurement, curriculum, and school administration into public discussion and debate. This has produced some evidence about pupil achievement test scores, mostly supporting the decline hypothesis, but some contradictory. These apparent contradictions demand investigation. Most importantly, the magnitudes and consistencies of test-score changes and their implications for educational policy demand a detailed analysis.

3. Research Design and Procedure

Information was gathered on major tests, their contents, scaling procedures, their changes over time, and on characteristics of the tested populations and their changes over time. Such data are reported for the following tests over the indicated time spans:

Scholastic Aptitude Test (SAT), 1957-1975
Preliminary Scholastic Aptitude Test (PSAT), 1959-1974
American College Testing Program (ACT), 1965-1974
Minnesota Scholastic Aptitude Test (MSAT), 1958-1972
Iowa Tests of Educational Decelopment (ITED), 1962-1974
Iowa Tests of Basic Skills (ITBS), 1965-1975
Comprehensive Tests of Basic Skills (CTBS), 1968-1973
National Assessment of Educational Progress (NAEP), 1968-1974

For each test, data were collected for at least two points in time that were amenable to interpretation because of minimal changes in content, scaling, or test population or because evidence was available on the extensiveness of changes in these characteristics. Data which were severely flawed in those aspects were omitted.

Descriptive data are presented for each test with respect to content, scaling, and test population. Analyses are based primarily on mean scores reported for subscales, grade or age level, sex, or other test population characteristics where such data were available. Most data are presented graphically to reveal trends over time. Discussion and interpretation are based on differences but no statistical tests of significance are reported. Some data are reported as percentages or proportions of students' attaining certain test scores. Test-population characteristics and changes in these characteristics are reported as available.

In a search for possible explanations of the apparent decline in test scores, data are reported on the social and educational context within which the scores were obtained. Such factors as school curricula, course enrollment patterns, amount of schooling, ethnic distributions, socio-economic status, pupil-teacher ratios, school attendance, and teacher characteristics are examined.

4. Findings

Nearly all reported test data showed declines for grades 5 through 12 over the past decade and this was true for all tested achievement areas. No evidence of decline was found for the lower grade levels (2-4)--in fact, there was some evidence to suggest slight increases in achievement at these levels. The declines observed were more pronounced at higher grade levels. Analyses of the tests and test-takers indicated that the declines were real, not artifacts of sampling either in test content or test populations.

Specific findings of most probably interest to readers of I.M.E. include:

(a) Both the verbal and mathematics scores on the SAT peaked in 1963 and declined steadily from that year through 1975. The decline in verbal scores was clearly more pronounced than that in mathematics and was more drastic for females than for males. The proportion of scores above 700 declined 52 percent for the verbal test and 15 percent for mathematics between 1966-67 and 1974-75.

(b) Results from the ACT follow the SAT trends showing sharper declines in English achievement than in mathematics achievement. Even more pronounced was the downward trend in social science achievement. Among the four achievement areas tested by ACT, only natural science scores seemed untouched by time--a finding not supported by the NAEP data.

(c) On the ITED tests, using data only from the state of Iowa, the mean scores on all seven subtests have declined since the mid-sixties at all assessed grade levels, grades 9 through 12. The decline in Quantitative Thinking scores was as pronounced as that of any other area tested.

(d) National averages of all test scores on the ITBS were available for 1955, 1963, and 1970. The pattern was one of general increase, on all subscales, from 1955 to 1963. From 1963 to 1970, the national data show consistent drops on the majority of the subscales: Reading, Language and Mathematics Skills. Continued increase existed only for Vocabulary and Work-Study Skills.

Examination of the social and educational context of the achievement test-score decline failed to reveal any clear causal relationship, although the authors conclude that curricular change and changing enrollment patterns are likely strong influences on tested achievement in mathematics and English. The data reveal a considerable decline in academic course enrollment which is largest for English, followed by mathematics and then natural sciences--closely paralleling test-score decline patterns.

5. Interpretations

The authors interpret the data as revealing a real decline in achievement test scores, essentially independent of test content or population changes. They identify curricular change as a likely factor contributing to the decline. They offer several conjectures (e.g., the differential effect of television viewing habits on different age groups) as to other potential sources of the apparent decline. A principle interpretation is that the data present a compelling argument for further research to provide more sensitive assessment and data on the social and educational context which might reveal cause-and-effect relationships.

Critical Commentary

The authors of this monograph have made a valuable contribution in bringing together in one source such voluminous achievement test data. Perhaps the volume of that data is its most serious limitation. It may be that an abbreviated summary, written for a more general audience, would have greater potential for the impact desired. The data suggesting decline are so consistent across tests and grade levels that only the most pure at heart would fault them for not applying statistical tests of significance. They have raised important questions to which the educational research community should respond. They have even offered reasonable conjectures to guide that response. While they try hard to avoid value judgments, I think their answer to the title question, "Should we worry?", is clearly "Yes!".

F. Joe Crosswhite
The Ohio State University

TRANSFER OF LEARNING ON SIMILAR METRIC CONVERSION TASKS. Houser, Larry L.; Trueblood, Cecil R. Journal of Educational Research, v68, pp235-237, February 1975.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by James Fey and Linda Rosen, University of Maryland

1. Purpose

The purpose of this study was to determine whether students who develop facility in making unit conversions between selected metric units for length could demonstrate mastery of similar conversions involving other units of length, volume, and mass without explicit instruction on those tasks.

2. Rationale

The strongest argument for adoption of the metric system of measurement in the United States is that the system's uniform procedure for constructing sub-units and multiples of base units greatly simplifies learning, retention, and use of the measurement scheme. While this hypothesis receives wide intuitive support among mathematics teachers, there is little experimental confirmation of the proposed benefits.

3. Research Design and Procedures

Subjects for the study were 99 prospective elementary school teachers who scored less than 7 correct on a 21-item metric conversion pretest. Each subject studied a computer-mediated tutorial instruction module on basic terminology and seven of the possible conversions between linear units of the metric system. Upon completion of the instruction, each subject took a 14-item linear conversion posttest (7 items on conversions covered by instruction, 7 not covered by instruction). Those subjects who mastered linear conversion then received a 7-item mass conversion posttest and a 7-item volume conversion posttest.

4. Findings

Only 54 of the 99 subjects demonstrated mastery (85% or better) on the linear conversion posttest. Of these 54 subjects, 44 demonstrated mastery on the mass conversion posttest and 48 demonstrated mastery on the volume conversion posttest.

5. Interpretations

The authors interpret their results as support for the hypothesis that learning conversions between selected metric units for length enables subjects to perform other linear conversions, mass conversions, and volume conversions without further instruction. They suggest that the findings might be a guide for design of metric instruction and that the results lend some support to the argument that metric system relations are easier to learn than the complex English system.

Critical Commentary

For mathematics educators who have long been urging adoption of the metric system because its inherent simplicity and uniformity facilitate instruction in measurement, the present study offers encouraging support. But several important questions are left unanswered.

First, the subjects of the study (prospective elementary school teachers) might have learning difficulties typical of young adults facing the metric system after years of studying and using the English system. However, their experience and reaction to the training procedure is very likely quite different from younger students who encounter the metric system for their first measurement experience. The authors carefully qualify each conclusion with the phrase "under the conditions of this study"; the point is that similar studies with other age groups of learners would be informative.

Second, the procedure of administering a one-half hour CAI instructional treatment and immediate 7-item posttests raises further questions about the depth of learning demonstrated by "mastery". The training and test items are all of the form $n \cdot (\text{unit A}) = k \cdot (\text{unit B})$, requiring solution for n or k. It is certainly conceivable that subjects who knew the basic concept of conversion between units of measurement simply committed the metric prefixes to memory and solved the transfer problems in a rote fashion. There is certainly more to understanding measurement conversion. The present study is a start toward studying efficiency of the metric system, but its generalizability is limited.

James Fey and Linda Rosen
University of Maryland

LOW-STRESS SUBTRACTION. Hutchings, Barton. Arithmetic Teacher, v22, pp226-232, March 1972.

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

1. Purpose

To present an algorithm for subtraction which permits computation with a minimum of stress.

2. Rationale

Low-stress algorithms have two basic attributes:

(a) A concise, definable, easily read supplementary notation is used to record every step.

(b) The learner can complete an intermediate step entirely rather than alternating between different kinds of alternate steps.

The author claims that if algorithms can be developed with these attributes that computation can be less stressful for the child doing the computing and that the teacher can more easily identify specific errors and analyze error patterns in computations.

3. Procedure

The low stress algorithm for subtraction requires the following:

(a) Facility in renaming (e.g., 64 can be written 5'4) taught formally

(b) ALL regrouping to be done before any subtraction takes place

(c) All digits in a minuend containing regrouping must be recorded as part of the renamed minuend

(d) Renamed minuend to be written between the minuend and subtrahend

Hutchings then shows the following example but with every step specified.

$$\begin{array}{r} 6\,4\,3\,5\,2 \\ -\ \underline{1\,7\,4\,5\,7} \end{array}$$

```
  6 4 3 5 2
        4 2
- 1 7 4 5 7
```

```
  6 4 3 5 2
      2 4 2
- 1 7 4 5 7
```

```
  6 4 3 5 2
  5 3 2 4 2
- 1 7 4 5 7

  4 6 8 9 5
```

One other example showing all the steps but not involving regrouping across zeros in the minuend, is given:

For regrouping across zeros the procedure is as follows:

```
  7 4 0 0 0 3 2

- 1 5 6 0 2 4 9
```

```
  7 4 0 0 0 3 2
  6 3       2 2
- 1 5 6 0 2 4 9
```

Note the zeros being "borrowed over" are skipped in the first regrouping. After the regrouping they are simply replaced by 9s and the subtraction proceeds.

```
  7 4 0 0 0 3 2
  6 3 9 9 9 2 2
- 1 5 6 0 2 4 9

  5 8 3 9 7 8 3
```

4. Findings

No data are reported.

5. Interpretations

Hutchings claims that error location and diagnosis can be accomplished more readily because all steps in the computation have been recorded. He then gives some examples of some length subtraction and suggests that problems be made to "fit" them.

Although he claims that more research needs to be done, it appears that "low-stress" algorithms appeal to different kinds of learners. Pupils who are usually good at computation accept these algorithms as alternate. On the other hand he reports that children who are having trouble show a marked increase in computational performance when using a "low-stress" form.

Critical Commentary

Teachers who are faced with the responsibility of helping children learn to use algorithms will be interested in Hutchings' "low-stress" forms. Although the reports abstracted here refer only to the "low-stress" form in subtraction, the 1976 yearbook of the National Council of Teachers of Mathematics contains a non-thematic essay in which Hutchings outlines the "low-stress" forms for all the operations.

I have not seen any research which reports in detail how children respond to the "low-stress" algorithms, but I would expect they would have some difficulties in keeping the record tidy. A neat record would be essential if the child, particularly the slow learner, were not to be hopelessly mixed up in maintaining numerals in their proper place and in handling the half-space superscripts. On the other hand, once those problems could be overcome there is no doubt that the algorithm is more explicit and leaves less load on the memory. There is also no doubt that it would be easier for the teacher to spot errors and to take specific remedial steps.

There needs to be some research information about the usefulness of "low-stress" algorithms in teaching computation.

This report appeared in the "Using Research in Teaching" department of the <u>Arithmetic Teacher.</u> The department was designed to present materials and ideas from research studies in a form in which their applicability in classrooms is apparent. However, this focus need not preclude the presentation of information about the study: for instance, brief descriptions of the design of the study and of the data on which the conclusions are based. Hutchings' report contains only one paragraph alluding to research, but presents no specific information. The validity of his statements about the success of the "low-stress" algorithms cannot be determined from this report. Need classroom teachers (as well as researchers) be shielded from details and data to this extent?

Doyal Nelson
University of Alberta

PREDICTION OF PERFORMANCE BY LOW ACHIEVERS: THE USE OF NON-VERBAL MEASURES. Jones, W. P.; DeBlassie, R. R. California Journal of Educational Research v26, pp11-15, January 1975.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by George W. Bright, Northern Illinois University.

1. Purpose

To investigate the efficiency of a non-verbal ability test in predicting future achievement of presently low-achieving students.

2. Rationale

Previous research indicated that a non-verbal measure of reasoning might be an effective predictor for quantitative and vocabulary test scores (but not for reading scores) of low-achieving eighth graders. The present study sought to extend these findings.

3. Research Design and Procedure

The measures used for predicting achievement were the Perceptual Reasoning (P), Spatial Relations (S), and Figure Composite (P + S) subscales of the Primary Mental Abilities Tests (PMAT). "Representative reliabilities" for these subscales ranged from .79 to .90. Subtest P "is comprised of figure grouping items," and subtest S "is comprised of square completion items (4-6 level) and square completion plus figure rotation items (6-9 level)" (p. 12). High (low) performance was defined to be an S + P score above (below) the mean for that grade. The achievement measures were the total reading stanine and the total arithmetic stanine on the SRA Achievement Series. Low achievement was defined to be a stanine of 5 or less.

Data came from grades 5 and 7 of a single school district which participated in the field test of the PMAT in May 1971. Achievement measures were administered in 1971 and in Spring 1972 as part of the regular testing program for that school. Only low-achieving students were included in the sample. At each grade level two pairs of "matching groups" were formed:

Pair 1: (high S + P, low reading) - (low S + P, low reading)

Pair 2: (high S + P, low arithmetic) - (low S + P, low arithmetic)

"Some students with extreme scores were dropped at each grade level to provide equivalent mean scores in the 1971 achievement tests between appropriate groups" (p. 13). The distribution of subjects was as follows: grade 5 reading, 26; grade 5 arithmetic, 32; grade 7 reading, 62; grade 7 arithmetic, 61. The amount of overlap between the two achievement groups at each grade level was not reported.

The 1972 achievement data were analyzed for each pair of groups using a _t_-test for correlated means for matched samples and a chi square test for contingency tables measuring the direction of movement of individual test scores (i.e., 1972 score greater than 1971 score, or 1972 score less than or equal to 1971 score).

4. Findings

At grade 5, the 1972 arithmetic score was higher for the high P + S group than for the low P + S group ($p < .05$). At grades 5 and 7 the proportion of high P + S subjects who attained increased arithmetic scores was greater than the proportion of low P + S subjects ($p < .05$).

5. Interpretations

"Results of this study add cautious credence to the utility of figure test scores in longer term prediction of achievement" (p. 13). All differences favored higher 1972 scores of the high S + P group.

Critical Commentary

1. The description of the research was so brief that it was inadequate.

   a. What are "figure grouping" items, "square completion" items, and "figure rotation" items? Examples would have been extremely useful.

   b. How many items were in each subtest?

   c. What kinds of samples were used to compute "representative reliabilities"?

   d. How were groups "matched"? Why were the "matched" groups of unequal size?

   e. How were "extreme scores" selected for exclusion?

2. The particular sample used in the study is subject to several severe limiting constraints:

   a. The school district, and hence the instruction used in that district, may be substantially unusual. (Why did the district participate in the PMAT field test?)

   b. What was the ethnic make-up of the sample groups?

   c. Was the grade mean cut-off score for high/low S + P groups the national grade mean or the school district grade mean?

   d. How many subjects were classified as low achievers in both reading and arithmetic?

3. The report would have been improved by an early statement of the hypotheses. The data analysis seems to have been done at least as much because of convenience as because of appropriateness.

4. Although "prediction" appears in the title, there was no analysis to determine how well the PMAT scores actually predicted achievement gains. The reported analyses do in fact suggest that the PMAT subtest scores may be effective predictors, but why didn't the researchers complete the analysis by using linear regression?

5. The χ^2 analysis supports the conclusion that high P + S subjects were more likely to attain higher arithmetic scores. The magnitude of the increase was not clearly reported, however. Were the increases large enough to be practically important? Too, one would like to know the P + S scores for the groups that increased and those that decreased. Efficient prediction would seem to demand that higher P and S scores be associated with the groups that attained increased scores.

6. If the hypothesis being tested in the χ^2 analysis had been clearly stated, it would have been clear that a one-tail test (such as Wilcoxen signed-rank) would have been more appropriate than the two-tail chi square test. The observed direction of results of this analysis was the only one consistent with the central direction of the research.

7. The presentation of this study seems to suggest post hoc investigation rather than experimentation. The data were gathered as part of the PMAT field test, which was conducted by one of the researchers. One wonders whether the patterns in the data were noticed before the study was conceptualized. If so, the results would not have importance unless there were further substantiations of the conclusion. If not, the researchers are to be faulted for lack of clarity in explaining the sequence of events.

The overall reaction to this study is skepticism. Too many details are omitted. The conceptualization is not clearly communicated to the reader. The findings, while statistically significant, do not appear to be strong enough to be educationally important. Also, no attention seems to have been given to the relation of the PMAT scores to other measures of reasoning ability which might be used to predict achievement. Do the P and S scores do any better job of predicting than more typical measures? Hopefully the referenced report of PMAT field test data contains this information. As one part of a very long-range investigation of the use of non-verbal measures for predicting achievement, this study may have a useful role. As a single investigation reported on its own merit, however, the study is quite weak.

George W. Bright
Northern Illinois University

THE INFLUENCE OF TWO TYPES OF ADVANCED ORGANIZERS ON AN INSTRUCTIONAL UNIT ABOUT FINITE GROUPS. Lesh, Richard A., Jr. Journal for Research in Mathematics Education, v7 n2, pp87-91, March 1976.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Werner Liedtke, University of Victoria.

1. Purpose

a. To determine whether organizers have a greater facilitating effect when they are given before a unit (advance organizers) than when they are given after a unit (post organizers).

b. To compare the effectiveness of examples and counterexamples as organizers.

c. To determine the relative effectiveness of advance and post organizers for students of different ability levels.

2. Rationale

The term "advance organizer" refers to a type of instructional material that has been hypothesized to be effective in introducing meaningful topics. Meaningful topics are those in which the new material that is to be learned is related in a nonarbitrary fashion to ideas that have already been mastered by the learner.

Since conflicting results of various studies led to the conclusion that "the ability of an advanced organizer to aid learning is debatable" (Peterson et al., 1973), it was hypothesized that advance organizers may be most effective in instructional situations where structural integration is a problem.

3. Research Design and Procedure

A six-hour self-instructional unit on finite groups was constructed. Two fifty-minute video tapes were produced to serve as organizers. One of the tapes dealt with examples, the other with counterexamples of the main ideas of the unit.

Forty-eight students were selected from two university algebra classes over a two-year period. During each of the two years, the same procedure was followed by the same instructor. A midterm examination was given. The results of this test were used to assign subjects to four treatment groups:

	Advance Organizers	Post Organizers
Examples	Group 1	Group 2
Counterexamples	Group 3	Group 4

The treatment was administered during four consecutive two-hour class sessions. Fifty minutes of the last session were used for the writing of the posttest.

A 2x2 analysis of covariance procedure was used to analyze the data. Posttest scores measured the criterion variable and scores on the midterm examination were used as the covariate.

4. Findings

Posttest Adjusted Mean Scores

Organizers:	Advance	Post
Examples	79.16	75.27
Counterexamples	83.71	76.85

The difference between means for the organizers was found to be significant at the .01 level. It was also reported that, at the .10 level of significance, students who received the counterexamples organizer scored better than students who received the examples organizer. No significant interaction was found.

The reliability coefficients (KR_{21}) for the midterm and posttest were .91 and .83 respectively.

The test for homogeneity of regression yielded an insignificant value for F (.46).

5. Interpretations

The concluding remarks deal with the fact that the counterexamples organizer was found to be more effective than the examples organizer. It is suggested that counterexamples can play an important role. Whenever a stock of familiar examples is available, counterexamples may be able to furnish a means of helping the students become aware of the relevant abstractions. It is suggested that further research is needed to determine the best advance organizer and under what conditions examples and/or counterexamples are most effective.

Critical Commentary

In reading through the study, several kinds of questions and comments come to mind. Among these are:

a. In the introduction the author points out that concepts such as "red" cannot be learned or taught only by looking at objects which display this characteristic. He feels that in learning about "redness", it might be helpful to point out other colors as counterexamples. However, if a young child successfully selects a red object from several objects, the child is in fact considering counterexamples or "non-red" objects, even if the fact is not verbalized. Is then the referral to counterexamples during the early stages of learning merely an attempt to have the child verbalize this fact?

The topic of finite groups was selected for this study. It could be argued that this topic is not the most appropriate one when attempting to determine the effect of counterexamples as organizers. It is difficult to talk about the properties (commutativity, closure, et cetera) and especially the systems (cyclic groups, isomorphisms between groups) included in the instructional unit without focusing on counterexamples. Typically, properties or conditions are defined and systems are tested to determine whether or not they meet these conditions. The results of the testing then yield examples and counterexamples.

b. Concrete materials or models were used to prepare the example and counterexample organizers for the unit. Were these materials used by the subjects during the study?

c. Why were the subjects selected over a two-year period? How many were selected each year? (Another study has to be referred to in order to ascertain the sampling procedure.)

d. What is the justification for use of the analysis of covariance design? The results of a midterm examination were used to assign the subjects to the four treatment groups (for equivalence in ability - Randomized Block Design). The same midterm scores were then inappropriately used as a covariate. Why was not an analysis of variance procedure used? The author seems to imply that homogeneity of variance for the four groups existed. However, the results of a test for this were not reported.

e. The first sentence in the conclusion states that a small value of F (.46) indicates that the effectiveness of the treatments did not depend on the sbility level of the subjects. How was this F calculated? Was this test based on standard score coefficients?

f. The value of $p < .10$ was considered to be significant and the discussion is based on this result, which could be classified as rather weak. The main finding which deals with the effect of organizers ($p < .01$) is virtually ignored.

The statement in the conclusion, "In fact most college students also have some intuitive familiarity with subgroups and isomorphic groups" somehow seems to violate the assumption which was made about the students and the topic in the first place.

g. It is true that more research is needed about advance organizers. As an attempt to determine the effect of examples and counterexamples as advance organizers, no definite answers are provided in this study. Perhaps the inclusion of a control group into a study of this type could yield some valuable information.

Werner Liedtke
University of Victoria

ACQUISITION OF UNDERSTANDING AND SKILL IN RELATION TO SUBJECTS' PREPARATION AND MEANINGFULNESS OF INSTRUCTION. Mayer, Richard E.; Stiehl, C. Christian; Greeno, James G. Journal of Educational Psychology, v67, pp331-350, June 1975.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Suzanne K. Damarin, The Ohio State University.

1. Purpose

To examine the effects of specific mathematical aptitudes, background experience, and methods of instruction on subjects' learning of topics in probability as measured by different types of posttest items.

2. Rationale

In earlier studies (Egan and Greeno, 1973; Mayer, 1974; Mayer and Greeno, 1972), interactions between s, cific aptitudes related to subject matter with method of instruction and interactions of method of instruction with type of posttest were observed. The current series of four studies examines these interactions further by manipulating the amount of experience with the subject matter given to subjects prior to instruction, various aspects of instruction, and type of posttest.

3. Research Design and Procedure

. The research plan common to the four experiments used subjects from a paid subject pool. Subjects were given pretests or pre-instructional treatments followed by instructional treatments and composite posttests. Data collected in each experiment were submitted to analyses of variance in order to examine interactions among parts of the experimental treatments or tests. The specific details of each experiment are as follows:

Experiment 1. Four aptitude measures (arithmetic, probability, and permutations test scores; SAT-Math score) were collected from each of 44 subjects who were then assigned to treatment groups. The four treatments varied on two dimensions in a 2x2 design: (1) general (theoretical) vs. formula approach, and (2) presence vs. absence of progress tests periodically throughout instruction. Instruction on binomial probabilities was administered by computer and lasted approximately 1 to 1-1/2 hours. A 30-item posttest contained one item for each combination of levels of three variables: formula vs. story problem (2 levels), type of question (5 levels), and content (3 levels). Analyses of variance were performed and interactions between and among aptitudes, treatments, and posttest subscales were examined.

Experiment 2. Ninety subjects were assigned to six treatment groups for instruction on Bayes' Theorem. Treatments varied in amount of experience with problems prior to formal instruction (no experience, problems without feedback, problems with feedback), and in the instructional approach (general vs. formula). Instruction was administered in booklet form and a 20-item posttest designed to incorporate the same posttest variables as experiment 1 was administered. Interactions between and among experience, instructional treatment, and posttest were examined.

Experiment 3. Fifty subjects were assigned to cells in a 2x2 design in which the first variable was pre-instructional experience (general vs. formula) and the second variable was related to instruction. Subjects in both levels of instruction were trained to solve binomial probability problems using a sequence of eleven steps; in the "cued" treatment these steps were given names, while in the "uncued" treatment they were not. A control group was given cued instruction without any pre-instructional experience. Errors made during instructional treatments were classified and compared; data for the non-control groups were submitted to analysis of variance.

Experiment 4. Forty subjects were assigned to four levels of pre-instructional experience (none, formula only, general only, both formula and general). Subjects were then given cued instruction as in experiment 3, and finally a posttest designed as in experiment 1. Data were submitted to analyses of variance and interactions between and among pre-instructional experience, errors during instruction, and posttests were examined.

4. Findings

Several significant interactions are reported for each experiment. Only experiment 1 dealt directly with aptitudes; interactions of probability and permutations aptitudes with general vs. formula instruction were marginally significant; interactions of aptitudes with other variables were found to be weak at best.

Pre-instructional experience had a significant effect on subjects' performance in experiments 3 and 4, but no main effect in experiment 2. Interactions between experience and treatment were observed in all three experiments; subjects with no prior experience performed better when given the formula or cued instruction than when given general or uncued instruction.

In all experiments formula experience or instruction was at least as effective as general experience or instruction in overall error reduction. However, instruction did interact with posttest item type; subjects in general treatment groups performed better on "story" problems than subjects given formula treatments.

Main effects for cuing, studied in experiments 3 and 4, were found. Cuing was more beneficial to subjects receiving general instruction than to those receiving formula instruction.

5. Interpretations

Formula instruction is more effective in training subjects to solve problems similar to those used in instruction, but general instruction is superior in preparing subjects to apply knowledge to problems in new contexts. Subjects who have specific aptitudes which are related to the topic of instruction learn better from a general treatment. Pre-instructional experience can be used to nurture those specific aptitudes. The nature of subjects' background experiences can seriously affect how they organize new material in an instructional sequence.

Critical Commentary

The authors have carefully manipulated several variables in this intricate sequence of studies, and the results obtained are of potential importance to the design and selection of instructional materials, as well as to researchers.

However, the care exercised in the manipulation of variables in the experimental design is not extended to the reporting of the data. No test data (means, SDs, reliabilities) are reported, and it is not entirely clear how data were treated. It appears that data from each experiment were submitted to several analyses of variance, but the designs of these analyses are not described. The design is a matter of some concern, especially in experiment 1 where the number of subjects is small in comparison with the number of variables, and in experiments 1, 2, and 4 where several nonindependent criterion measures are used. This concern is increased by the marginal significance of many findings.

Controls on the experience of subjects prior to the experiments do not appear to have been very strict. Some subjects were eliminated because they exhibited or reported knowledge of formulae being taught; however, it did not appear that subjects were screened for prior knowledge of or experience with probabilistic concepts as presented in the general instructional treatments.

Suzanne K. Damarin
The Ohio State University

DIFFERENT PROBLEM-SOLVING COMPETENCIES ESTABLISHED IN LEARNING COMPUTER PROGRAMMING WITH AND WITHOUT MEANINGFUL MODELS. Mayer, Richard E. Journal of Educational Psychology, v67 n6, pp725-734, December 1975.

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

1. Purpose

To investigate the effects of using different types of prerequisite experiences as learning situations for different problem-solving competencies selected from elementary computer programming. In addition, the role, amount, and type of practice and its relation to posttest performance were examined for each of the different instructional treatments. A third area investigated was that of an aptitude-by-treatment interaction involving the Ss' mathematical ability and the different forms of instruction.

2. Rationale

Ausubel (1968), Rothkopf (1970), Mayer and Greeno (1972), and others have focused investigations on the role of prior experiences, "meaningful learning sets", in one's memory on the acquisition of new knowledge. Little is presently know concerning how people learn elementary computer science, especially the aspects of programming. Findings from this study could speak to the role of models, rules, types of practice and questions, and other points in developing a pattern for future technical instruction in this area.

3. Research Design and Procedure

The study reported in the article consisted of three separate experiments. The first experiment focused on how two different types of prior learning experiences can be incorporated into instruction. The two types used were the "model" approach which made use of scoreboards, ticket windows, et cetera, and the "flow chart" approach which called on a person's prior work with charts composed of geometric symbols.

The Ss were 80 university students from an introductory psychology course. The experimental design was that of a completely crossed 4x2 factorial design. The four instructional treatments were titled rule, model, flow, and model-flow. The levels of the second factor were related to the presence or absence of a set of eight practice problems.

The Ss were given the instruction via programmed materials. The rule book contained the information but no conceptual framework. The model and flow booklets contained the same information as the rule booklets, but they were supplemented with the appropriate conceptual models. The model-flow group received the conceptual materials given to both the model and flow groups.

The Ss receiving the practice materials were given a set of problem cards containing two problems requiring the generation of a statement, two problems involving the interpretation of a statement, two problems involving the generation of non-looping programs, and two problems involving the interpretation of looping programs.

The instructional materials were followed by an 18-item posttest created around a 2x3 design. The first factor in the test design was whether the S was to write a statement or short program (generation) or interpret a given statement or program (interpretation). The second factor was the level of complexity of the material in the question. This factor had the levels of simple statement, non-looping program, and looping program.

The results of the ANOVA for the data from Experiment I were:

a. Practice had no significant effect overall or in any of the interactions. Hence all practice data were pooled with the non-practice data.

b. Ss using materials incorporating the model approach did significantly better than those using materials not using the model format.

c. There was no overall difference in the performances of the Ss in the model group and the rule group.

d. There was a significant interaction between the method of instruction and the type of posttest item. Ss with instruction in the model excelled on items using interpretation while those in the flow and rule groups excelled on items requiring generation.

e. There was a significant 3-way interaction between instruction, problem type, and problem complexity. The Ss in model and model-flow excelled on interpretation-looping problems, while those in rule and flow excelled on generation-nonlooping problems.

f. Analysis of the results for the Ss in the flow and model-flow groups indicated that the flow approach resulted in poorer transfer to items requiring extension of the material to novel situations (interpretation or looping), while being very good at preparing Ss for doing work similar to that found in the learning materials.

The second experiment was concerned with the effects noted in Experiment I and whether they would be maintained if the Ss were given feedback in working with an example program.

This experiment used a 4x2 factorial design. The first factor was the instructional type: model, rule, model-program, rule-program. The second factor was the S's score classification from the quantitative portion of the SAT: ($\overline{SAT}$ = 560). The instructional materials were the same as the rule and model materials of Experiment I. The rule-program versions were supplemented with an example program.

The analysis of the ANOVA for the data of Experiment II showed that:

a. A significant two-way interaction existed between instruction and type of problem. The model and model-program Ss did better on problems involving interpretation items, while the rule and rule-program Ss did better on items requiring generation activities.

b. The approaches using the model tended to improve the performances of the low-ability students while retarding the performances of the high-ability students.

A third experiment focused on whether the type of practice activity would serve to call forth different learning sets and have a differential effect on learning.

Fifty-six Ss were placed in the cells of a 2x2x2 factorial design. The levels of the first factor were model or rule text. The levels of the second were the types of practice exercise, interpretation or generation. The levels of the third were the ability levels for the Ss.

The results of the corresponding ANOVA showed that:

a. Ss using the model approach were helped most by practice items of the generation type, while the reverse was true for the students in the rule groups.

b. Similar results held true for ability levels and Ss performances.

A supplementary study was conducted to find good predictors for overall posttest and posttest subtest performance. In addition to the SAT score, the score on a set of algebraic story problems seemed to have the best predictive validity.

4. Interpretations

The results of the expereiments indicate that:

a. Initial instruction in computer programming might best employ the model diagram approach as an "advance organizer."

b. Sample programs and flow charting elicit less productive associations than the model approach.

c. Ss using the model materials tend to function well on items requiring interpretation of novel materials, but Ss using the rule materials seemed to function better in situations requiring a more straightforward transfer of their learning situations.

48

d. The model approach may interfere with the learning of high-ability subjects while helping in the learning of low-ability subjects.

Critical Commentary

The results of this set of investigations are of current interest to many mathematics educators. The conduct and design of this set of studies appeared to be well thought-out and conceived. However, several questions remain for the abstractor.

a. With the small number of Ss per cell in the design, what was the power of the tests used in the ANOVA?

b. What was the reliability of the tests used in the experiment?

c. Was the information conveyed to the learner in the models materials really the same as that conveyed to the learner in the rule materials? A move may consist of a statement move accompanied by a picture move, while a rule move may only consist of a statement move. Are these materials of an equivalent nature then?

d. The scores on the posttests appeared to be very low. Are they high enough to state that much learning had really taken place?

e. Much of the hoped-for statistical information was missing, i.e., tables of means, ANOVA tables, and graphs for significant interactions. Such material is helpful to the attainment of a full grasp of the findings.

The experimenter should be congratulated for his careful building of one study into the next to provide us with partial replications of his series of studies.

John A. Dossey
Illinois State University

PERCEPTUAL INFORMATION IN CONSERVATION: EFFECTS OF SCREENING. Miller, Patricia H.; Heldmeyer, Karen H. Child Development, v46, pp558-592, June 1975.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Douglas T. Owens, University of British Columbia.

1. Purpose

These questions were investigated in conservation of liquid situations:

(a) Whether totally screening the liquid would produce a high percentage of "conservation" answers.

(b) Whether children of different ages would respond differently upon removal of the screen.

(c) Whether the extremeness of the transformation (degree of change in height and width) would influence performance.

2. Rationale

In Piaget's conservation tasks, a child is required to ignore several misleading perceptual cues and provide a logical explanation before the child is classified as a conserver. Perhaps these requirements are too stringent and lead to a false diagnosis of non-conservation in some children. Prounounced stimulus cues may draw nonconservers and some new conservers toward a nonconserving answer. Systematic removal of the cues should reveal more clearly how the perceptual information influences conservation performance.

3. Research Design and Procedure

There were 108 kindergarten and 84 first grade-children from two predominantly white, middle-class schools tested. Four children who failed the verbal pretest were rejected. The verbal pretest consisted of giving the child a plastic bag of uncooked popcorn and then asking the child to determine from three other bags which one had more, less, and the same amount of popcorn as the first bag.

In the experiment proper, five sizes [illegible] beakers were used. Each beaker held 1 quart of water with 1 inch of space at the top. The standard container was 15.2 cm tall and 9.4 cm wide. There were two shorter-wider containers 10.5 cm x 11.9 cm and 5.2 cm x [illegible].4 cm and two taller-thinner beakers 21.3 cm x 7.8 cm and 27.4 cm x 6.8 cm. A screen with an opaque curtain on the front was also used.

There were three stimulus conditions:

Condition One, with a typical conservation procedure: After establishing the equality of water in two standard beakers, the water was poured from one of the beakers into one of the shorter-wider or taller-thinner beakers. On the fifth trial the water was poured into a glass identical to the standard.

Condition Two, with fewer perceptual cues: In this condition the glass into which the water was poured in each of the five trials was screened from the Ss view. In each case the child was also provided an empty container identical to the one behind the screen. The five trials of this condition using the same beakers as the previous condition were followed by three typical conservation tests as posttest.

Condition Three, with three progressively increasing levels of perceptual cues: After establishing the equality of water in two standard beakers, (A) water was poured into a different container which was screened, and the conservation question was asked. (B) An empty container, identical to the one behind the screen, was shown and described as "the same as the one behind the screen." The conservation question was asked. (C) The screen was removed to reveal the beaker of water, and the conservation question was asked. The three trials of this condition (each having steps A, B, and C) used a shorter-wider container, a taller-thinner container, and a standard container. The three trials were followed by three posttests--typical conservation tests.

In each of the conditions 1, 2, and 3, four orders of container sizes were used. These four orders were balanced within each grade in each condition.

There were two kinds of criteria for conservation: a conservation judgment and a conservation judgment with adequate explanation. Adequate explanations included compensation, previous equality, irrelevancy of transformation, reversibility, and no addition or subtraction. Two scorers had 94% agreement on whether an explanation was adequate and 98% agreement with respect to the type of adequate explanation.

4. Findings

(a) A comparison of the screening condition with the typical conservation test revealed that fewer misleading perceptual cues yielded more conservation responses among kindergarten children only.

(b) Increasing the amount of misleading perceptual information produced response patterns in kindergarten children which were distinctly different from response patterns of first graders. In particular, at the beginning of the first trial with screening (trial 1A) the majority of children asserted conservation. When the kindergarteners were shown a beaker identical to the one behind the screen (trial 1B) most of them switched to a nonconservation answer, a significant change (McNemar chi square $p < .001$). When the screen was removed (trial 1C), kindergarteners showed a slight increase in conservation which was significant for conservation judgments (binomial test, $p = .03$), but not for conservation with adequate explanations.

At the beginning of the second trial (trial 2A) kindergarteners had a high level of conservation judgments, but it was significantly lower than on trial 1A (chi-square, $p < .05$). Conservation with explanation scores were low and remained low throughout trial 2. Thus experience on trial 1 affected conservation explanations more than conservation judgments on trial 2A. Again conservation judgments decreased sharply in trial 2B (binomial test, $p = .008$), but the decrease from 2B to 2C was not significant.

First-grade children were relatively unaffected by changing amounts of perceptual information. They demonstrated more conservation on the typical conservation posttests than at any other time in the sessions.

(c) The four shapes of beakers had no differential effect, but produced significantly less conservation than pouring water into another standard beaker.

(d) There were no significant differences due to sex or order of presentation of containers.

5. Interpretations

Many kindergarten children appear to hold two conflicting beliefs--a belief in nonconservation and a belief in conservation which can be supported by a logical explanation. The belief expressed in a particular situation depends upon the amount or type of perceptual information available.

If logical explanations reflect operations, then many of the children in this study possessed the underlying cognitive operations normally attributed to "true conservers." However, later in the experiment, many of the children regressed to nonconservation and did not use these operations. Perhaps they did not always realize when the operations available to them applied.

Conservation of liquid quantity is not an all-or-none ability, but consists of several levels. Many young children considered to be nonconservers by the standard procedures may have a rudimentary understanding of the invariance of liquids which they can demonstrate under facilitating conditions. Thus to categorize a child as a "conserver" or "nonconserver" on the basis of the standard test is inaccurate. This study contributes a step in the direction of a more refined test consisting of a number of items, varying from full perceptual support to items with many irrelevant cues at the other.

Critical Commentary

The idea of systematically varying the amount and kind of perceptual information available in a conservation test has been and can be revealing. This concept opens up a whole series of questions about any number of conservation settings in addition to conservation of liquid quantity.

What explanation did children give for changing from a conservation answer to a nonconservation response in the face of new perceptual information? It appears that this would be helpful in the interpretation of these results.

Different levels of performance on conservation tasks have been observed and acknowledged by Piaget and many others. The interpretation is a matter of definition. Is one willing to classify a performance as "conservation" when it only occurs in the absence of distracting cues, or must the same behavior be obtained in all situations including the presence of misleading perceptual cues?

More information regarding the research design, statistical procedures, and data would have been helpful. For example, was a given child tested under only one or all three stimulus conditions? How large were the groups for each condition? Results were reported only in terms of significant chi squares and percentages. What were the numbers of children who responded in certain ways?

Douglas T. Owens
The University of British Columbia

DIVISION INVOLVING ZERO: SOME REVEALING THOUGHTS FROM INTERVIEWING CHILDREN. Reys, R. E.; Grouws, D. A. School Science and Mathematics, v75 n7, pp593-605, November 1975.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by James M. Sherrill, University of British Columbia.

1. Purpose

The objective of the study was to learn more about how pupils think about division in general and division involving zero in particular.

2. Rationale

The article reports a follow-up to the study described in Grouws and Reys (1975). In the first study, two instructional sequences were implemented to develop the concept of division involving zero. Although significant changes resulted from the instruction, the level of performance on both the post and retention tests left much room for improvement. The limitations of the paper/pencil tests convinced the authors that a clinical approach was also needed.

3. Research Design and Procedure

The nature of the classes, a description of the testing instruments, and a discussion of the instructional lessons are found in the Grouws and Reys (1975) article. Three or four pupils (N is approximately 60) from each of the classes (grades 4, 6, and 8) participating in the first study were randomly selected and interviewed after the division posttest and again six weeks later.

Each interview followed the same format. When the class had completed the division test, the selected students were taken individually to a separate testing area. Four questions, each printed on a separate index card, were asked of each student. The questions were presented in the following order:

1. What is 12 divided by 3?
2. What is 0 divided by 4?
3. What is 8 divided by 0?
4. What is 0 divided by 0?

Each question was read aloud by the interviewer. Subsequent, additional questions were asked only when a correct response was given, a pupil seemed to have trouble verbalizing, the interviewer decided that the pupil's responses were no longer productive, or the pupil was becoming frustrated. The length of the interviews ranged from 7 to 18 minutes. Each interview was tape-recorded and transcribed.

Questions similar to the first three had been included in the instructional lesson, on the practice problems completed in class, and on the division tests. Prior to the interview, the last question was never presented nor discussed in the instruction for the study. The omission was intentional because the investigators felt there were advantages to examining this indeterminate form in the interview setting.

4. Findings

Nearly 20 hours of tapes and the associated transcripts were examined. It would be relatively easy to select excerpts that support many positions. For example, the authors suggest that interviews could be chosen to support both the position that division by zero is well developed for fourth graders and that eighth graders have difficulty with the same concept. They also mention that none of the interviews were representative of any group. The only common element was variety.

The findings stated in the article were:

(a) In order to understand why a non-zero number divided by zero has no solution, a pupil must first have clearly comprehended the inverse relationship between multiplication and division.

(b) One of the most frequent misconceptions encountered centered around whether or not zero is a number.

(c) The question involving zero divided by zero was difficult for all pupils. Zero was the most popular response. The most popular justification for the incorrect response was, "That's what my teacher says." Whether teachers actually say that zero divided by zero is zero should be investigated.

5. Interpretations

The results of the interviews are interpreted in the final section entitled "Summary of Classroom Implications:"

(a) Zero is a number and it should be developed accordingly.

(b) A necessary prerequisite to being able meaningfully to handle zero in division situations is competence in constructing related division and multiplication sentences.

(c) Division by zero is not permissible.

(d) Division by zero is a complex concept. It will not likely be developed in one day or even in one year.

Critical Commentary

First, one must read the Grouws and Reys (1975) article before reading the present article!

The present article should not be read with the expectation of finding a research study using an air-tight research design to test hypotheses. The two articles together present a first-step, hypothesis-generating study. Reys and Grouws have delved into an area of concern to teachers and have offered the teachers many suggestions.

On the technical side, the number correct for zero as a divisor and the number correct for zero as a dividend have been interchanged in Table 1. In the present article it is stated that "the investigators personally interviewed approximately sixty pupils" and this was done by interviewing "Three or four pupils from each of the participating classes...". In the Grouws and Reys (1975) article, however, it is stated that 30 classes participated.

Finally, quite a bit of teaching takes place in the interviews, but this may be justified by the types of information the investigators were trying to collect in the study. The authors do an admirable job of presenting interview data in article format.

James M. Sherrill
University of British Columbia

Grouws, Douglas A. and Reys, Robert E. Division Involving Zero: An Experimental Study and Its Implications. Arithmetic Teacher, v22, pp74-80, January 1975.

PROCESS MODELS FOR PREDICTING THE DIFFICULTY OF MULTIPLICATION PROBLEMS USING FLOW CHARTS. Romberg, Thomas A.; Glove, Richard. Technical Report No. 337. Wisconsin Research and Development Center for Cognitive Learning, July 1975.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Michael Bowling, Denison University.

1. Purpose

The purpose was to determine whether process models contructed using steps identified from flow charts would account for more variance in predicting the difficulty of two-digit multiplication problems than did a process model developed by Cromer (197?).

2. Rationale

In attempting to predict the difficulty of two-digit multiplication problems, Cromer used 14 variables such as: "TDF", the value of the tens digit of the first number; "DCM", the number of digits carried in multiplication; and "NDP", the number of digits in the product. He administered two forms of an 84-problem multiplication test to 238 fifth-graders. The problems were of the form

$$\begin{array}{r} ab \\ \underline{\times\ cd} \end{array}$$

where a, b, c, d $\in \{0, 1, 2, \ldots, 9\}$. Problems with a = 0 but b, c, and d $\neq$ 0, were not included. The problems were generated using a random number routine. The dependent variable for each problem, general difficulty (DIFF), was defined as the proportion of students who failed to obtain the correct solution to the problem. Hence, for a given problem P, $0 \leq \text{DIFF(P)} \leq 1$.

Values of the 14 predictor variables were computed for each problem and DIFF was expressed as a linear combination of those values. Regression weights were used as coefficients. Of the complete model, the factor models (principal components, oblique rotation), and the other reduced models, the "best" alternative accounted for about 75% of the variance.

The authors of the present study hypothesized that certain of Cromer's variables could be expressed as a combination of simpler, nontrivial variables and account for more of the variance.

3. Research Design and Procedure

A flow chart description of the two-digit multiplication algorithm (Rombert and Anglin, 1973) was used to produce the augmentations of Cromer's lists of predictive variables. In particular, Cromer's variables "OA", the number of operation steps in addition, and "OM", the number of operation steps in multiplication, appeared to be insensitive to certain problem differences. For example, the problems 42 x 2, 82 x 41, and 15 x 20 all have OM value 2. The flow chart description suggested consideration of ten new variables, such as: "NDM(NDA)" = the number of decisions that an individual would have to make when going through the multiplication (addition) routine.

Cromer's basic and reduced models were re-evaluated to fit the (computer) statistical package available to the researchers. For each old and new basic model, multiple linear regression weights were used as coefficients to express DIFF as the linear combination of the appropriate variables.

The seven basic and four reduced models were each factor-analyzed (principal components, orthogonal rotation) to produce factor models. R^2 and corrected R^2 were computed for each basic, reduced, and factor model to determine the problem variance accounted for by that model. "Independent" variables which could be expressed as linear combinations of other variables were excluded from the factor analysis.

4. Findings

Corrected R^2 values ranged from .57 to .76 for the various models. Cromer's complete model (14 variables) accounted for 75.58% (corrected) of the variance; the new complete model (24 variables), 76.64%; and a model composed of nine of Cromer's variables plus the ten new variables, 75.85%.

Factor analysis of the latter model produced nine factors for rotation. Corrected R^2 was .7187 for this model; no other factor model accounted for as much as 70% of the total variance. The model comprised of only the ten new variables yielded four factors for rotation. These four factors accounted for 57.75% of the total variance.

5. Interpretation

(a) "The new flow chart variables do produce models that account for somewhat more of the variance in difficulty than do Cromer's models."

(b) All of the process models accounted for less of the variance in difficulty than did the corresponding models comprised of process <u>and</u> digit variables.

(c) In each case the factor model accounted for less of the variance than did the corresponding complete model.

(d) Most of the variables in the basic and reduced models did not account for a "significant" percentage of the independent variance (not unexpected, since some independent variables were "interdependent"). However, the factor analyses fairly consistently yielded four factors which seemed to correlate highly with: a set of multiplication variables, a set of addition variables, a set of variables related to number of digits carried in addition, and a set of variables related to number of digits carried in multiplication. Less consistent factors seemed to relate to order of the numbers, size of the numbers, and the number of digits in the product.

Critical Commentary

1. Since the best of the new models accounted for less than 1% more corrected variance than Cromer's complete model, the authors' interpretation that "somewhat more of the variance" has been accounted for seems rather ambitious, if not naive.

2. Prior to the factor analyses, a correlation matrix for the 25 variables was formed. The authors' claim that all but two of the 24 independent variables correlated "significantly" (no level specified) with DIFF is interesting. For N = 168 and $\frac{24 \cdot 25}{2} = 300$ correlations, what significance level was considered acceptable?

3. This study represents an extended replication of Cromer's study with an overwhelming use of statistics. In addition to a re-interpretation of Cromer's data, why were not the 168 problems administered to a new sample in order to define DIFF better? Surely Cromer's sample is subject to teacher and pupil variables constraining generalizability.

4. The independent variables are discrete with underlying continuity. Would nonmetric multidimensional scaling have been more appropriate than factor analysis?

5. It is questionable that the results "will prove useful in developing a general theory of mathematics learning," as the authors hope. The results tell us little about <u>how</u> children <u>learn</u> two-digit multiplication. They may tell us of the difficulty involved in various steps of the algorithm, but only if we assume that the models investigated closely approximate the models used by the students. Of greater value would seem to be research in which process models were used to develop instructional strategies; then the measured effectiveness of the strategies would provide evaluation of the models. (For an example of such a study, see Holzman et al., 1976).

Michael Bowling
Denison University

References

Cromer, F. E. Structural Models for Predicting the Difficulty of Multiplication Problems. Nashville, Tenessee: Peabody College, 1971.

Holzman, T. G.; Glaser, R.; Pellegrino, I. W. "Process Training Derived from a Computer Simulation Theory," Memory and Cognition, v4, pp349-356, 1976.

COGNITIVE STYLES, SPATIAL ABILITY, AND SCHOOL ACHIEVEMENT. Satterly, David J. Journal of Educational Psychology, v68 n1, pp36-42, February 1976.

Expanded Abstract and Analysis Prepared Especially for I.M.E. by Leslie P. Steffe, The University of Georgia.

1. Purpose

Satterly's purpose is to investigate "the relation among a group test of field independence, a test that assesses preference for analytic cognitive style in a picture-grouping task ..., intelligence and spatial tests, and measures of school achievement."

2. Rationale

Field independent subjects experience information as discrete from the organized field of which it is a part, whereas the perception of field dependent subjects is dominated by the overall organization of the field. On the one hand, a literature review has led to the conclusion that group pencil-and-paper tests of field independence do not define a factor distinct from general intelligence and spatial ability. Thus, since field independence is an example of cognitive style, it seems unlikely that tests of cognitive style can make a contribution to school achievement beyond that predictable from traditional reasoning tests.

On the other hand, various researchers have found that correlations between field independence and verbal comprehension are low in adult populations; one researcher extracted a factor of cognitive style separate from general intelligence among boys; and significant correlations between field independence and mathematical ability have been reported among college students. Although there is no information to link field independence to achievement in mathematics, work in mathematics does seem to demand analytic operations similar to those described as necessary for success on field-independence tests.

3. Research Design and Procedure

Two hundred one boys (mean age 10.8 years, s.d. 3.4 months) in four English primary schools representing the full ability and socioeconomic range in the schools were used as subjects (excepting those whose reading level was two years below age norms). Eleven tests were administered to the subjects: (a) an embedded figures test (EFT); (b) the Gottschaldt Simple Figures Test; (c) a test of preference for analytic cognitive style; (d) a test of mathematics attainment (test C1 of the National Foundation for Educational Research); (e) a test of English comprehension; (f) a vocabulary test (the English Picture Vocabulary test); (g) the Shapes Test of the Differential Test Battery; (h) a spatial judgment test; (i) a test of haptic perception of shape; (j) the Primary Verbal Reasoning Test; and (k) a general ability test (Perceptual Part I of the Differential Test Battery).

A correlational analysis and a principal components analysis were carried out on the eleven tests. Various ANOVAs were conducted using EFT as the independent variable. The groups were defined by 30 field independent (FI), 30 field dependent (FD), and 30 intermediate (I) boys. ANCOVAs, as well as ANOVAs, were reported where linearity of regression between the variate and covariate was indicated. IQ was used as the covariate. Pairwise differences of means were tested using the Tukey test in the ANCOVAs.

4. Findings

(a) The first-order partial correlations (IQ removed) between (1) EFT and mathematics (.26) and (2) EFT and haptic perception (.31) were significant (p .01). The corresponding correlation between EFT and spatial judgment was significant (p .05). The corresponding correlations between EFT and the two verbal tests were not significant.

(b) One-way ANOVA's revealed differences between means in favor of FI boys in mathematics ($F_{2,87} = 9.13$, $p < .01$); vocabulary ($F_{2,87} = 6.25$, $p < .05$), spatial judgment ($F_{2,87} = 3.99$, $p < .05$); and haptic perception ($F_{2,87} = 6.25$, $p < .05$).

(c) EFT was significant in the ANCOVAs only for mathematics and haptic perception. In the case of mathematics, the mean scores were 35.31, 33.50, and 29.47 for the I, FI, and FD groups, respectively. The Tukey test revealed that only the I-FD difference was significant.

(d) A varimax rotation of the first four principal component factors revealed the following four factors:

Factor 1--the verbal tests (5 and 6), the intelligence test (10), and the mathematics test (4).

Factor 2--the two tests of cognitive style (1 and 3).

Factor 3--spatial factor (tests 2, 7 and 8).

Factor 4--perceptual speed (test 11).

5. Interpretations

Satterly, in his discussion of the results, stated:

(a) "...considerable overlap exists between field independence and verbal intelligence (a correlation of .41 existed between the two tests).

(b) The analysis offers support for the existence of a small factor of cognitive style distinct from intelligence and spatial ability...but the factor...is comparatively small and derived from the four-factor solution.

(c) The relationship of the EFT with Test 3 (analytic cognitive style preference test) and their unexpected separability from spatial ability...is, perhaps, explicable by the order of presentation of these tests....

(d) ...cognitive style...does not make an appreciable addition to the prediction by IQ scores of the majority of tests in the battery.

(e) The data suggest that exceptional field independence does not confer advantage in the learning of mathematics, but, rather, that highly field-dependent behavior inhibits high attainment.

(f) ...cognitive style characteristics do affect the responses of children..., albeit only in minor ways when end products, as distinct from strategies of learning, are investigated."

Critical Commentary

The cognitive styles of children have been singled out in publications in the field of mathematics education as being potentially useful to teachers of school mathematics. Of course, the hope is that a teacher's knowledge of the cognitive styles of a group of children would lead to improved mathematics instruction through accommodation of the teacher's instructional style to the children's cognitive style. Satterly's study is a first step in realizing the potential of cognitive style to instruction in mathematics. Being only a correlational study, the end results of learning were investigated--not the dynamics of the learning-teaching process. While the results were weak, they are encouraging. Satterly, as noted in the rationale, correctly hypothesizes that work in mathematics seems to demand analytic operations similar to those described as necessary for success on field-independent tests. This hypothesis was barely tested in his study due to its correlational nature.

Because cognitive style cannot be varied systematically, Satterly's hypothesis is not empirically testable. However, various studies are justifiable due to Satterly's work. One hypothesis is that FI students would be able to acquire information more rapidly than FD students and the FD subjects acquire information best under slower-paced instruction. Interaction between instructional pace and FI-FD needs investigated. Moreover, longitudinal work could be done, where the interest is in achievement in mathematics. In such studies, investigators must be cautious not to attribute causality to FI or FD if results are found due to the correlations reported by Satterly.

I also suggest that investigators interested in cognitive style include investigations of the relation of FI and FD to abstraction in mathematics learning. One hypothesis is that the FI subjects would be capable of making mathematical abstractions in less time and with fewer experiences than FD subjects. Another hypothesis is that the FI subjects would have a better long-term memory of mathematical abstractions than would FD subjects. One last hypothesis is that the FI subjects would be capable of more powerful abstractions than FD subjects, everything else being equal. Obviously, the latter suggestion is fraught with the difficulty of construct definition.

Leslie P. Steffe
The University of Georgia

LEARNING BASIC PRINCIPLES OF PROBABILITY IN STUDENT DYADS: A CROSS-AGE COMPARISON. Schermerhorn, S. M.; Goldschmid, M. L.; Shore, B. M. Journal of Educational Psychology, v67, pp551-557, August 1975.

Critical Abstract and Analysis Prepared Especially for I.M.E. by Gerald D. Brazier, Virginia Polytechnic Institute and State University.

1. Purpose

This study explored the effectiveness of the learning cell, or student dyad, for the acquisition of principles of probability in grade 5, grade 9, and university students. It was hypothesized that the activities of the learning cell would help all students to learn, but would be most effective with the older students. In addition it was hypothesized that mastery of the content could be predicted from students' ratings of the effectiveness of the learning cell activities and their ratings of how much they enjoyed the activities.

2. Rationale

The authors note that the current trend toward individualizing instruction deemphasizes the social aspects of school learning. Since Piaget and others have argued that student-student interaction is important in developing critical thinking and objectivity, effective instructional techniques relying on that interaction should be investigated.

A body of research indicates the effectiveness of the dyadic learning situation. Some investigators have hypothesized that the attention paid by the learner to the development of the teaching steps in a dyadic situation is the important factor in making such a learning situation effective. If that is the case, older students might benefit more, since research seems to indicate that they may be more perceptive.

Probability was chosen as a topic that could be learned at varying degrees of complexity by all subjects in the study. Whether children not yet at Piaget's formal operations stage can learn probability concepts has been questioned, but some research indicates that this is possible.

3. Research Design and Procedure

The learning cell consisted of students teaching other students using orally presented study questions. Each of the subjects was given two homework assignments on the subject of probability to read in preparation for the two days of in-class participation in the learning cell. For the fifth graders (n = 46), assignments included concrete examples and simple experiments to be performed at home. The ninth graders (n = 35) and university students (n = 40) read excerpts from books and articles which treated probability with minimal mathematical sophistication.

The experiment took place during three classes spread over 5 to 7 days. During the first hour a pretest (Form A) was given on the material of both assignments. The students were then instructed to read assigned material on probability and prepare five written questions to be shared with a student partner during the next class. At the beginning of the next class, a test (Form B, parallel to Form A) was given on the first reading assignment.

The subjects were then assigned to partners (self-selected in fifth and ninth grade) and for approximately one-half hour the partners alternated in answering each others' prepared questions. Afterward, the written questions were collected and a test (Form C) was given. Then the second assignment was made. The activities on the third day were identical to those of the second.

Data from the two assignments were pooled. Analyses of variance for age and sex differences were conducted with the following variables: the test scores, scores on the student-prepared questions, ratings of partners by the subjects, ratings of their own learning by the subjects, ratings of enjoyment, and ratings of the readings. In addition, a multiple regression analysis was performed using the final test score as the criterion and all the above variables together with the initial score as predictors. There were separate control groups at each age level to check for unequal difficulty within the three test forms, effects on learning of multiple testing, and effects of time between first and third tests.

4. Findings

The results for the control groups indicated no significant effects in the controlled factors. A repeated-measures 3x2x3 ANOVA for test form, sex, and age, using the test scores as criterion, yielded a significant three-way interaction, a significant age x sex interaction, and significant main effects for age, sex, and test form. Simple t tests on the test form pairs B-A and C-B showed significant differences, with the pair B-A having an appreciably larger t value.

Separate 2x3 analyses of variance on the remaining variables yielded these results:

(a) Questions prepared by females were rated higher by their partners than those prepared by males; however, there were no age differences.

(b) Ratings of partners showed no significant effects.

(c) Grade 5 subjects gave the learning cell substantially higher ratings than did the other groups.

The regression results showed that the initial test scores (Form A) made a significant contribution ($R^2 = .64$), while the remaining variable did not.

5. Interpretations

The authors concluded that learning did take place and that the learning cell is an effective means of learning some probability principles, even with children at the pre-formal operations stage. Even though a significant age effect was indicated by the ANOVA, an analysis of gain scores did not support the hypothesis that older subjects would learn more. The authors contend that confounding effects of greater enthusiasm among fifth graders and ceiling effects for university students "washed out" what might have been significant differences in gain scores.

Finally, several recommendations are made concerning ways in which the learning cell might be modified (e.g., include less reading) and ways in which learning outcomes other than mastery of content (e.g., development of critical thinking) might be measured. It is again reiterated that the success of the learning cell may be due to heightened student awareness of the teaching process.

Critical Commentary

The student dyad is an instructional technique that is certainly worth investigating. The body of literature quoted by the authors raises some interesting questions. It is unfortunate that the study does not address those questions very well. Why would the learning cell be expected to be successful? If it is because of the heightened awareness of the teaching process, as hypothesized several times, then why not deal with that issue? Instead, attitudinal variables were measured. If it is proposed that the social interaction itself is critical, then why was there no control to isolate that variable? The device of each student being his own control is certainly inadequate, because the A-B gain is not independent of the B-C gain.

As a test of the effectiveness of the learning cell, the experiment falls short. Since there were no controls, the learning that occurred cannot be attributed to the dyadic nature of the learning situation; the critical variables were not isolated. It is unfortunate that the ceiling effect eliminated gain scores as a way of obtaining a cross-age comparison. Without extensive piloting of the test instruments, such a result was beyond the control of the experimenters.

The appeals to Piaget--citing him to justify social interaction and then criticising his statements that probability is a concept at the formal operations stage--leave me befuddled. Without a careful examination of what the fifth graders were asked to learn, it is impossible to judge whether formal operational thought was required. The authors' implication that somehow Piaget has been disproved in this instance is not justified by the data.

The report provides a clear presentation of the student dyad and contains an excellent analysis of the literature. The experiment itself sheds little or no light on the questions raised, however, because the one critical variable tested showed inconclusive results--the cross-age comparison "washed out" because of the inadequacy of the test instruments.

Gerald D. Brazier
Virginia Polytechnic Institute and State University

MATHEMATICS EDUCATION RESEARCH STUDIES REPORTED IN RESEARCH IN EDUCATION
July - September 1976

ED 119 678 Blazek, Ron. Influencing Students Toward Media Center Use; An Experimental Investigation in Mathematics. Studies in Librarianship No. 5. 176p. Not available from EDRS. Available from American Library Association, 50 East Huron, Chicago, Illinois 60611. ($6.50)

ED 120 001 Higgins, Jon L. (Ed.) Research Reporting Sections, Annual Meeting of National Council of Teachers of Mathematics (54th, Atlanta, Georgia, April 21-24, 1976). 117p. MF and HC available from EDRS.

ED 120 010 Horwitz, Stephen. Effects of Amount of Immediate and of Delayed Practice on Retention of Mathematical Rules. 35p. MF and HC available from EDRS.

ED 120 011 Martin, Rogers William. An Analysis of Selected Characteristics of the Secondary Mathematics Teachers and Mathematics Curricula in Caddo Parish, Louisiana. 214p. MF and HC available from EDRS.

ED 120 020 Kaffine, Lawrence; Paulsen, Russell. Evaluation of Secondary Mathematics Curriculums in District 15 as They Relate to the Math Background Requirements of Post-Secondary VTAE Students. Final Report. 77p. MF and HC available from EDRS.

ED 120 027 Lesh, Richard A. (Ed.) Number and Measurement. Papers from a Research Workshop. 238p. MF and HC available from EDRS.

ED 120 126 Dunlop, David L.; Fazio, Frank. The role of Abstract Preferences in Determining Student Behavior in Problem Solving Tasks Within a Piagetian Framework. 19p. MF and HC available from EDRS.

ED 121 605 Kilborn, Wiggo; Lundgren, Ulf P. A Contribution to the Analysis of Arithmetic Teaching and Learning. 26p. Document not available from EDRS. Available from ERIC/SMEAC, The Ohio State University, 1200 Chambers Road, 3rd Floor, Columbus, Ohio 43212 (on loan).

ED 121 615 Trent, John; Lee, Ivan E. The Imminent Meeting with Metrics, II. A Report on a Survey of Nevada In-Service Teachers. 6p. MF and HC available from EDRS.

ED 121 624 Suydam, Marilyn N. Compilation of Research on College Mathematics Education. 287p. MF and HC available from EDRS.

ED 121 625 DiConstanzo, James L. The Integration of Logical and Empirical Procedures for Hierarchically Structuring an Individualized Curriculum. 76p. MF and HC available from EDRS.

ED 121 626 Shores, Jay H.; Underhill, Robert G. An Analysis of Kindergarten and First Grade Children's Addition and Subtraction Problem Solving Modeling and Accuracy. 35p. MF and HC available from EDRS.

ED 121 627 O'Reilly, Robert R. Attitudes Concerning the Nature of Mathematics, Mathematics Teaching and Learning: A Partial Replication of the I.E.A. Study. 38p. MF and HC available from EDRS.

ED 121 631 Gregory, John W.; Casteel, J. Doyle. Verbal Environment Correlates with Student Growth in Logical Reasoning Ability. Final Report Summary. 11p. MF and HC available from EDRS.

ED 121 663 Sells, Lucy W. The Mathematics Filter and the Education of Women and Minorities. 20p. MF and HC available from EDRS.

ED 122 858 Klein, Roger D.; Thompson, Murray. The Use of an Individualized Motivation System to Modify Accuracy and Rate on a Computer-Based Number Facts Program. 18p. MF and HC available from EDRS.

ED 123 088 Weaver, J. F. Calculator-Influenced Explorations in School Mathematics: Number Sentences and Sentential Transformations I, II. Project Paper 76-1. 53p. MF and HC available from EDRS.

ED 123 089 Weaver, J. F. Calculator-Influenced Explorations in School Mathematics: A Further Investigation of Third-grade Pupils' Performance on Open Addition and Subtraction Sentences. Project Paper 76-3. 24p. MF and HC available from EDRS.

ED 123 091 Greene, John F.; And Others. The Effect of Extended In-Service Training Curricula Upon the Mathematics Achievement and Attitudes of Elementary Teachers. 31p. MF and HC available from EDRS.

ED 123 132 Cooney, Thomas J. (Ed.) Teaching Strategies: Papers from a Research Workshop. 213p. MF and HC available from EDRS.

ED 123 216 Gordon, Marshall. Choice of Rule-Example Order Used to Teach Mathematics as a Function of Conceptual Level and Field-Dependence-Independence. 14p. MF and HC available from EDRS.

MATHEMATICS EDUCATION RESEARCH STUDIES REPORTED IN JOURNALS AS INDEXED BY CURRENT INDEX TO JOURNALS IN EDUCATION
July - September 1976

EJ 132 354 Goldman, Roy D.; Hewitt, Barbara Newlin. The Scholastic Aptitude Test "Explains" Why College Men Major in Science More Often Than College Women. Journal of Counseling Psychology, v23 n1, pp50-54, January 1976.

EJ 132 668 Smith, Deborah Deutsch; Lovitt, Thomas C. The Differential Effects of Reinforcement Contingencies on Arithmetic Performance. Journal of Learning Disabilities, v9 n1, pp21-29, January 1976.

EJ 133 238 Scott, Lloyd F.; Neufeld, Herman. Concrete Instruction in Elementary School Mathematics: Pictorial vs. Manipulative. School Science and Mathematics, v76 n1, pp68-72, January 1976.

EJ 133 398 Schoen, Harold L. Self-Paced Mathematics Instruction: How Effective Has It Been? Arithmetic Teacher, v23 n2, pp90-96, February 1976.

EJ 133 399 Green, Robert W. Analysis of Questions Teachers Ask About Teaching Modern Mathematics in the Elementary School. Arithmetic Teacher, v23, n2, pp98-102, February 1976.

EJ 133 523 Martin, J. Larry. A Test with Selected Topological Properties of Piaget's Hypothesis Concerning the Spatial Representation of the Young Child. Journal for Research in Mathematics Education, v7 n1, pp26-38, January 76.

EJ 133 524 Kidder, F. Richard. Elementary and Middle School Children's Comprehension of Euclidean Transformations. Journal for Research in Mathematics Education, v7 n1, pp40-52, January 1976.

EJ 133 525 Carpenter, Thomas P.; Lewis, Ruth. The Development of the Concept of a Standard Unit of Measure in Young Children. Journal for Research in Mathematics Education, v7 n1, pp53-58, January 1976.

EJ 133 526 Post, Thomas R.; Brennan, Michael L. An Experimental Study of the Effectiveness of a Formal Versus an Informal Presentation of a General Heuristic Process on Problem Solving in Tenth-Grade Geometry. Journal for Research in Mathematics Education, v7 n1, pp59-64, January 1976.

EJ 133 578 Black, Virginia. Logic: They Can Do It in the Fifth Grade. Pre-College Philosophy, v1 n3, pp37-45, July 1975.

EJ 133 633 Cronbach, Lee J.; Webb, Noreen. Between-Class and Within-Class Effects in a Reported Aptitude X Treatment Interaction: Reanalysis of a Study by G. L. Anderson. Journal of Educational Psychology, v67 n6, pp717-724, December 1975.

EJ 133 892 Robitaille, D. F. Classroom Personality Patterns of Teachers of Secondary Mathematics. Alberta Journal of Educational Research, v21 n4, pp249-254, December 1975.

EJ 134 900 Eisenberg, Theodore A. Computational Errors Made by Teachers of Arithmetic: 1930, 1973. Elementary School Journal, v76 n4, pp229-237, January 1976.

EJ 135 013 Linville, William J. Syntax, Vocabulary, and the Verbal Arithmetic Problem. School Science and Mathematics, v76 n2, pp152-158, February 1976.

EJ 135 014 Austin, Joe Dan. Do Comments on Mathematics Homework Affect Student Achievement? School Science and Mathematics, v76 n2, pp159-164, February 1976.

EJ 135 028 Berrill, Renee. Mathematics and Slow Learners. Mathematics in School, v5 n1, pp26-28, January 1976

EJ 135 059 Cooper, Charles R.; Pertosky, Anthony. Secondary School Students' Perceptions of Math Teachers and Math Classes. Mathematics Teacher, v69 n3, pp226-233, March 1976.

EJ 135 125 Lesh, Richard A., Jr.; Johnson, Howard. Models and Applications as Advanced Organizers. Journal for Research in Mathematics Education, v7 n2, pp75-81, March 1976.

EJ 135 126 Lesh, Richard A., Jr. The Influence of an Advanced Organizer on Two Types of Instructional Units About Finite Geometrics. Journal for Research in Mathematics Education, v7 n2, pp82-86, March 1976.

EJ 135 127 Lesh, Richard A., Jr. The Influence of Two Types of Advanced Organizers on an Instructional Unit About Finite Groups. Journal for Research in Mathematics Education, v7 n2, pp87-91, March 1976.

EJ 135 128 Dossey, John A. The Relative Effectiveness of Four Strategies for Teaching Algebraic and Geometric Disjunctive Concepts and for Teaching Inclusive and Exclusive Disjunctive Concepts. Journal for Research in Mathematics Education, v7 n2, pp92-105, March 1976.

EJ 135 129 Knifong, J. Dan; Holtan, Boyd. An Analysis of Children's Written Solutions to Word Problems. Journal for Research in Mathematics Education, v7 n2, pp106-112, March 1976.

EJ 135 130 Allen, Layman E.; Main, Dana B. The Effect of Instructional Gaming on Absenteeism: The First Step. Journal for Research in Mathematics Education, v7 n2, pp113-128, March 1976.

EJ 135 365 Tokar, Edward B.; Stofflet, Frederick. The Validity of the SRA Achievement Series, Multilevel Edition; Reading, Language Arts, and Arithmetic Subtests for Minority and Non-Minority Group Fourth Grade Pupils. Educational and Psychological Measurement, v35 n4, pp1029-1031, Winter 1975.

EJ 135 466 Bean, Andrew G.; Mayerberg, Cathleen Kubiniec. Attitude-Aptitude Relationships in the Quantitative Domain: A Canonical Analysis. Journal of Experimental Education, v44 n1, pp4-8, Fall 1975.

EJ 136 696 Pinker, Aron. Induction and Well Ordering. School Science and Mathematics, v76 n3, pp207-214, March 1976.

EJ 136 702 Fehlen, Joan E. Mastery Learning Techniques in the Traditional Classroom Setting. School Science and Mathematics, v76 n3, pp241-245, March 1976.

EJ 136 852 Dayoub, Iris Mack. An Investigation and Evaluation of Goals of Mathematics Education for Prospective Elementary Teachers. American Mathematical Monthly, v83 n1, pp43-52, January 1976.