DOCUMENT RESUME

ED 309 032 SE 050 657

AUTHOR	Blosser, Patricia E., Ed.; Helgeson, Stanley L., Ed.
TITLE	Investigations in Science Education. Volume 14, Number 1, 1988.
INSTITUTION	Ohio State Univ., Columbus, Ohio. Information Reference Center for Science, Mathematics, and Environmental Education.
PUB DATE	88
NOTE	70p.
AVAILABLE FROM	SMEAC Information Reference Center, The Ohio State University, 1200 Chambers Road, 3rd Floor, Columbus, OH 43212 ($8.00 subscription per year; $2.75 single copy).
PUB TYPE	Collected Works - Serials (022) -- Reports - Research/Technical (143) -- Information Analyses (070)
JOURNAL CIT	Investigations in Science Education; v14 n1 1988
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Cognitive Development; *Cognitive Structures; College Science; Computer Uses in Education; *Experiential Learning; Higher Education; Learning Processes; *Piagetian Theory; *Science Education; *Science Instruction; Teaching Methods
IDENTIFIERS	*Science Education Research

ABSTRACT

The reviews of articles contained in this issue have been grouped into those dealing with research related to cognitive development and to instruction. In addition, seven science education researchers have provided responses to critiques of their published research. Included are: (1) an analysis of an article designed to persuade researchers to abandon Piagetian psychology in favor of the alternative conceptions movement as the psychological framework for science education research (Gilbert and Swift); (2) a report of a study designed to see if the strength of the relationship between student performance and the presence/absence of manipulative models in a testing situation varied with student reasoning level (Staver and Halsted); (3) a study to document students' ideas or conceptual frameworks and to investigate the consistency with which the ideas were used by individuals in different contexts (Clough and Driver); (4) a report which deals with the testing of the ability of several learning theory models to explain college students' performance on drill and practice microcomputer programs in chemistry (Zitzewitz and Berger); (5) a second report describes the investigation and the effects of concept-related instructional organization and locus of control orientation on meaningful learning achievement (Sherris and Kahle); and (6) a study which reports on the relative effectiveness of two hypothesis-testing strategies (disconfirmation, confirmation) using games (Gorman). In the Responses to Critiques section, Yore, Dreyfus, Hill, and Lawrenz provide reactions to articles critiqued in past issues. Responses from Clough and Driver, Staver, and Berger relate to critiques contained within this issue. (CW)

INVESTIGATIONS IN SCIENCE EDUCATION

National Association for Research in Science Teaching

ERIC Clearinghouse for Science, Mathematics, and Environmental Education

INVESTIGATIONS IN SCIENCE EDUCATION

Volume 14, Number 1, 1988

Published Quarterly by

SMEAC Information Reference Center
The Ohio State University
1200 Chambers Road, 3rd Floor
Columbus, Ohio 43212

Subscription Price: $8.00 per year. Single Copy Price: $2.75.
Add $1.00 for Canadian mailings and $3.00 for foreign mailings.

Volume 14, Number 1, 1988

NOTES FROM THE EDITOR:

The reviews of articles contained in this issue have been grouped into those dealing with research related to cognitive development and to instruction. In addition, seven science education researchers have provided responses to critiques of their published research.

In the Cognitive Development section is found an analysis of an article designed to persuade researchers to abandon Piagetian psychology in favor of the alternative conceptions movement as the psychological framework for science education research (Gilbert and Swift); a report of a study designed to see if the strength of the relationship between student performance and the presence/absence of manipulative models in a testing situation varied with student reasoning level (Staver and Halsted); and a study to document students' ideas or conceptual frameworks and to investigate the consistency with which the ideas were used by individuals in different contexts (Clough and Driver).

In the Instruction section, one report deals with the testing of the ability of several learning theory models to explain college students' performance on drill and practice microcomputer programs in chemistry (Zitzewitz and Berger); a second report describes the investigation of the effects of concept-related instructional organization and locus of control orientation on meaningful learning achievement (Sherris and Kahle); and the third study reports on the relative effectiveness of two hypothesis-testing strategies (disconfirmation, confirmation) using games called 2, 4, 6 and Eleusis (Gorman).

In the Responses to Critiques section, Yore, Dreyfus, Hill and Redden, and Lawrenz provide reactions to articles critiqued in past issues of ISE. Responses from Clough and Driver, Staver, and Berger relate to critiques contained within this issue.

Patricia E. Blosser
Editor

Stanley L. Helgeson
Associate Editor

COGNITIVE DEVELOPMENT

Gilbert, John K. and David J. Swift. "Toward a Lakatosian Analysis of the Piagetian and Alternative Conceptions Research Programs." Science Education, 69 (5): 681-696, 1985.

Descriptors--Elementary School Science; Elementary Secondary Education; *Models; *Program Evaluation; Research Design; *Research Methodology; Science Education; Secondary School Science

Expanded abstract and critical analysis prepared especially for I.S.E. by Anton E. Lawson, Arizona State University, Tempe, AZ.

Purpose

The purpose of this article is to persuade researchers to abandon Piagetian psychology in favor of the alternative conceptions movement as the psychological framework for science education research. A philosophy of science developed and expounded by I. Lakatos is used as the basis of the argument.

Rationale

A Lakatosian Perspective -- A Lakatosian research program has three component parts called the Negative Heuristic, the Protective Belt, and the Positive Heuristic. The negative heuristic or "hard core" consists of the basic postulates of the theory assumed to be true by researchers operating within the research program. The abandonment or alteration of these basic assumptions is not possible without abandonment of the program itself. The protective belt consists of a set of auxiliary hypotheses which are compatible with the hard core but can be altered without danger to the hard core. They serve the dual function of operationalizing and protecting the hard core. They operationalize the hard core by allowing for specific predictions to be made and tested and they protect the hard core by deflecting attention away from it in cases where predictions are not validated by empirical test. The formulation and testing of auxiliary hypotheses make up the primary activity of a research program.

Finally, the positive heuristic is a set of suggestions or hints on how to change the auxiliary hypotheses when they fail the empirical test.

Lakatos' notions can be applied to individual research programs and to a comparison of rival programs. The basis for comparison lies in the relative placement of a program on an evaluative continuum between progressive and degenerative. A research program is said to be progressive as long as it keeps predicting novel facts. It is degenerative if its theoretical growth lags behind empirical growth, that is, if it gives only post-hoc explanations of chance discoveries or of facts anticipated by, or discovered in, a rival program. Consequently one of Lakatos' major claims is that he gives rules for the elimination of whole research programs: "If a research programme progressively explains more than a rival it 'supercedes' it, and the rival can be eliminated" (p. 685).

Findings

A Lakatosian View of Piagetian Psychology -- According to the authors, the hard core of Piagetian psychology (PS) consists of the following irrefutable postulates:

1. Living organisms strive for equilibrium states of adaptation: they autoregulate through the process of equilibration which is seen as a process analogous to the evolution of species via genetic assimilation.
2. Intellectual development is analogous to embryological epigenesis (i.e., mental structures are formed in a way similar to embryo formation by the development of a series of novel structures due to successive differentiations - as opposed to performism).

The process of equilibration consists of two complementary processes: assimilation of events to existing mental "schemas" and accommodation, or change of schemas to allow assimilation of initally disequilibrating events. Epigenesis has four essential features. First, the process of development occurs in a casual sequence such that successive steps are dependent upon those preceding. Second, the sequence results in increased organization, differentiation, and complexity. Third, during the process something new emerges: qualitatively different structures appear. And fourth, there is stepwise growth through a series of stages, each stage being marked by qualitatively different emergent structures.

According to Gilbert and Swift, the protective belt - the refutable variants of the program - consists of the increasing sophistication of the four stages of development (sensori-motor, preoperational, concrete operational, and formal operational). They view the program's positive heuristic to be attempts, such as that of Arlin (1975), to develop modifications of the stage theory to make it acceptable to apparently inconsistent results.

A Lakatosian View of the Alternative Conceptions Movement -- In the authors' view the hard core of this movement consists of the following assumptions:

1. The world is real.
2. All observations are theory-laden.
3. Individuals use personally appealing explanatory hypotheses to cope with events in their environment.
4. The individual tests these hypotheses through interaction with reality against personally appealing criteria.
5. Reality provides guidance as to the adequacy of these hypotheses as tested.
6. When hypotheses are judged inadequate by such testing, either the hypotheses or the test criteria are modified or replaced.

The protective belt of the alternative conceptions movement (ACM) is not well developed, yet the authors speculate that there exists "children's science", "teachers' science", and "scientists' science" and that "children's science" can remain virtually unchanged by science teaching. In their view, Hewson's ideas on "conceptual exchange" may have great potential to the movement. Likewise, the positive heuristic of the movement is ill defined, nevertheless the authors suggest that the following statement may provide guidance on how to modify and sophisticate the protective belt: "An individual will test the adequacy of his/her hypotheses against the criteria of prediction and control of events" (p. 690).

A Comparison of the Piagetian and Alternative Conceptions Research Programs -- The authors view the research programs as rivals, yet both concern themselves with the question "Why do students fail to learn the things we want them to learn?" and both are founded on a constructivist philosophy. To introduce differences in the programs, the authors summarize criticisms of the Piagetian program (PS) as follows:

1. There is evidence of very young children showing behavior anticipated in far older people without having demonstrated prerequisite stages of development.
2. Ambiguity exists regarding the timing of stage characteristics. Are specific abilities acquired at the onset, during, or at the end of the stage?
3. Ambiguity exists regarding competence and performance. Evidence suggests that competence and performance diverge for children, yet stage characteristics are supposed to describe general abilities.
4. Methodological fiat renders the stage theory untestable.
5. The concept of equilibration lacks sufficient precision to delineate the adaptations of thinking and fails to adequately explain the effects of training on stage transitions.

Turning their attention to the alternative conceptions movement (ACM) Gilbert and Swift state that this research program is primarily descriptive, devoted to identifying the existence of alternative conceptions. The primary distinction between it and the Piagetian program is that ACM abandons the notion of "stage" as this notion allegedly creates more problems than it solves.

Interpretations

With Lakatos' notion of progress in mind, Gilbert and Swift conclude that the PS shows little or no empirical or theoretical progress. They view the construct of decalage as evidence of this stagnation. Further they view the very appearance of the ACM as evidence of the gradual abandonment of PS by researchers.

ABSTRACTOR'S ANALYSIS

In my view Gilbert and Swift's claim that PS and ACM are in fact rival programs, in the Lakatosian sense, is unfortunate. First, it seems clear to me that ACM is just that - a research movement with an interest in students' alternative conceptions. This movement, as Gilbert and Swift admit, is primarily descriptive - not explanatory - hence fails to constitute a major theory-based research program in the Lakatosian sense with a well defined hard core, protective belt, and positive heuristic. Therefore, it cannot claim to be a bonafide rival to PS. Indeed, the movement can be clearly traced to Piaget's own interest in children's alternative conceptions and to Piaget's basic "constructivist" premise that knowledge is primarily a personal construction. For these reasons I view the authors' use of Lakatosian philosophy of science as a highly convoluted and largely irrelevant excuse to laud ACM research and criticize Piagetian psychology. This, of course, is not to say that aspects of Piagetian psychology are not worthy of criticism. Most assuredly they are, yet this paper fails to do so at anything but the most superficial level.

The fact that Gilbert and Swift view PS and ACM as rival, as opposed to complementary, research perspectives is unfortunate in the sense that their view argues against analysis, modification, and synthesis which I view as desirable from the point of view of instructional theory. Clearly, their remarks fail to discredit the hard core of PS except to highlight the fact that these notions of psychological equilibration and epigenesis lack precision, a fact acknowledged by Piaget himself. One must keep in mind that these psychological constructs were based on analogies drawn from biology by Piaget in the early part of this century. The biological constructs of genetic assimilation and embryological epigenesis are not in question among biologists. Neither is the fact that intellectual development is in some ways analogous to these biological processes. Imprecision is the problem and the ultimate solution requires going well beyond the primitive analogies to develop models based on valid neurological principles. See, for example, an initial attempt to do so entitled "A Neurological Model of Sensory-Motor Problem Solving with Possible Implications for Higher-Order Cognition and Instruction" (Lawson, 1986a).

It should be pointed out that Gilbert and Swift's characterization of epigenesis as resulting in a stepwise growth through a series of stages marked by "qualitatively different emergent structures" is inaccurate. Rather, the concept implies that development proceeds from one stage to another (more complex) stage, due not to qualitatively different emergent structures but due to novel combinations of previous structures such that these novel combinations lead to behaviors with novel emergent properties. Berieter (1985) discusses what he refers to as the "learning paradox" in this context. Berieter asks "How can a structure generate another structure more complex than itself?" (p. 204). As he views the issue, it would not seem possible for a simple self-regulating system to become more complex without some external ladder to climb on, which presumably does not exist - hence the learning paradox.

Lawson and Staver (1987) suggest that a solution to the learning paradox lies in the notion of emergent properties in the natural sciences. Emergent properties are defined as qualitatively unique properties of an object or a system of interacting objects which are derived from a unique combination and/or configuration of the system's component parts. Consider, for example, graphite and diamond, two substances composed only of carbon atoms. In graphite, the carbon atoms are arranged in layers which slide past one another easily, giving graphite a soft, greasy feel and an opaque, black appearance. Diamond, too, is composed of carbon atoms, and can be produced from graphite under conditions of extremely high temperature and pressure. The conversion of graphite to diamond results from a different arrangement of the carbon atoms. In diamond, each carbon atom is bonded to four others in a three-dimensional structure. This three-dimensional array contains no layers and makes diamond a very hard, brittle, crystalline material. Thus, the properties of diamond are emergent from those of graphite. The key point, as far as the present argument is concerned, is that those emergent properties do not arise from novel parts. Rather they arise from a novel arrangement of the same parts.

The natural sciences are replete with examples of emergent properties from embryology and evolution in biology to the collision of gaseous molecules to produce sound waves in physics. The point is that emergent properties exist in many fields in the natural sciences and most likely do in neurosciences as well. See Allman (1986) for an introduction to such work. Behaviors with emergent properties which arise during intellectual development consist first of successful sensory-motor actions in the environment. Second language is acquired which allows the child to acquire knowledge transmitted from adults who speak the same language. A third level begins at the moment the individual begins to ask questions, not of others, but of himself and through the gradual internalization of patterns of argumentation acquires the ability to "talk to himself" which constitutes the essence of hypothetico-deductive thought. Thus, novel behaviors which emerge during development in a stage-wise fashion consist of successful sensory-motor actions, language to name and describe the world of interacting objects and verbally acquire information about that world from others, and finally the internalization of the language to allow the solution of problems in a hypothetico-deductive manner from within oneself (c.f. Piaget, 1976; Lawson, Lawson, and Lawson, 1984; Luria, 1961; Vygotsky, 1962).

It suffices to say that, contrary to Gilbert and Swift's assessment, I believe psychological stages, so defined, are real. This fact, however, in no way diminishes the importance of students' alternative conceptions. Rather, it serves to enrich our understanding of intellectual development by forcing us to consider the relationship between these general stages of development and children's specific conceptions. It is here that Piaget's theory provides little guidance, yet it is here that I believe the primary instructional payoff lies.

What, if any, relationship exists between students' misconceptions and stages of intellectual development? Parmenedes, the ancient Greek philosopher stated that "the senses deceive us." He believed that one can only reach truth through reason. Personal experience provides the basis for knowledge that is at times inaccurate (e.g., optical illusions, Piagetian nonconversation responses). Leading naturalists of the past advocated ideas such as spontaneous generation, special creation, and the inheritance of acquired characteristics. These ideas have their roots in personal experience. Maggots appear to be spontaneously generated from rotting flesh; people create objects so living objects must also be created by "people" (with special God-like properties); children look like their parents so changes in the appearance of the parent will cause a change in the appearance of an as-yet-to-be-born child.

The rejection of these ideas during the past required the generation of alternative hypotheses and their testing through logical reasoning (as Parmenedes had claimed), experimentation, data collection, and considerable argumentation. Open minded scientists who became aware of these alternative ideas (e.g., evolution, natural selection, genetics), the available evidence, and were able to follow the lines of reasoning used to argue the cases, were generally convinced and were able to overcome prior "misconceptions" in favor of the more scientifically accurate conceptions.

By analogy, it can be hypothesized that the same thing happens in the science classroom. For students to overcome their misconceptions they must become aware of the scientific conceptions, as well as their own alternative conception(s), and they must become aware of the evidence <u>and</u> reasoning which bears on the validity of the alternative conceptions. In other words, they must be able to logically "see" how the evidence supports the scientific conceptions and contradicts the naive misconception (cf., Posner, Strike, Hewson and Gertzog, 1982). Logically "seeing" this requires the use of formal operational reasoning patterns.

Because formal reasoning patterns are precisely those used to evaluate alternative conceptions in a logical hypothetico-deductive manner (e.g., combinational reasoning, control of variables, probabilistic and correlational reasoning), concrete operational students who lack these skills would be expected to hold more misconceptions than their formal operational peers. A recent study by Lawson and Thompson (1987) found this to be precisely the case. On a test following instruction on concepts of evolution and genetics, a sample of concrete operational seventh grade students revealed an average of 1.67 misconceptions per students while their formal operational classmates held only 0.43 misconceptions per student. The conclusion is simply this: Students who lack formal reasoning skills hold more misconceptions than their formal peers because formal reasoning skills are necessary to overcome prior misconceptions.

The preceding discussion leads to the possibility that PS and ACM need not be seen as rival programs but rather can be synthesized into a coherent framework for the articulation of a theory of instruction. Clearly space does not permit a detailed discussion. Rather, I will simply list possible postulates of such a theory below. The interested reader may refer to Lawson (1986b) for details.

1. Students often hold misconceptions, i.e., knowledge derived from extensive personal experience which is incompatible with established scientific theory.
2. Misconceptions may be deeply-rooted, instruction-resistant impediments to the acquisition of scientifically valid conceptions.
3. The overthrow of misconceptions requires students to move through a phase in which a mismatch exists between the misconception and the scientific conception and provokes a "cognitive conflict" or state of mental "disequilibrium".
4. The improvement of reasoning skills arises from situations in which students are engaged in exchanges of contradictory conceptions where arguments are advanced and evidence is sought to resolve the contradiction.
5. Argumentation provides verbal experiences from which particular forms of argumentation (i.e., patterns of

reasoning) may be abstracted (i.e., internalized).

6. The learning cycle is a method of instruction which consists of three phases called exploration, term introduction, and concept application.
7. Use of the learning cycle provides the opportunity for students to reveal prior conceptions/misconceptions and the opportunity to argue and test them, thus become "disequilibrated" and develop more adequate conceptions and reasoning patterns.
8. There are three types of learning cycles (descriptive, empirical-inductive, hypothetical-deductive) which are not equally effective at producing disequilibrium and improved reasoning.
9. The essential difference among the three types of learning cycles is the degree to which students either gather data in a purely descriptive fashion or initially set out to explicitly test alternative conceptions (hypotheses).
10. Descriptive learning cycles are designed to have students observe a small part of the world, discover a pattern, name it, and seek the pattern elsewhere. Normally only concrete operational reasoning is required and little or no disequilibrium occurs.
11. Empirical-inductive learning cycles require students to describe <u>and</u> explain a phenomenon, thus allowing for misconceptions, argumentation, disequilibrium, and the development of formal reasoning patterns.
12. Hypothetical-deductive learning cycles require the immediate and explicit statement of alternative conceptions/hypotheses to explain a phenomenon and require formal reasoning patterns in the test of the alternatives.

The result of instruction based upon the previous theoretical postulates is essentially two-fold. First, students will leave the classroom with conceptions of nature more compatible with those of the practicing scientists and, second, the students will have possession of reasoning skills (mediated by the use of internalized language) that will allow the evaluation of alternative conceptions of nature that may arise in the future and will allow them to make decisions based upon sound reasoning and evidence as opposed to potentially misleading appearances.

REFERENCES

Allman, W. F. "Mind Works." Science 86 (15) 23-31, 1986.

Bereiter, C. "Toward a Solution of the Learning Paradox." Review of Educational Research, 55 (2): 201-226, 1985.

Lawson, A. E. "A Neurological Model of Sensory Motor Problem Solving with Possible Implications for Higher-Order Cognition and Instruction." Journal of Research in Science Teaching, 23 (6) 503-522, 1986a.

Lawson, A. E. "Integrating Research on Misconceptions, Reasoning Patterns and Three Types of Learning Cycles." Paper presented to the US-Japan Seminar on Science Education, Honolulu, Hawaii, September, 1986b.

Lawson, A. E., D. I. Lawson and C. A. Lawson. "Proportional Reasoning and the Linguistic Abilities Required for Hypothetico-Deductive Reasoning." Journal of Research in Science Teaching, 21 (2), 119-131, 1984.

Lawson, A. E. and J. R. Staver. "Solving the Learning Paradox: Emergent Properties and a Neurological Model of Constructivism." Unpublished paper, Arizona State University, 1987.

Lawson, A. E. and L. Thompson. "Relationships among Biological Misconceptions, Reasoning Ability, Mental Capacity, Verbal I.Q. and Cognitive Style." Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, April 1987.

Luria, A. R. The Role of Speech in the Regulation of Normal and Abnormal Behavior. J. Tizard (Ed.). Oxford: Pergamon, 1961.

Piaget, J. The Grasp of Consciousness. Cambridge: Harvard University Press, 1976.

Posner, G. J., K. A. Strike, P. W. Hewson and W. A. Gertzog. Accommodation of a Scientific Conception: Toward a Theory of Conceptual Change." Science Education, 66 (2), 211-227, 1982.

Vygotsky, L. S. Thought and Language. Cambridge, MA: The M.I.T. Press, 1962.

Staver, J. R. and D. A. Halsted. "The Effect of Reasoning on Student Performance on Different Sections of a Posttest." Science Education, 68 (2): 169-177, 1984

Descriptors-*Academic Achievement; *Chemistry; Cognitive Development; *Cognitive Processes; *Developmental Stages; High Schools; Pretests Posttests; Science Education; Science Instruction; *Science Tests; *Secondary School Science; Testing

Expanded abstract and analysis prepared especially for I.S.E. by Geoffrey J. Giddings and Barry J. Fraser, Curtin University of Technology, Perth, Western Australia.

Purpose

The authors state their purpose in terms of the following question: "When sections of a posttest differ in the use of manipulative models by students to answer questions, are differences in performance attributable to differences in the reasoning levels of the students?" We interpret this to be a question about whether the strength of the relationship between student performance and the presence/absence of manipulative models in the testing situation varies with student reasoning level.

Rationale

The study arises out of research relating the Piagetian level of students' thinking patterns with the reasoning demands of the instruction. The study is centered around the teaching of chemical bonding (in particular, molecular geometry and shape) at the high school level. The authors suggest that previous research relating the effects of instructional strategies and students' reasoning level on achievement has largely ignored the reasoning demands of the test settings themselves. Thus, the study attempted to provide us with evidence about the differential effect of students' reasoning levels on achievement when differing evaluative formats were used on posttest items.

An important assumption in the study is that students who use predominantly concrete reasoning patterns are able to make direct inferences from their own observations, but are unable to make inferences "twice removed" from the observations because this requires a formal thought pattern.

Research Design and Procedure

This study was carried out in a large upper to middle class suburban public high school. Instruction in chemical bonding for the 105 students in the sample occupied a total of four weeks, with a one-week segment on molecular geometry and shape being the focus for this study. Students remained in their own classes (four in all) throughout the investigation, with each class receiving similar instruction from the same instructor. All students took tests on chemical bonding including a special posttest on molecular geometry and shape.

The Piagetian logical operations test (PLOT) (Staver and Gabel, 1979) was administered to all students prior to the four-week instructional sequence on chemical bonding. PLOT is a group-administered test of formal thought which enabled the authors to assign students into groups broadly labelled "concrete thinkers" and "formal thinkers." In separate analyses, PLOT scores were used: (1) as a continuous variable and (2) as a dichotomous variable (concrete/formal).

Understanding this study requires the reader to identify clearly the teaching strategies used in the instructional sequence on molecular geometry and shape. The most significant point in relation to the instructional strategy is that three-dimensional molecular models were used <u>extensively</u> throughout the instructional sequence by both instructor and students. Specifically, in a hands-on laboratory session, the students assembled a "type of ball and stick model" of four basic molecules. Extensive teacher-student discussion occurred both during this laboratory session and during a post-laboratory discussion, which was directed at relating electron dot and dash representations to the actual molecular shapes that had been constructed.

Immediately following the segment of instruction on molecular geometry and shape, all students were given a two-part stereochemistry quiz as a posttest. It is important to note that the test had two distinct sections. In section one, students were permitted to construct and manipulate molecular models to help answer the questions whereas, in section two, students answered without recourse to any molecular models.

The first section of the posttest required students to draw electron dot and dash structures for the same four molecules which the students had previously examined in their laboratory model-manipulation exercise. Students were also required to construct models of the same four molecules, using a less fixed molecular model system (straight pin and polystyrene sphere). Students were asked, too, to determine the formula and shape of each molecule. A total of 28 points was awarded for the various segments of this first section.

The second section required the students to write an appropriate electron dot and dash structure for a more complicated molecule not previously encountered by the class. No model materials were available to the students for this section of the posttest. A single question worth a total of only two marks comprised section two of the posttest.

Findings

The authors report student mean performance on the two sections of the stereochemistry quiz (the posttest) as well as their PLOT scores. Bivariate regression analyses, using PLOT scores as a continuous independent variable in the first analyses and as a dichotomous independent variable in the second analyses, were performed separately for each of the two sections of the stereochemistry quiz as dependent variables.

Students labelled as formal thinkers performed better on both the first section of the posttest (model section) and the second (no-models section) than did students identified as concrete thinkers. The relationship between performance and student reasoning level was only marginally stronger for the no-models section. The authors interpret these results as suggesting that the reasoning level of students could affect their performance on sections of a test differing in reasoning demands (under conditions of constant instruction).

Interpretations

The authors conclude that, in the same instructional environment, the effects of students' reasoning levels are evident in performance on tests that differ in the extent to which models can or cannot be used. Thus achievement might be affected by the actual format of a test. The authors conclude that teachers and instructors should carefully analyze their tests for the reasoning demands of the questions and should compare and/cr adjust these to the reasoning levels of the students.

ABSTRACTORS' ANALYSIS

The authors are to be commended for delineating a question - namely, whether test performance depends differentially on test format (in terms of presence or absence of manipulative models) depending on student level of (Piagetian) reasoning ability - which is interesting, educationally important and which has the potential of providing fruitful insights in science education. Unfortunately, however, the study's design limitations do not permit an unconfounded test of this important question and the study's results, contrary to claims made by the authors, don't provide convincing evidence to support the presence of this differential effect.

There are some sources of confounding in the study which limit the possibility of obtaining credible conclusions. First, as the authors themselves note, the two sections of the test differ not only in terms of the presence/absence of molecular models, but also possibly in the cognitive requirements of the test items. Thus, it is not possible to know whether any observed differences in performance should be attributed to the cognitive demand of the test items or to the presence/absence of molecular models. Second, because the section of the test not involving models contained only a single two-mark item, limited confidence can be placed in the dependent variable.

The main data analysis involved two bivariate regressions with student reasoning level as independent variable. Whereas the dependent variable for the first analysis consisted of scores on the section of the test involving models, the second analysis involved as dependent variable scores on the section of the test not involving models. This pair of regression analyses is reported, first, with reasoning (PLOT) scores retained as a continuous variable and, second, with PLOT scores dichotomized in order to classify students into either "formal" or "concrete" thinkers. Whereas each analysis yields the strength of the relationship between performance and reasoning ability, a comparison of the relative strengths of the two relationships provides information relevant to the study's main question about the differential effect of test format on performance for students of various reasoning levels.

The reporting of the regression analyses is unnecessarily complex given that only two variables (performance and reasoning level) are involved in any analysis. Because the regression analysis is equivalent to a simple correlation analysis for this case of two variables, the reader could have been presented with the simple findings that the correlation between performance and reasoning level was 0.26 for the models section and 0.25 for the no-models section, using continuous PLOT reasoning scores, and was 0.20 for the models section and 0.25 for the no-models section using dichotomous reasoning scores.

Although not large, all of these correlations represent a statistically significant association between reasoning level and performance on both the models and the no-models tests. This result is interesting in its own right, but is nothing new as past research often has established links between achievement and Piagetian level.

Moreover, these results are not relevant to a key question in this study: namely, whether the effect of the use/non-use of models in tests on performance depends on student reasoning level. In fact, the study's design does not permit an effective test of this hypothesis, which would involve the detection of a (reasoning level) x (presence/absence of models) interaction in a factorial design.

The analysis involving dichotomous PLOT scores (formal/concrete) revealed a correlation of 0.20 for the models test and of 0.25 for the no-models group. The authors tend to interpret this small difference as suggesting the existence of a differential relationship between test performance and student reasoning ability depending on the presence/absence of models. Given the various limitations of the analysis (e.g., the arbitrariness of the choice of marks for defining formal/concrete; the loss of statistical power associated with dichotomizing a continuous variable; and the problem that the no-models test had only a single two-mark item), it is dangerous to interpret this small difference between the correlations of 0.20 and 0.25 as educationally significant. Furthermore, for the more dependable analyses involving PLOT scores as a continuous variable, the correlations had virtually identical values of 0.26 for the models test and 0.25 for the no-models test.

A related problem is the implicit assumption by the authors that concrete reasoning requires the presence of objects (in this case the molecular models) and that formal thinking is required to operate without the models (as in the second section of the posttest). These assumptions do not appear to fit entirely with the original literature. For example, in Piaget and Inhelder (1967) we find:

> "Let us begin by removing a possible source of misunderstanding. Once they have evolved fully, all operations...can be performed abstractly." (p. 457)

This statement suggests that even the concrete operational structures, once formed, can be carried out <u>without</u> objects being present - even twice removed! Thus, a more appropriate definition of concrete level of reasoning could be that the reasoning level cannot go <u>beyond</u> what has already been done with objects.

One clear and useful result of the study is that the combination of teacher demonstrations with molecular models during instruction, and with student manipulation of models during instruction and posttesting, resulted in satisfactory achievement by virtually all the students in the sample. However, the interpretation by the authors that "formal thinkers are more successful" is hardly proven by this study. The relatively small differences found between formal and concrete students, together with a number of design weaknesses, make such a conclusion unwise.

Overall, the research forms a good basis as a pilot study of an important question in science education. What is needed now is a comprehensive and more detailed examination of the relationship between reasoning demands of test settings and the reasoning levels of the students than is provided in this article.

REFERENCES

Piaget, J. and B. Inhelder. The Child's Conception of Space. New York: Norton, 1967.

Staver, J. R. and D. L. Gabel. "The Development and Construct Validation of a Group-Administered Test of Formal Thought." Journal of Research in Science Teaching, 16: 534-544, 1979.

Clough, Elizabeth and Rosalind Driver. "A Study of Consistency in the Use of Students' Conceptual Frameworks Across Different Task Contexts." Science Education, 70 (4): 473-496, 1986.

Descriptors--Biology; *Cognitive Development; *Concept Formation; Evolution; Heat; Learning Theories; Physical Sciences; Pressure Physics; Science Education; *Science Instruction; Secondary Education; *Secondary School Science

Expanded abstract and analysis prepared especially for I.S.E. by Anton E. Lawson, Arizona State University.

Purpose

This study explored the issue of the consistency of use of students' conceptions across a number of different tasks which probed understanding of aspects of pressure, heat, and biological evolution. The purpose was not only to document students' ideas or conceptual frameworks, but to investigate the consistency with which the ideas were used by individuals in different contexts. Specific questions raised were:

1. Was there any commonality in students' conceptions and, if so, what frameworks were identified?
2. What was the prevalence of these frameworks among students of different ages?
3. To what extent were the same frameworks used in different contexts by the sample as a whole?
4. To what extent were the same frameworks used in different contexts by individual students?

Rationale

A number of recent studies have shown that individuals construct their own explanations for phenomena they encounter in their lives. Many of these personal theories are inconsistent with, even contradictory to, modern scientific explanations. Various authors have referred to these constructions as informal theories, naive theories, alternative frameworks, naive principles, or intuitive ideas. In most cases, the assumption has been made that personal theories are stable, they are shared by many students, and they are used to interpret a wide variety of phenomena, such that it would be worth planning instructional sequences to change them.

A number of authors have pointed out the similarity of students' personal theories to scientific theories of the past. This similarity is seen as evidence of the experiential basis and pervasiveness of personal theories. On the other hand, other researchers have been more skeptical about the stability or generality of these theories, and therefore have not advocated widespread instructional interventions to deal with them. Hence, the central issue of this study was to determine the consistency with which students used their own personal theories to explain a variety of phenomena that varied in external features but involved the same basic scientific principles. The basic prediction tested by Clough and Driver was that students would use the same theory across contexts that are different, but that scientists construe similarly. The alternative prediction is, of course, that they would not because context cues, which vary, and the lack of any profound personal commitment to a theory would instead lead students to change their explanations from context to context.

Research Design and Procedure

Sample. The sample consisted of 30 12-year-olds, 30 14-year-olds, and 24 16-year-olds from three schools in one urban area. A spread of ability was obtained by selecting a stratified random sample of students based upon responses to a short relational reasoning test.

Tasks. Each student was individually interviewed in two sessions lasting 45 minutes each. At least two tasks, which involved presentation of some phenomenon, or photographs of the phenomenon, were administered in six areas:

1. Changes in pressure with depth.
2. Pressure in different directions.
3. Movement of fluids from regions of high to low pressure.
4. Heat conduction in solids.
5. Heritability of acquired characteristics.
6. Natural selection.

Each student's conception of the phenomenon and explanation were probed. Interviews were audiotaped and transcribed.

Data analysis. Data were analyzed in two stages. In the first stage, responses were grouped into mutually exclusive categories based upon type of explanation offered. When a response contained multiple confused explanations, it was designated "uncodeable." In the second stage, categories of responses that occurred across parallel tasks were identified. These were designated as "frameworks." Frameworks were identified across question contexts for all groups of questions. Contingency tables and coefficients were then used to determine the consistency of student explanations across tasks.

Findings

Frameworks were identified and listed for each area investigated. For example, the following frameworks were identified for the tasks related to heat conductivity:

1. Heat energy travels through different materials at different rates.
2. Metal attracts/conducts coldness.
3. Conductivities depend on some observable property of material.
4. Metals let heat in and out more easily.

In general, the results indicated a relatively small change toward more scientific understandings among the older students in the sample.

The proportion of students using each framework was similar in the different task contexts relating to a given idea. However, consistency of the use of frameworks by individuals was not always as high as expected. It was more likely to occur for phenomena for which there were a limited number of frameworks and where these were congruent with intuition. There was consistency of use of both the scientifically accepted frameworks and the students' alternative frameworks, although the degree of consistency appeared to depend on contextual aspects of the tasks.

Contingency coefficients, which reflected the overall level of consistency for the six task areas mentioned previously, were: .52, .53, .54, .69, .38, and .72, respectively.

Interpretations

This study indicates that students do have alternative frameworks (personal theories) in all areas investigated, and that the prevalence of these is reasonably predictable across different contexts, but that the frameworks are used less consistently at the individual level. The consistency with which the ideas are used also appears to depend on the topic area. This would need to be taken into account in curriculum planning. Science curricula could be developed in such a way as to pay explicit attention to these various ideas, to provide other opportunities for students to make their ideas explicit so they are open to inspection, and to provide counter examples to stimulate conceptual change. There is evidence that some alternative conceptions are resistant to instruction even when the teaching has been deliberately structured to incorporate or contrast students' ideas. Detailed classroom-based research is needed to devise appropriate teaching strategies to deal with this problem.

ABSTRACTOR'S ANALYSIS

My comments will center on four general issues:

1. The generality of conceptual frameworks and thinking skills.
2. Sources of alternative frameworks.
3. Factors which influence conceptual change.
4. Appropriate teaching methods.

Generality of conceptual frameworks and thinking skills. In their introduction, Clough and Driver raise the issue of the influence of context-dependent task cues on student responses by calling attention to such influences on Piagetian tasks of formal operational reasoning. In short, their argument was that general "formal" thinking skills do not exist because task performance varies too much from task to task (i.e., intertask correlations are too low to indicate the existence of general thinking skills). It was very interesting to note that Clough and Driver's general conclusion in the present study was that alternative frameworks do exist. While I certainly have no quarrel with this general conclusion, it should be noted that the empirical basis for their conclusion rested primarily on contingency coefficients, which ranged from .72 to .38 (x = .56). Clough and Driver's willingness to draw the conclusion that general conceptual frameworks exist based upon coefficients which can, at best, be described as moderate, must be contrasted to their apparent unwillingness to accept the conclusion that general thinking skills (i.e., Piagetian formal operational schemata) exist based upon coefficients of equal or large magnitude when responses to various Piagetian formal tasks are intercorrelated.

A substantial number of studies reviewed by Lawson (1985) have found coefficients larger than those reported by Clough and Driver, even when Piagetian tasks involving different contexts and different reasoning patterns were intercorrelated. The appropriate conclusion to draw is that both conceptual frameworks of some general applicability and general thinking skills exist. The interesting issues that arise from the acceptance of this position involve the interactions among the two types of knowledge, e.g., How does the presence or absence of thinking skills influence the nature of and/or modifiability of conceptual frameworks? And how does the presence or absence of conceptual frameworks influence the nature and/or modifiability of thinking skills? Future research should attempt to answer these questions.

Sources of alternative frameworks. Although it seems safe to conclude that alternative conceptual frameworks do exist, there appears to be a considerable variation among them with regard to the extent to which they are used and their stability. Their consistency of use and stability may in large part be a function of their source and their relationship (or lack of relationship) with other conceptual systems. Many students, for example, believe in the Christian account of special creation. This alternative framework is often intimately linked to a whole host of religious ideas which are all pervasive in the life of the individual. Here the source is societal and the framework is all pervasive and extremely stable (for many). On the other hand, the idea of suction has its origin in personal experience and is far less pervasive. The scientific views of air pressure is difficult to teach, but one would certainly not have to battle religious doctrine to do so. Still other personal explanations may need only a quick remark to modify. For example, if a student theorizes that water rises in trees due to little heart-like pumps in their roots, a comment to the contrary by a science teacher may be all that is needed for the student to discard the idea.

The point is simply this. Although a variety of frameworks no doubt exist, care must be exercised not to jump on a bandwagon and assume that all such frameworks are of equal relevance. Alternative views arise from personal experience, from society in general, from specific peers, from poor instruction, etc. The main research issue here should not be to simply identify alternative frameworks but to identify the important ones (defined as ones which are serious impediments to science literacy).

Factors which influence conceptual change. Of course, the degree to which personal theories are linked to other concepts and conceptual systems will play a large part in the ease with which they can be modified and/or discarded through instruction. A recent study by Lawson and Weser (1989), for example, tested the prediction that one factor which influences the modifiability of the alternative biological concepts of special creation, teleology, orthogenesis, the soul, reductionism and vitalism is students' hypothetico-deductive thinking skills. This prediction was based upon the hypothesis that in order to change one's view, the view must be contrasted with the alternatives and data must be gathered to compare and contrast with deductions from one's view and those alternatives. The "logical" comparison of alternatives with their deduced consequences and the evidence allows one to decide which of the views should be rejected and which should be accepted. If students do not possess the necessary thinking skills to assimilate arguments of this nature, then they are not expected to change their minds.

In a sample of nearly 1,000 college students enrolled in a one-semester biology course, Lawson and Weser found that the better thinkers did indeed move more away from the nonscientific conceptions, provided the conceptions were not tied too closely to emotionally laden religious beliefs. For instance, when the issue of special creation/evolution was considered, the better thinkers did not change any more than the poor thinkers. The reason for this was not clear. There is, no doubt, more than one reason to change one's mind. Some peope, for example, may change their minds not because they understand the reasons for one idea and against another, but because they are impressed by the authoritative position of the person advocating the view. It is possible that the evidence and arguments presented against special creation and for evolution were not sufficiently well presented to convince the good thinkers, but that the instructor's status was sufficient to sway some of the poor thinkers. A number of studies have shown that poor thinking skills and field dependence are linked, and it seems likely that field dependent people would be more likely to be impressed by a person's stature than field independent people. Clearly much research remains to be done to clarify the nature of conceptual change and the influence of factors such as these.

Teaching methods. In my view, the alternative conceptions research tradition, within which Clough and Driver's paper clearly belongs, represents an extremely worthwhile and potentially productive tradition in terms of its implications for teaching methodology. In short, the research task is to identify pervasive student personal theories, explicate the manner in which they are constructed and the way they can be modified. Instruction can then be designed in such a way to accomplish this end. Clearly the work of authors such as Hewson and Hewson (1984), Champagne, Klopfer and Gunstone (1982), and Anderson and Smith (1986) and others is relevant.

Since conceptual chang was very much the aim of the earlier curriculum developement efforts of the Science Curriculum Improvement Study (e.g., SCIS, 1973), it should come as no surprise that the SCIS program embraced a teaching methodology which encourages conceptual change to take place. That teaching methodology is called the learning cycle and it consists of three phases called exploration, invention, and discovery. During exploration, students encounter new phenomena which often provokes them to use and verbalize personal theories. During the invention and discovery phases of the learning cycle, the various theories put forth by the students and/or the teacher are then discussed and tested through actual experiments. Examination of the views of others, the data and the relevant arguments allow for conceptual change. Another very important result occurs as well. Students become more conscious of and skilled in the general process of theory construction and test. In other words, they become better thinkers.

In my view, the learning cycle represents a flexible model for science instruction which is appropriate for students of all ages in all areas of science. Further, it allows for a synthesis of research into students' alternative frameworks with that into students' thinking skills. A much more detailed discussion of this synthesis can be found in Lawson, Abraham and Renner (1989).

REFERENCES

Anderson, C. W. and E. R. Smith. "Teaching Science." In V. Koehler [Ed.], The Educator's Handbook: A Research Perspective. New York: Longman, Inc., 1986.

Champagne, A. B., L. E. Klopfer, and R. F. Gunstone. "Cognitive Research and the Design of Science Instruction." Educational Psychologist, 17: 31-53, 1982.

Hewson, P. W. and M. G. A. Hewson. "The Role of Conceptual Conflict in Conceptual Change and the Design of Science Instruction." Instructional Science, 13: 1-13, 1984.

Lawson, A. E. "A Review of Research on Formal Reasoning and Science Teaching." Journal of Research in Science Teaching, 22 (7): 569-618, 1985.

Lawson, A. E., M. R. Abraham, and J. W. Renner. A Theory of Instruction: Using the Learning Cycle to Teach Science Concepts and Thinking Skills. Cincinnati: National Association for Research in Science Teaching, 1989.

Lawson, A. E. and J. Weser. "The Rejection of Nonscientific Beliefs as a Function of Instruction and Reasoning Ability." Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco, April, 1989.

Science Curriculum Improvement Study. SCIS Omnibus. Berkeley, California: Lawrence Hall of Science, 1973.

INSTRUCTION

Zitzewitz, B. S. and C. Berger. "Applications of Mathematical Learning Models To Student Performance on General Chemistry: Microcomputer Drill and Practice Programs." Journal of Research in Science Teaching, 22: 775-791, 1985.

Descriptors--*Academic Achievement; *Chemistry; *College Science; *Computer Software; Higher Education; *Mathematical Models; Microcomputers; Science Education; Science Instruction

Expanded abstract and analysis prepared especially for I.S.E. by Ann C. Howe, North Carolina State University.

Purpose

The first purpose of this study was to test the ability of several learning theory models to explain college students' performance on four drill and practice microcomputer programs in chemistry. A second purpose was to suggest a method of evaluating microcomputer drill and practice programs and to show one way that the microcomputer can be used as a research tool in science education.

Rationale

The rapid increase in the use of microcomputers for instructional purposes has not been accompanied by a corresponding increase in our understanding of how, or whether, learning takes place. This study tested learning outcomes of the use of four computer programs against mathematical equations that had been proposed as models of learning under similar conditions. Thus, the study built on previous work in seeking a model for predicting outcomes of the use of microcomputers in science education.

Research Design and Procedure

A hypothetical case, using data from three subjects, was employed to illustrate the method of analysis, which will not be familiar to some readers. In this method, curvilinear regression analysis, using a polynomial regression equation, is used to analyze group learning data (Kerlinger and Pedhazur, 1973). This analysis is carried out through a series of steps, testing at each step to determine whether a higher order polynomial adds significantly to the variance of the dependent variable accounted for. In this case the dependent variable is the score (in percentage points) on a learning task and the independent variable is time on task. By a rather involved procedure, the quartic (fourth degree) equation was shown to give the best fit to the hypothetical data and could be used, according to the authors' interpretation, to predict the amount of learning achieved <u>before</u> the start of the program. Before turning to actual data collection and analysis, the authors also explain the method used for determining whether an outlying data point could be discarded in the analysis of data collected.

Data were collected from 54 students in a first-semester general chemistry course who volunteered to participate. Equivalent groups of subjects were formed by random selection to use one of four selected drill and practice microcomputer programs. The programs were used after a lecture in place of a recitation assignment. Subjects' scores on the items in the programs and cumulative time on task, recorded on the computer, formed the data base for the analyses.

Data for each of the parts of the programs (with a few exceptions) were analyzed separately, using the technique described above. In addition to the polynomical regression equation (model) used for the hypothetical case, two other mathematical models were tested for goodness of fit to the actual data. These models were the exponential saturation learning curves of Bush and Mosteller, (1951) and of Aldridge (1983).

Findings

Results of the analyses were complex and, in some cases, ambiguous, but the general conclusion was that the best fit to data was given by the linear model. It was also found that the percentage of variance explained varied greatly among the microcomputer programs and parts of programs, from 75% for one section of one program to 2% for another section of the same program. In 65% of the cases neither the Bush and Mosteller model nor the Alridge model gave significantly better fits to the data than the linear equation.

Interpretations

The investigators drew two major conclusions from the study. The first was that learning from computer drill and practice program appears to be linear; that is, when a student works on a problem based on a small number of algorithms, the percentage of problems answered correctly increases linearly with time until a ceiling is reached. Students learn at different rates, and different programs teach at different rates.

This information should be useful in evaluating programs, since the rate of learning is an important variable in any instructional program.

The second conclusion is that the use of microcomputers to study [sic] and gather data is a valuable technique in research on the use of microcomputers.

ABSTRACTOR'S ANALYSIS

This study has done us a service by reminding us that models are as necessary as instruments and experiments. It also raises many interesting questions. Can higher order learning be modeled by a mathematical equation? Does learning from a computer follow the same pattern as other learning, if indeed other learning follows a predictable pattern? How can the microcomputer be used as a research tool in science education?

In spite of our obvious need to know more about how and what people learn from computer programs, one has to question whether the current state of the art allows us to further our understanding by means of the analytical techniques used in this study. Are such complicated mathematical procedures, now made relatively easy by the computer, necessary or even useful when we have the kind of data collected for this study? One is reminded of a comment by Quinn McNemar, formerly a professor of psychology, statistics and education at Stanford. Referring to the use of cubic, quartic, quintic, etc. components he wrote, "Since these polynomial forms of relationship are scarce in the empirical data of psychology and are even more scarce in the minds of psychological theorists, there would seem to be no good reason for going beyond the second degree polynomial (quadratic) ..." (McNemar, 1962, p: 361). One suspects that these polynomial forms are also scarce in the minds of science educators and that most readers are relieved to be told that a linear equation fits the data as well as, or better than, the higher order equations.

Leaving aside the non-mathematician's reluctance to confront curvilinear regression analysis, the question of data overkill remains. It is not clear that the same care was exercised in planning the procedures that provided the data as in analyzing the data after they were collected. More details about the computer programs and the way they were used would clear up some of the questions and help the reader form a judgment about the significance of the results. It appears, however, that there were so many differences in the programs, in pupils' approaches to test taking, and in other aspects of the procedures that it is not surprising that no strong generalizations can be made and it seems premature to dismiss the Bush and Mosteller or the Aldridge models.

The need to study the educational uses of microcomputers is urgent. The advance of computer technology is outstripping our ability to harness it or adapt it as a means of improving teaching and learning. For this reason one welcomes any attempt to further our understanding of instructional uses of computers. Science educators, particularly, should certainly welcome the attempt to develop or test models of learning, since our field has been notably atheoretical in its approach to most problems. Other researchers are encouraged to continue this line of work, taking care to refine their instruments and experimental designs as they do so.

REFERENCES

Aldridge, B. G. "A Mathematical Model for Mastery Learning." Journal of Research in Science Teaching, 20: 1-17, 1983.

Bush, R. R. and F. Mosteller. "A Mathematical Model for Simple Learning." Psychology Review, 58, 313-323, 1951.

McNemar, Q. Psychological Statistics. New York: John Wiley Publishers, 1962.

Kerlinger, F. and E. J. Pedhazur. Multiple Regression in Behavioral Research. New York: Holt, Rinehart and Winston, 1973.

Sherris, J. D. and J. B. Kahle. "The Effects of Instructional Organization and Locus of Control Orientation on Meaningful Learning in High School Biology Students." Journal of Research in Science Teaching, 21 (1): 83-94, 1984.
Descriptors--Biology; *Concept Formation; *Concept Teaching; High Schools; *Instructional Design; Learning Theories; *Locus of Control; Science Education; *Science Instruction; *Secondary School Science; Sex Differences

Expanded abstract and analysis prepared especially for I.S.E. by Lynn Dierking, University of Maryland.

Purpose

The purpose of this article was to describe a study which investigated the effects of concept-related instructional organization and locus of control orientation on meaningful learning achievement. Specifically, researchers attempted to determine the effects of conceptual cues and concept mapping experience on subsequent meaningful learning. Researchers also focused on the interactions between this instructional treatment and students' locus of control orientation.

Rationale

The authors reviewed literature on instructional organization and locus of control orientation, concluding that there was a need to broaden the knowledge base regarding instructional techniques which enhance meaningful learning and in particular, to investigate the relationship between locus of control orientation and achievement. They also felt that there was a lack of meaningful learning-related studies that have investigated the effects of interaction between instructional treatments and individual attributes.

Research Design and Procedure

The design of the study was quasi-experimental ("reversed treatment non-equivalent control group design with pretest and post-test"; Cook and Campbell, 1979); a 2 x 2 factorial design with treatment type as one independent variable, locus of control orientation as the second independent variable, and achievement test scores as the dependent variable. There were 541 ninth and tenth grade subjects; 282 in the experimental group (13 classes) and 259 in the comparison group (12 classes). One-way ANOVA analyses by treatment group of Otis Lennon Mental Ability Test (OLMAT) scores and of pretest scores indicated that the two groups were equivalent on the OLMAT variable but not on the pretest variable. When treatment groups were used as an independent variable, ANCOVAs with pretest scores as the covariate were calculated to account for nonequivalence on the pretest variable (homogeneity of regression requirements evaluated by Cochran's C test were met in all ANCOVAs).

The instructional materials developed for the study were entitled "Homo sapiens and the World Environment" and consisted of an explicit study guide, a reading supplement and activities supplement. The guides were designed to take about five weeks to complete, and were identical for both groups except that the experimental study guide began with an introduction which explained how concepts and principles are important in the learning process and how to prepare a concept map. In addition, the experimental study guide contained conceptual cues which reminded students of the importance of major concepts to related activities and readings.

Meaningful learning achievement was measured by three alternate form tests developed by the investigator and requiring application of newly learned concepts or principles; a pretest, a posttest and a six-week retention test. Content and theoretical validity were established for each of the tests. Multiple choice items were computer scored and the short answer questions were scored by the investigator and a colleague, and an intergrader correlation calculated. Students' locus of orientation was measured shortly before the instructional treatment began utilizing the Nowicki-Strickland internal-external (ANS-IE) locus of control scale and students divided into internally and externally oriented groups.

Real classrooms were used and schools differed in time taken to complete the units. Despite these inter-school differences, researchers felt that students in the study had access to the same instructional materials and more than adequate teacher assistance throughout the project's implementation.

Findings

Main effects. There were no differences in achievement of retention patterns between the two groups with the instructional treatment as independent variable. In general, the experimental treatment did not enhance subjects' meaningful learning achievement more than the comparison treatment. ANOVA with posttest and retention test scores as dependent variables and locus of control orientation as the independent variable resulted in a statistically significant F ratio, suggesting that meaningful learning achievement was influenced by locus of control orientation. Subjects with an internal locus of control orientation achieved at a higher level than subjects with an external locus of control orientation.

Interaction effect. The treatment/locus of control interaction effect was not statistically significant with retention test scores as the dependent variable, but was statistically significant with retention test scores as the dependent variable. Externally oriented subjects in the experimental group retained more information than externally oriented subjects in the comparison group. Internally oriented subjects retained about the same level of achievement regardless of treatment type. Overall, internally oriented subjects retained more information than externally oriented subjects.

Because previous studies had indicated that the effects of locus of control orientation on achievement may differ between males and females, ANCOVA analyses by treatment and locus of control orientation were performed on retention test scores for female and male subjects separately. The only statistically significant effect (at the 0.05 level) of these analyses was the locus of control effect resulting from the analysis with males. Although not stated in the text, the abstractor was able to determine from one of the graphs that males in the experimental group who were externally oriented performed better on the retention test than externally oriented males in the comparison group. Internally motivated males, however, performed better on the retention test if they had not received the experimental treatment. Graphs and discussion seemed to center on the results of the analysis with females, although the article indicated that that analysis was non-significant.

Interpretations

The researchers outlined four conclusions. They suggested that the experimental treatment did not enhance subjects' meaningful learning because the presence of the learning cues may not have been as important as the instructional materials read, activities completed and questions answered. Examination of study guides and interviews with subjects and teachers indicated that even though concept mapping tasks were attempted ' ' most subjects in the experimental group, many became frustrated and/or pressed for time after completing the first few concept maps. In the later study guide sections, concept maps were often not completed. They suggested that investigations that focus more upon the interaction of students with the instructional materials are needed before this result can be unequivocally accepted.

They also suggested that the possible difference in the treatment/locus of control interaction effect between female and male subjects indicated that externally oriented females may have been particularly responsive to specific instructional learning cues as a result of differential socialization experiences. For students with an external locus of control orientation, and especially for externally oriented females, the experimental instructional techniques may be useful in enhancing meaningful biology learning.

The treatment/locus of control orientation effect observed for retention test scores, but not for posttest scores, was explained by discussing the initial learning shock syndrome described by Ausubel, Novak and Hanesian (1978). The authors felt that the six week time period between completion of instructional materials and administration of the retention test would have allowed any learning shock effects to subside. The six week period would not have affected the stability of meaningfully learned concepts of principles, however.

They also felt that students with an internal locus of control orientation may always achieve at a higher level than externally oriented students, in part because they may be more likely to apply specific skills when working to succeed in a school subject. Because externally-oriented students may not be so likely to apply appropriate skills to a learning task, instructional guidance and specific learning suggestions could be especially helpful to them.

ABSTRACTOR'S ANALYSIS

There is no question that the study described in this article represents very important and thought-provoking science education research. As the authors conclude in their purpose and rationale, there is a need to broaden the knowledge base regarding instructional techniques which enhance meaningful learning, to investigate the relationship between locus of control orientation and achievement and to design meaningful learning-related studies that investigate the effects of interaction between instructional treatments and individual attributes. In this regard the researchers are to be commended for their effort to do just that. They have asked some intriguing questions and have attempted to design an experiment which will systematically answer those questions. Or have they?

They chose to study concept mapping as an instructional technique and to investigate the effects of this strategy on meaningful learning achievement as a function of students' locus of control orientation. One might question, why concept mapping? The abstractor assumed that the researchers were suggesting that this strategy is highly structured, perhaps providing the externally controlled learner with a structured processing skill that might facilitate their meaningful learning achievement. However, in thinking about it and also in reading the researchers' observations about student use of the technique, it would seem that the process itself is structured, but the student has free choice in deciding whether to use the technique or not. If students are externally controlled individuals, feeling that efforts on their part do not really affect their achievement, why would they choose to use this technique? In fact as indicated earlier, examination of study guides and interviews with subjects and teachers indicated that even though concept mapping tasks were attempted by most subjects in the experimental group, many became frustrated and/or pressed for time after completing the first few concept maps. In the later study guide sections, concept maps were often not completed.

Perhaps designing an instructional treatment that actually subtly forced students to use it would have been a better aproach. One example is a subsequent study attempting to investigate the effects of computer-assisted instruction (CAI) versus text mode programmed instruction (PI) and locus of control orientation on preservice teachers' achievement of science process skills (Wesley, Krockover and Devito, 1985). Although the subjects and outcomes are different in their study; Wesley, Krockover and Devito (1985) describe at length why they selected these particular treatments. They felt that both CAI and PI represent highly structured methods of instruction, but that the external discipline inherent in CAI insures that learners utilize instructional materials in the highly structured manner in which they were designed to be used. Therefore, they hypothesized that externally controlled students would perform better if exposed to CAI, rather that PI.

Whether one agrees with this logic or not, at least there was an attempt on the part of researchers to justify the choice of experimental treatment with the literature in their purpose and rationale. There was little effort to do this in the current article being reviewed. Interestingly enough, even with the effort to design an appropriate treatment, Wesley, Krockover & Devito (1985) observed few interaction effects either, concluding that the two treatments (computer-assisted instruction and text mode programmed instruction) were not significantly different enough to have a measurable effect on the dependent variable.

One can also question the design and instrumentation of the study and the resulting effect on the internal validity of the research. The groups compared were non-equivalent as measured by the pretest and the treatments not similarly implemented. The researchers justify this fact by suggesting that there is value gained from including as much data as possible from typical classroom situations and that this partially counteracts the imprecision resulting from variation among schools. Once again, the researchers are to be commended for their effort to use typical classroom situations and to involve as many subjects as they did in the study, particularly important when conducting ATI studies where Cronbach and Snow (1977) suggest at least 100 subjects for each comparison group. However, the Otis Lennon Mental Ability Test (OLMAT) scores actually indicated that groups were equivalent, while it was the investigator-designed pretest that indicated non-equivalence. Why were measures not taken to revise the pretest? Instead, this "non-equivalence" was handled by statistically manipulating the data, a dubious use of statistics at any time, but in this case particularly dangerous. The difference between the equivalence results as measured by the standardized OLMAT and the investigator-designed pretest should have been a red flag to revise the alternate form tests. It is difficult, therefore, to trust any of the results measured utilizing these instruments.

However, even if groups were non-equivalent, the justification given for proceeding with the study, is unjustifiable in this abstractor's opinion. There are ways to conduct more naturalistic "controlled" experimental studies that do not violate internal validity factors. Matching subjects, attempting to at least normalize implementation of treatments, and throwing out skewed data represent a few ways to do this. This abstractor has been a longtime advocate for "meaningful" naturalistic research but, in our effort to conduct such research, we should not sacrifice rigor in an effort to eliminate reductionism. In fact, we need to be even more rigorous with ourselves methodologically in order to justify the importance and meaningfulness of our data. In particular, if we choose to analyze our data quantitatively with parametric statistics, we need to methodologically be true to ourselves so that what we are saying happened (or did not happen) is truly justified.

One final methodological point to be made about this particular study deals with the interpretation, or perhaps better stated, the misinterpretation of the gender differences related to locus of control orientation/treatment interaction effects. Because previous studies had indicated that the effects of locus of control orientation on achievement might differ between males and females, ANCOVA analyses by treatment and locus of control orientation were performed on retention test scores for female and male subjects separately. The authors state that the only statistically significant effect (at the 0.05 level) of these analyses was the locus of control effect resulting from the analysis with males. Despite this, the researchers proceeded to discuss and graph the results of the non-significant female-subjects-only analysis, emphasizing these results in their findings and subsequent discussion when actually the only significant results, observed with males and supporting much of the data on external vs. internal controlled individuals, were not emphasized at all.

Methodological flaws aside, it was extremely interesting to the abstractor that what seemed to correlate most highly with meaningful learning achievement was not the presence of learning cues but were the number of instructional materials read, activities completed, and questions answered. What does this suggest but further evidence that time on task is still one of the most important variables, other than prior knowledge, associated with subsequent learning.

This study does represent an admirable attempt to investigate some extremely important issues to science education. As the researchers conclude, there still seems to be much data suggesting that internally motivated students will always perform better than externally motivated ones. However, it would be encouraging to discover instructional techniques that might, even in a small way, enhance the meaningful learning of externally motivated students. A subsequent study (Lehman, Carter & Kahle, 1985) has investigated the effects of concept mapping and vee mapping on the achievement of black high school students, and although these treatments were also not significant as compared to standard outlining, the results provide further insights. Obviously, there is still much to be done to understand the complex relationships that may exist between selected instructional strategies and individual differences of students.

REFERENCES

Cook, T. D. and D. T. Campbell. Quasi-experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally College Publishing Co., 1978.

Cronback, L. J. and R. E. Snow. Aptitudes and Instructional Methods: A Handbook for Research on Interactions. New York: Irvington Publishers, Inc., 1977.

Lehman, J. D., C. Carter, and J. B. Kahle. "Concept Mapping, Vee Mapping and Achievement: Results of a Field Study With Black High School Students." Journal of Research in Science Teaching, 22 (7): 663-673, 1985.

Wesley, B. E., G. H. Krockover, and A. Devito. "The Effects of Computer-Assisted Instruction and Locus of Control Upon Preservice Elementary Teachers' Acquisition of the Integrated Science Process Skills." Journal of Research in Science Teaching, 22 (8): 687-698, 1985.

Gorman, M. E. "Falsification in Experimental and Classroom Simulations." School Science and Mathematics, 86 (4): 306-321, 1986.

Descriptors--*College Students; *Educational Games; Educational Research; *Grouping Instructional Purposes; Higher Education; Problem Solving; *Science Education; *Science Instruction

Expanded abstract and analysis prepared especially for I.S.E. by Steven W. Gilbert, Ball State University, Muncie.

Purpose

The relative effectiveness of two hypothesis-testing strategies was compared using games called 2,4,6 and Eleusis. The two strategies were disconfirmation, in which a possible solution is evaluated by testing and eliminating alternatives, and confirmation, in which a player proposes and tests one correct possibility at a time. It was believed that a greater number of correct solutions would be found using disconfirming strategies than confirming strategies, mirroring Popper's concept of falsification.

Research Design and Procedure

Experiment 1. College students (N=120) in introductory psychology were randomly assigned to either disconfirmatory, confirmatory or control groups. Controls were given no strategy, confirmatory subjects were asked to solve a problem by validating hypotheses thought to be correct, and disconfirmatory subjects were instructed to try to invalidate ideas they thought to be correct. The simulation was 2, 4, 6. In this game, students are given a triplet of numbers arranged according to a rule unknown to them. By proposing subsequent triplets, they attempt to find the rule. The triplets they propose are shown to a judge, and students modify their ideas according to what is or is not accepted. In this case, no feedback was given about the correctness of the rules the students proposed until the end of the experiment.

Experiment 2. College students worked in groups of four playing Eleusis. In this game, playing cards are laid down in a row according to some rule known only to a dealer. Players propose cards to add to the row, which are either accepted or rejected. Rejected cards are laid alongside the row. Accepted cards become part of the row. Play continues until players are certain of their rule.

This experiment compared the solution rate for groups of students using one of three strategies: confirmatory, disconfirmatory (described for experiment 1) or combined. In the combined condition, students used a confirmatory strategy until they felt convinced of a hypothesis, then they switched to a disconfirmatory strategy. Four different, progressively more complex rules were sought in each of four games. After a group felt it had correctly identified a given rule, it went on to try to find the next one.

Experiment 3. This experiment was set up in the same way as experiment 2, except that subjects were told that a random error would be introduced into the dealer's confirmation. That is, the dealer would say "yes" to a wrong choice or "no" to a correct choice in 0-20% of his or her responses. This was intended to simulate the effect of experimental error in research. In fact, no error was introduced for any condition. Only three rules were tested for, rather than four, and a control condition, in which students were free to use their own strategy, replaced the combined condition. Twenty-four groups of four students were tested.

Findings

In both experiments 1 and 2, students using disconfirmatory strategies for finding rules were successful far more often than students in either of the other two conditions (Chi square significance equal to $p < .005$ and $p < .02$ respectively). Interestingly, the mean for the confirmatory condition was lower in both experiments than for either of the other two groups. Of the four consecutive rules sought in Eleusis (Experiment 2), the most pronounced differences in solution rates among groups were observed on the later, more difficult rules. The mean proportion of incorrect cards played was significantly higher for the disconfirmatory than for the confirmatory groups, indicating that the groups were following their assigned strategies. The introduction of possible error in Experiment 3 resulted in a dramatically lower number of solutions in all three groups. No group found rule 4, four groups found rule 3, and five groups found rule 1. Because of these figures, the significance of the difference among the groups could not be determined.

Interpretations

The authors conclude that both Eleusis and the 2,4,6 task can be used to illustrate the concept of falsification in the classroom. Conducting 2,4,6 activities as a class is suggested as one way of teaching problem-solving skills. This has been done by the author with classes of 12-80 college students, and in some junior high and high school classrooms.

The results of experiment 2 illustrate that falsification, in the form of disconfirmation, can be translated into a problem-solving strategy useful for finding hidden rules in Eleusis. It is suggested by direct observation that in some cases high school students may be more proficient at using disconfirmation than college students. These students are able to recognize ways that the game is both simiiar to, and different from, real science. The possibility of error hindered performance, even when groups were able to demonstrate that there were no errors in their proposed solutions. The author felt that introducing the possibility of error made a more difficult and realistic simulation of scientific reasoning.

ABSTRACTOR'S ANALYSIS

This study is actually a series of three studies, each related to, but not dependent upon, the others. Evidence is presented to support claims that (a) college students can be involved in problem-solving using 2,4,6 and Eleusis; (b) that different strategies for finding the rules in these games may produce significantly different solution rates; (c) that disconfirming hypotheses may be a more successful strategy for finding the rules of these games than confirming them; (d) that disconfirming is more productive than strategies normally used without instruction; and (e) that introducing the possibility of error dramatically decreases the success rate in finding rules underlying the game of Eleusis.

One of the most positive aspects of this research is that the analyses are straightforward and clear. The numbers clearly show the superiority of one approach over the other without complex statistical tests. Data are provided not only to support the outcome, but also to verify that the subjects were actually using the strategies they were instructed to use (as reflected in number of wrong cards played).

More attention could have been given to providing a clear description of the subjects, and the conditions under which the trials were run. The reader is forced to hunt for data that should be overt. For example, it is only by reading the results of experiment 1 that one learns of the number and division of subjects for that part of the work. Similarly, the total number of students participating in experiment 2 is not reported, although it is known that they are divided into groups of four ...and so on.

A more serious theoretical question is whether or not Popper's concept of falsification should have been used as the framework within which to report the research. In its larger sense, falsification is a criterion, not a problem-solving strategy per se. As long as a proposed solution can potentially be shown to be false, it may be considered a part of the scientific domain. This does not mean that it will be shown false, as the author implies nor does Poppler's idea lead to the conclusion that falsification through disconfirmation is the best scientific problem-solving strategy. Many ideas may never be directly or deliberately falsified. Disconfirmation in the history of science appears to be more a cumulative product of research, rather than a problem-solving strategy for a single research endeavor.

It has yet to be demonstrated whether scientists use disconfirmation in a significant amount of their research, or whether disconfirmation occurs de facto by the emergence of overwhelmingly supported theories. Thus, while the research is very interesting as a study of a specific kind of problem-solving, can it really be said to model scientific work? At present, it is difficult to say, but perhaps the answer is reflected in the comment that "...psychological research has shown that both scientists and college students do not falsify consistently on tasks that model scientific reasoning." If this is the case, it is not unreasonable to question the validity of the model. Interestingly, the author uses the results of these games to confirm his hypothesis that they are useful for teaching falsification, and does not attempt to disconfirm the idea. His approach reflects the notion that in a single research product, disconfirmation may not actually be either useful or convincing.

From a practitioner's standpoint, the incorporation of 2,4,6 and Eleusis into the precollege classrooms was particularly interesting. However, it needs to be pointed out that success at this level was not substantiated in this report. It does not appear to be the intention of the author to make such a claim, although the title of the paper alludes to "classroom simulations." It would certainly be interesting to see experimental results from high school or junior high school that are as concrete and dramatic as those reported here for college students. It would also be interesting to know how the authors actually connected the games, as simulations, to the practice of science.

Several questions suggest themselves as extensions of this research: (1) Can similar findings be produced by in-classroom research at the high school, junior high school, and upper elementary levels? (2) To what extent will students, without prompting, transfer the understanding and skill gained through these games to other problems solving situations? (3) To what extent will or do these games contribute to knowledge of the scientific process? (4) To what extent do these games even actually mirror the scientific process? (5) Will students who are unsuccessful using confirmatory strategies switch, without prompting, to using disconfirmation? (6) To what extent does 2,4,6 and Eleusis training affect problem-solving skills in "real" science situations? (7) How can these simulations be incorporated to best advantage into the science curriculum? (8) What are the effects of the games on student attitudes? (9) Why is disconfirmation more successful than confirmation?

Simulations such as Eleusis have been proposed for the science curriculum before. What is particularly interesting about this study are its findings regarding the apparent effectiveness of disconfirmation on problem-solving, and the paralysis induced by even the possibility of error. The application of this idea to illustrate scientific process is unfortunately not well-developed here, since the direct correspondence of the processes utilized in Eleusis and those used in scientific research has not been convincingly demonstrated.

Even with this in mind, the report is particularly fruitful, not only as to its findings, but also in the questions it raises, some of which have been asked above. A research program based on the use of simulations is clearly a priority in science education, given the potential such models have for developing the students' knowledge of science and scientific processes. The transferability of the skills thus gained is probably the most pressing question that needs to be answered now.

RESPONSES TO CRITIQUES

IN RESPONSE TO THE ANALYSIS OF

Yore, Larry D. "The Effect of Lesson Structure and Cognitive Style on the Science Achievement of Elementary School Children" by John W. Renner. Investigations in Science Education, 13 (2): 56-61, 1987.

Larry D. Yore
University of Victoria

Renner's abstract is an accurate, concise distillation of the research reported. It is impressive and somewhat embarrassing to see that one's efforts can be described in approximately one third the space as used to originally report it and without much loss of detail.

Renner's analysis is likewise concise and on target. His comments would improve the research design and likely yield data that would more completely illustrate the relationships between lesson structure, learner structure and science achievement. The question regarding whether the two science topics and their associated instruction were sequentially contiguous was not clearly answered in the article but was implied by the diagram of the research design. The two four-week instructional units were contiguous. It must be reiterated that the low-structure (X) and high-structure (X') treatments were reversed across the two groups after the completion of the first science topic, as illustrated below.

Group 1

X	0	X'	0
Magnets-Low	Magnets test	Mystery Powders-High	Mystery Powders Tests

Group 2

X'	0	X	0
Magnets-High	Magnets test	Mystery Powders-Low	Mystery Powders Tests

It is important that readers realize that each group was exposed to approximately four weeks of low-structure inquiry and four weeks of high-structure inquiry.

The concerns regarding length of treatment and classroom context are important considerations for future research. The short length of treatment issue is compounded when it is stressed that each group only received about four weeks each of low- and high-structured inquiry. Regardless of the actual duration, this researcher was "expecting too much achievement from two treatments which [were] too short" (Renner, 1987, p 61). The classroom context issue is interesting and clearly illustrates the need to report other qualitative information that will allow readers to more accurately interpret results. It is not clear that Renner's hypothesis, which suggests science inquiry embedded in an inquiry-oriented classroom would be more effective, is reasonable. It seems equally likely that larger and consistent differences would have been apparent if the students had not had other inquiry teaching/learning experience. If they were exposed to inquiry in other subject areas such as mathematics and social studies, students might have developed inquiry skills and coping strategies that would override the science treatments presented. Again, this is a valid issue worthy of further consideration.

The final response is to an issue that Renner did not react to fully: that is, the sources of structure in science teaching/learning. The two treatments used in the reported study were structurally different, but were they significantly different? Abraham and Renner (1983), Renner, Abraham and Birnie (1983) and others at the University of Oklahoma have found that the effects of structured learning cycles have not been consistent across physics, chemistry and biology. The results for physics and chemistry and preliminary results for biology appear to indicate that the inherent content structure of these subject areas may influence the teaching/learning cycle's effect on science achievement. The structured sequence of exploration, invention, and application did not produce significant achievement differences in physics, while the sequence did produce significant differences in the less structured content of chemistry and biology. The University of Oklahoma researchers have attributed the observed differences to the cognitive development of the target learners. The second part of this study and other studies, which considered cognitive development as an independent variable, have found support for such a hypothesis, but have not consistently verified such an interpretation (Shymansky & Yore, 1980; Yore, 1984; Yore, 1986).

In summary, future research on inquiry teaching/learning strategies needs to continue to consider learner attributes and the degree and source of structure over longer treatment periods in a variety of classroom environments. Such research will likely need to utilize a combination of quasi-experimental and naturalistic methodologies. The science topics selected for such research will need to be of low inherent content structure so that the addition of varying degrees of external teacher structure will result in instructional treatments that are significantly different structurally.

REFERENCES

Abraham, M. R. and J. W. Renner. "Sequencing language and Activities in Teaching High School Physics." Norman, OK: University of Oklahoma, 1983.

Renner, J. W., M. R. Abraham, and H. H. Birnie. "Sequencing Language and Activities in Teaching High School Chemistry. Norman, OK: University of Oklahoma, 1983.

Shymansky, J. A. and L. D. Yore. "A Study of Teaching Strategies, Student Cognitive Development, and Cognitive Style as They Relate to Student Achievement in Science." The Journal of Research in Science Teaching, 17: 369-382, 1980.

Yore, L. D. "The Effects of Cognitive Development and Age on Elementary Students' Science Achievement for Structured Inductive and Semi-deductive Inquiry Strategies." The Journal of Research in Science Teaching, 21: 745-753, 1984.

Yore, L. D. "Science Achievement, Cognitive Development, and Degree of Structure in Inquiry Lessons." School Science and Mathematics, 86: 271-283, 1986.

IN RESPONSE TO THE ANALYSIS OF

Dreyfus, A., E. Jungwirth, and P. Tamir. "Biology Education in Israel as viewed by Teachers" by Cheryl Mason. Investigations in Science Education, 13 (1): 58-64, 1987.

Amos Dreyfus
The Hebrew University of Jerusalem

There are a few comments we would like to make. As to Dr. Mason's remarks on p.6, i.e., that the paper lacks details about the rationale of the sample-selection, etc., we would like to point out that this paper was intended to concentrate on the process of the study and its results, rather than on too many details of method. Readers interested in these are referred to our paper, "An Approach to the Assessment of Teachers' Concerns in the Context of Curriculum Evaluation" in Studies in Educational Evaluation 8 (1): 87-100, 1982. We should just state here that the sample was composed to represent proportionally the various types of schools, so that the perceived "imbalance" was intended.

No criticism of teachers' views was either intended or expressed (p.7).

There has been a plethora of papers on biology education in Israel. For example, Science Education, JRST, Journal of College Science Teaching, European Journal of Science Education, Journal of Biology Education, etc. have all published articles by the present authors. A rather comprehensive overview of the Israeli BSCS-adaptation can be found in Tamir, P. and E. Jungwirth "Students' Growth as a Result of Studying BSCS-Biology for Several Years," Journal of Research in Science Teaching 12 (3): 263-279, 1975, which includes a comprehensive bibliography. The interested reader is also referred to the bibliography in the present paper and the companion paper cited above.

Further information is available to interested persons from the undersigned.

IN RESPONSE TO THE ANALYSIS OF

Hill, Douglas H. and Michael G. Redden. "Spatial Puzzles and the Assessment of Children's Problem-Solving Performance" by Meghan M. Twiest and D. Daryl Adams. Investigations in Science Education, 13 (4): 7-12, 1987.

Douglas M. Hill
Riverina-Murray Institute of Higher Education

The GEFT was "designed to provide an adaptation of the original individually administered EFT which would make possible group testing" (p. 26, Witkin, Oltman, Raskin and Karp, 1971). All but one of the items were taken from the EFT. The authors report a split halves reliability of 0.82 and "reasonably high" correlations between GEFT & EFT.

The same 35-piece jigsaw was used throughout the study. This was described as the "standard jigsaw." This jigsaw was completed some days after the GEFT was administered. The results are shown in Table 1 (below):

Table 1

Means and standard deviations for GEFT scores and time in seconds taken to complete a 35 piece jigsaw puzzle

		MEANS			STANDARD DEVIATIONS		
	Task	boys	girls	Total	boys	girls	Total
(i)	GEFT	8.53	6.85	7.71	3.65	2.39	3.19
(ii)	Jigsaw completion time	625.1	729.7	675.5	262.4	402.8	339.3

$r_{i.ii} = 0.50$ ($p < .01$)

The differences in means on the GEFT were significant ($t = 2.2$, $p < .05$) but not in case of jigsaw completion time ($t = 1.3$, p. 20).

Ten student volunteers were interviewed to help interpret observations and assist in identifying the strategies used in solving jigsaw puzzles. This information was used to help provide support for the notion that scores on GEFT are functionally related to performance on the jigsaw completion task.

All students were asked about their experience and competency with jigsaw puzzles. This proved to be a difficult task. Students found it difficult to describe their level of competency and often could not remember when they last completed a jigsaw puzzle.

The test-retest reliability for completion of a similar jigsaw was 0.79.

REFERENCE

Witkin, H. A., P. K. Oltman, E. Raskin and S. A. Karp. Manual for Embedded Figures Test. Palo Alto, CA: Consulting Psychologists, Inc., 1971.

IN RESPONSE TO THE ANALYSIS OF

Lawrenz, Frances. "An Evaluation of the Effect of two Different Lengths of Inservice Training on Teacher Attitudes" by Constance M. Perry. Investigations in Science Education, 13 (1): 36-43, 1987.

Frances Lawrenz
University of Minnesota

I believe Perry has provided an insightful analysis of my study and I absolutely agree with her that other, more carefully controlled, research is necessary to clarify the contribution of the various components of training to goal attainment. I view my study as an attempt to begin this investigation using real-world projects. In real-world projects experimental design is limited because it is uncommon for funding agencies to support the extra effort, random assignment, and exclusiveness that true research requires. I believe the two projects described in my study were as similar as they could be, given the differences in their length. Certainly fifteen weeks of training will provide a different experience than five weeks of training, and Perry was accurate in her reiteration of my caution of this as a limitation of the research. The study does, however, begin to address some of the problems inherent in designing inservice training for energy education. Even though the results may be confounded, it does appear that it is possible to effect attitude change in a five week as opposed to a fifteen week program. Naturally, this does not preclude the possibility that a fifteen week program could achieve other goals as well.

Perry raised several specific questions about the study that can be answered. Most importantly, she questioned the use of repeated measurement analysis and suggested that an analysis of covariance might have been more appropriate. The answer is "no." The participants in both programs had been pre- and post-tested with the same instruments so that individual change (as opposed to group change) over the course of the program could be assessed. Looking at change for an individual helps to control for bias between groups. For example, it would matter less if one group had more positive attitudes to begin with (unless there was a topping out effect), because the analysis would be based on an individual's growth from wherever he or she began rather than based on post-test scores. An analysis of covariance is another technique for controlling for initial differences, but it is a somewhat weaker approach in that the covariate is only related to the post-test and is not the post-test itself. The covariate, therefore, is not as accurate a measure of an individual's initial knowledge or attitude. Covariates are best used when scores in one area are believed to have an effect on scores in other areas, e.g., covary on content knowledge while examining attitude. The repeated measure analysis allows the researcher to take into account the variance due to subjects as well as providing a comparison of the two groups on the basis of their change over the course of the program.

Perry had two separate questions about significance testing. One related to the use of chi squares or some other appropriate nonparameteric statistic to compare the demographic type data obtained from the two groups. No statistical comparisons were completed. The statement of "no substantial differences" was qualitative and based on the questionnaire data and personal experience with the participants. No consistent differences between the two groups were perceived. The second question related to whether or not explaining fifteen percent of the variance was significant. The difference had already been shown to be statistically significant. Whether or not fifteen percent is educationally significant is a qualitative judgement, and I believe the answer is "yes."

IN RESPONSE TO THE ANALYSIS OF

Clough, Elizabeth and Rosalind Driver. "A Study of Consistency in the Use of Students' Conceptual Frameworks Across Different Task Contexts" by Anton E. Lawson. Investigations in Science Education, 14 (1): 22-29, 1988.

Elizabeth Clough
Sheffield University

Rosalind Driver
Leeds University

Anton Lawson's abstract of our paper presents a clear account of the study. However, in his statement "The basic prediction tested by Clough and Driver was that students would use the same theory across contexts that are different, but that scientists construe similarly." Lawson does suggest that the research was conducted within a tight design of the kind involved in hypothesis testing. The study was in fact a more open exploration of the ideas that children use in a range of contexts, and the contextual factors which appear to be influential. We recognise that the problem area is a messy one and we did not wish the design of the study to artificially reduce the complexity of the factors explored.

Lawson's interesting discussion of the paper focuses on four main issues. We will respond to each of these;

Generality of Conceptual Frameworks and Thinking Skills

The argument in support of the existence of thinking skills or conceptual frameworks requires further scrutiny. What does it mean to say such thinking patterns "exist"? We offer a number of interpretations:

1. an identifiable thinking pattern is used by an individual in a particular situation.

2. an identifiable thinking pattern is used by an individual in a number of situations.

3. an identifiable thinking pattern is used by members of a group of people in various situations.

In our study, conceptual frameworks were identified through the procedures we adopted as existing at the group level (i.e., 3 above). The question of their use across contexts by individuals was the subject of the investigation.

A parallel exists here with claims made about general thinking skills. These, too, undoubtedly exist at the group level. The issue is what claims can be made about their use by individuals.

As far as the hypothesis put forward by Lawson that both general thinking skills and conceptual frameworks exist and interact, we would see this as a useful line of enquiry.

Sources of Alternative Frameworks

The question of the sources of alternative frameworks is also a useful line of enquiry. In our paper we identified, as Lawson does, a number of possible sources including direct sensory experience and cultural influences. These often interact in their influence on the progressive development of childrens' scientific frameworks. In terms of which factors may have the most powerful influence on children's ideas, we would question Lawson's surmise that societal influences are necessarily stronger than personal ones. Indeed, in some cases, we suspect the reverse situation may be the case. For example, experiences with certain physical phenomena from an early age such as pushing objects and seeing them slow down and stop, we suspect contribute to the strong alternative conceptions that children (and adults) have in the area of mechanics.

Factors Which Influence Conceptual Change

Lawson's account of a recent study on the modifiability of alternative biological concepts is an interesting illustration of the complexity of the problem. He indicates that there is more than one reason for people to change their mind and suggests that both the authority of a person advocating a view (external authority), and the persuasiveness of evidence and arguments presented for a view, (internal authority) may affect whether one changes one's mind. Whether adopting a view because of an external authority necessarily amounts to changing one's mind is, as we suggest, a moot point. One may adopt a certain position for certain purposes (e.g., passing examinations) without necessarily accepting it as one's own view. A further point we would add to the complex issue of changing one's mind: that is, that adopting a new view (on whatever authority) requires one first to comprehend that view. There is thus an important distinction to be made between constructing alternative positions and appraising them. Both have a part to play in conceptual change.

The factors which influence conceptual change have been explored in the conceptual change literature (Hewson, 1981; West and Pines, 1985) and studies designed explicitly to promote conceptual change in a principled way have had limited success (Nussbaum and Novick, 1982). It is an area of enquiry where it is wise to recognise the complexity of the problem.

Teaching Methods

We agree with Lawson that the research task is to identify children's alternative conceptions, understand something of how they arise, and then investigate ways in which teachers and students can work on these together to change them, bringing them closer to accepted scientific ideas. However, we suspect that we may differ from Lawson in our understanding of learning processes in the classroom setting, something which will profoundly influence the management of this enquiry.

Recognition of the importance of learners' existing conceptions means according children's frameworks some status in the teaching and learning process. To really accept that the student carries ultimate responsibility for his or her own learning implies a shift in the power relationship between student and teacher; ultimately it implies a measure of student control over the curriculum. We believe that prescriptive instructional sequences, derived from a rational model of curriculum (such as SCIS) where the focus of control lies with the teacher, cannot deliver this fundamentally different approach to learning science. Teachers will need to learn to work diagnostically and more democratically with students if learners' ideas are properly to be taken account of, as individuals change their conceptual understanding.

As a final point we would argue that promoting conceptual change requires much more than implementing a specific pedagogical algorithm such as the three phases of the learning cycle. In order to engage with the ideas pupils bring and to provide them with experience which extend or change their ideas, knowledge of the general trajectory in children's understandings in specific domains is necessary.

Recent research by the Children's Learning in Science Project in England suggests that children tend to follow certain predictable paths in the development of their understanding of particular domains (Brook and Driver, 1989). We would argue that it is knowledge of this developmental trajectory, (which expert teachers often have a tacit feel for) that will inform and underline effective conceptual change strategies.

REFERENCES

Brook, A. and R. Driver. Progression in Science: the Development of Pupils' Understanding of Physical Characteristics of Air Across the Age Range 5-16 Years. .ildren's Learning in Science Project, CSSME, University of Leeds, 1989.

Driver, R. "Changing Conceptions" In: Adey P. (ed.) Proceedings of International Seminar Adolescent Development and School Science. King's College, London. September 13-17, 1987. Falmer Press, 1989.

Hewson, P. "A Conceptual Change Approach to Learning Science." European Journal of Science Education, 1981.

Nussbaum, J. and S. Novick. "Alternative Frameworks, Conceptual Conflict and Accommodation: Toward a Principled Teaching Strategy." Instructional Science, 1982.

West, L. and L. Pines. Cognitive Structure and Conceptual Change. Academic Press, 1985.

IN RESPONSE TO THE ANALYSIS OF

Staver, J. R. and D. A. Halsted. "The Effect of Reasoning on Student Performance on Difference Sections of a Posttest" by Geoffrey J. Giddings and Barry J. Fraser. Investigations In Science Education, 14 (1): 15-21, 1988.

John R. Staver
Kansas State University

In their analysis of our paper, Giddings and Fraser correctly pointed out problems with the research design. Initially, we wanted to answer the question, "When sections of a posttest differ in the use of manipulative models by students to answer questions, are the differences in performance attributable to differences in the reasoning levels of student (p. 171)?" And so we plead guilty to delineating the introduction, rationale, and problem statement as the research was conceived. In the field, however, the experiment is sometimes modified or it is lost. We chose to modify rather than lose the entire experiment, and the method, results, discussion, and implication sections not only describe the experiment as it was carried out, but point out the design problem which Giddings and Fraser noted. We regret any confusion that may have resulted by sticking to our original question, which we and the abstractors thought to be important.

However, we do take issue with points made by Giddings and Fraser about the analysis of the data, the credibility of our conclusions, and our interpretations. First, we stand by our statements, "Regression analysis and nonparametric ANOVA again indicate that reasoning, defined as a continuous distribution of PLOT scores, significantly predicts performance, and that formal thinkers outperform their concrete counterparts. But, the design of this preliminary study does not isolate the effect of reasoning level in either the presence/absence of models or the cognitive requirements of the items. The models/cognitive requirements issue remains confounded. But one can conclude that when these two characteristics of tests are coupled, then the effects of reasoning are manifested. The next research step would be to design a factorial investigation in which reasoning level, presence/absence of models, and cognitive requirements of questions are independent variables (p. 176)." We make no claim for an unconfounded test of the original question.

Second, Giddings and Fraser state that we tend to interpret the beta weights of the regression of dichotomous PLOT scores (formal/concrete) for the models (b = .25) and no models (b = .20) groups as suggesting the existence of a differential relationship between test performance and student reasoning ability depending on the presence/absence of models. We neither make nor imply such an interpretation. Nowhere in our article do we compare the differences between the beta weights, then use such differences to suggest that an interaction exists. Giddings and Fraser describe clearly why no such interpretation should be made. We agree. The principal rationale for initially dichotomizing PLOT scores was the insistence of reviewers who prefer this approach. Thus, we dichotomized PLOT scores in a rational manner, then cautioned readers. We, too, prefer PLOT scores as a continuous distribution.

In closing, we focus on the final point of Giddings' and Fraser's analysis, the need for a more comprehensive examination than we provided in this preliminary research. We noted previously that our initial question was, in fact, not answered. We did, however, follow up on our original intent by carrying out a factorial experiment (Staver & Halsted, 1985). Readers who are interested in the results of the factorial study may find it in the published literature.

REFERENCE

Staver, J. R. and D. A. Halsted. "The Effects of Reasoning, Use of Models, Sex Type, and Their Interactions on Posttest Achievement in Chemical Bonding After Constant Instruction." Journal of Research in Science Teaching, 22 (5): 437-447, 1985.

IN RESPONSE TO THE ANALYSIS OF

Zitzewitz, Barbara S. and Carl F. Berger. "Applications of Mathematical Learning Models to Student Performance on General Chemistry: Microcomputer Drill and Practice Programs" by Ann C. Howe. Investigations in Science Education, 14 (1): 33-37, 1988.

Carl F. Berger
The University of Michigan

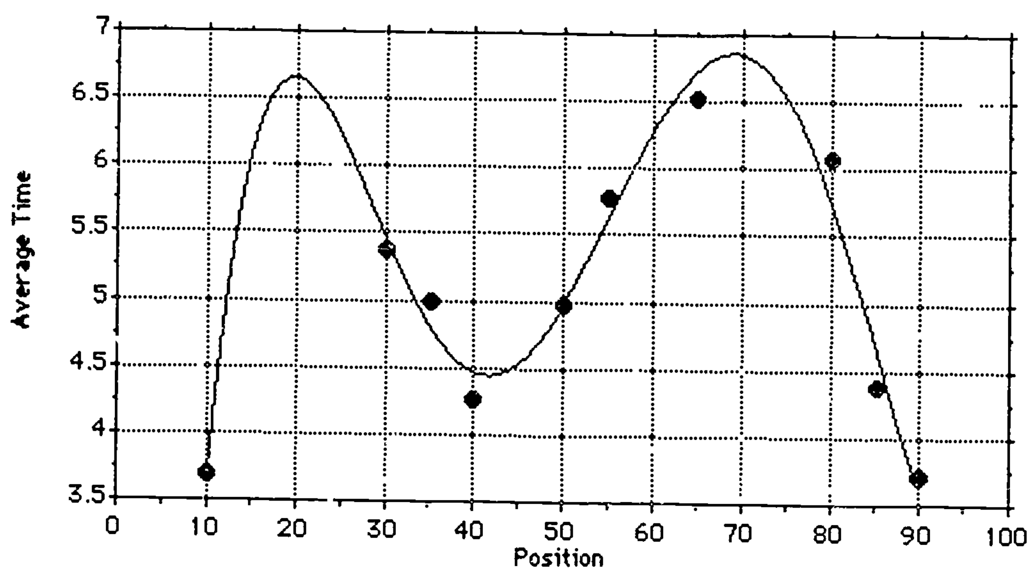
Analyzing data: How much "cooking" is necessary

Dr. Howe is certainly correct in citing the ease with which the microcomputer can be used to gather, analyze, and display data of learning studies and she is also correct in wondering if we do not, in fact, need to invoke the most parsimonious solution to such studies. In general, Dr. Zitzewitz and I would agree with her. It is interesting to note that she uses the comment by McNemar concerning his notion that there is no reason to go beyond the quadratic or second degree polynomial regression in order to find relationships in data. We were constantly surprised at the number of studies we examined in which people did not go beyond straightforward linear regression analysis. In today's world where we have more curvilinear relationships in education than linear relationships, we are often struck by the fact that researchers do not examine their data in any kind of graphic form to see if it is appropriate to look at polynomial regression. While in most instances we do not believe that one should necessarily continue to examine data ad infinitum, there are, indeed, instances where it is appropriate to look at higher order polynomial regressions.

In an analysis we did for another paper where we examined students' estimation on a number line, we found that students were using the center of the number line as an anchoring point in improving the time and accuracy of their estimation. Thus, in that case, a fourth order polynomial not only was correct but fit quite well with what we would predict such students to do: (1) have low estimation times at the bottom and top of the number line and (2) have a dip in the middle of the number line. (Please see illustration one)

Illustration One

Position of estimation on a number line versus average time to estimation



Thus we think that while it is important not to push data analysis to the limit, we are concerned that we should take a hard enough look at it to force us to consider the kinds of ways in which people think as they solve problems.

Testing Learning Models: Signal and Noise

Dr. Howe correctly raises the issue of testing learning models. Her issue concerns the problem of the learning on the task (signal) and its relation to the other multiple variables involved during that learning (noise). Until recently arguments were made that we did not have the capability for testing such models because the signal was too difficult to separate from the noise. More importantly, it was argued, we needed sufficiently specific presentation and testing conditions as well as analytic techniques to test the models. Due to these drawbacks, models were often presented as interesting applications developed from theoretical proposals of how people learned. Using such models, teaching and learning paradigms were developed that to this day remain taken on faith. If such models are touted as modeling real learning then we can no longer wait for techniques to so specifically define the conditions that they are unrealistic. Learning models must be tested in the real world with all the noise attendant in the surroundings. At the same time wwe must be careful to understand when such noise is so great that it prevents that analysis of a learning situation. In our study we could control the presentation and testing and could measure with accuracy the success and time taken to achieve results. To test whether we had a decent signal to noise ratio we observed and asked students to reflect on the task process. It may be argued that it would be as effective to focus on the learning model just by asking students to reflect as they are doing the problems, or after they are done. We believe such meta-cognitive research techniques suffer from several problems and should be carried out only in conjunction with graphical and statistical analysis. If meta-cognitive analysis is done while a student is doing his/her work, it tends to interfere with the process. In addition, we have heard of instances and have found with our own students that they modify what they believe they are doing after they have done it and have had an opportunity to reflect on it. We think both techniques are necessary and it was just such an analysis that encouraged us to believe that we could test models proposed by Aldridge and others. Under such conditions where we know the approaches taken by the subjects and where we have the technological capability to produce the graphics to illustrate student learning, we believe we can test and have tested specific models of learning. Using the techniques outlined in the study, others may more carefully define the conditions to more accurately test these models and possibly prove us wrong.

For these reasons we believe that the comments made by the reviewer should not be taken lightly. The review is excellent and, in the future, we will probably not publish higher order regressions unless we believe they theoretically and realistically enhance the discussion of the paper as well as more clearly defining the conditions under which subjects have carried out the task.