JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 40, SUPPLEMENT PP. S1–S7 (2003)

Dr. Novak has prepared his statement on science education research especially for the first issue of the Journal in the hope that it will stimulate thought and discussion by our readers, especially by members of NARST and AETS. Critical responses will be welcomed both by the author and by the Editor. In particular, Dr. Novak solicits comments useful to the NARST Committee on Criteria for Research of which he is currently the Chairman.

A Preliminary Statement on Research in Science Education

Joseph D. Novak

Purdue University, Lafayette, Indiana

Introduction

Research work in science education is a special area of scholarship within the scientific enterprise. The scientific enterprise ranks with the arts and religion as one of the major areas of human endeavor. Science education can be classified within science, albeit it stands as a poor cousin when compared with physics, biology, and other fields. The purpose of research in science education, nevertheless, is the same as that in other fields of science, *e.g.*, to advance the conceptual schemes which have been developed to explain events in the universe about us. Though science education is intellectually an infant in the scientific enterprise, there is reason to believe that advances in this field can eventually have as far reaching consequences as have developments in atomic theory and cell theory.

Until the advent of theory development based on experimentation, science, as we know it, advanced only slightly with the gradual accumulation of knowledge about the universe. As Nagel,[7] Crombie,[6] and others have pointed out, modern science began its rapid advance with the development of theoretical models which could be modified through experimentation. Alchemy might have persisted until today if a crude model for an atom had not been devised and subjected to experimental tests. Educational research, in general, and science education research, specifically, stand today in a position comparable to that of chemistry in the 18th century. Conant[4] has suggested that we are limited in our advance in education to the level obtained by skilled teachers working on the basis of past experience, just as the metallurgist of the 18th century was limited to the experience of the artisans Subsequent development of atomic molecular models has resulted in advances in metallurgy that the artisan could not hope for It is likely that substantial advances in education would result if we could develop models for learning (specifically learning o science) that have equivalent heuristic and ultimately, practical value. Unfortunately most research workers in education are a ignorant of the methods of science as th average lay person! Partly for this reason, research workers in the area of *science* education, at least some of whom have an understanding of science, are likely to make major contributions to educational research in general and indirectly to the scientific enterprise as a whole. This should be our goal.

The Role of Theory in Science Education Research

As is true for other areas in science, educational research must be based on the construction of model systems which can be submitted to test. An essential aspect of the test involves measurement or quantifica-

Originally published in Volume 1, Number 1, pp. 3-9 (1963).

DOI 10.1002/tea.10090
Published online in Wiley InterScience (www.interscience.wiley.com).

tion. In a recent symposium report, *Quantification,* problems of measurement in various fields were presented; it is significant that education was not included, though measurement in the field of psychology was discussed.[12] The omission of education as a field in this symposium conference points up the paucity of carefully devised measurements or quantification in this field. If one recognizes the inordinately crude manner in which measurement was attempted in most science education studies reported, the results are at best difficult to interpret. Except for status studies where direct "nose counts" are obtained, much of the research reported is empty. Unfortunately, results from status studies have primarily transitory value only and contribute little to our understanding of how science is learned.

It is not possible to develop accuracy in quantification unless we are clear as to the nature of the object or phenomenon we seek to measure. To obtain values for parameters we must first define the parameters we seek to estimate! We cannot estimate the mass of a proton unless we have some theoretical model which suggests how the measurement should proceed. Similarly, we cannot measure the level of understanding of a concept in science unless we have a model system which suggests the operations by which measurement of the level of understanding can be obtained.

There is a critical need for development of more adequate models for the learning process. The extensive application of computer theory to functioning of the brain in learning shows promise but more problems arise with this cybernetic model than are solved. Analysis of data processing, storage, and exchange by the brain is advancing rapidly; neurologists, psychologists, clinicians, and others are exploring output of the block box (brain). Almost no work by researchers using science classrooms for data collection has been reported with the intent of testing some cybernetic model of learning. We submit that the classroom provides a rich source of data on input and output information that could be analyzed for possible elucidation of the learning process.

Guidelines for Research in Science Education

Sources of Information

Science Education research has suffered from a lack of exploitation of research findings in other fields. A casual review of studies reporting work in science education will show many with no bibliography, uncritical acceptance of prior findings, and/or references limited to very similar studies. Basic research in science education can derive from analysis of literature in various areas of behavioral study; there should be no conspicuous boundary between research in science education and research in psychology, social psychology, animal behavior, neurology, clinical psychology, and related fields. A criterion of a carefully planned study which should be applied is the relative extent to which the researcher has drawn from findings in various fields of behavior. Though we need papers setting forth new ideas that have significance for research, more typically a research study that has no bibliography probably should not be published in any reputable journal.

Kinds of Research Studies

There is no classification system for research studies that does not contain serious flaws; at best, any system is artificial and does not represent a true relationship among the various kinds of research studies. The scheme proposed here was one discussed by a panel at the NARST meetings in 1961.[8] This scheme does present at least some suggestions for relating types of research studies—a consideration that needs attention.

1. Surveys and Status Studies

This type of study, often labeled as normative study, involves a survey or census to establish a norm or baseline from which trends could be discerned or which might be used as a basis for judging some aspect or quality in relation to a sample or a population. Except when unique data collection techniques are employed, the normative type of research would not be appropriate for doctoral students in science education. Obourn's point is well taken that most normative research should be done by local, state, or national agencies and research bureaus.[5]

2. Analytic Surveys

Closely related to the survey or normative study is the *analytic* survey. This type of research is similar to the normative study in that data are obtained regarding some aspect or quality of a sample or population; it differs in that there may be some measuring devices employed to discern possible relationships between parameters. Anderson points out that a survey may typically report a finding such as "the average teacher of chemistry has an average of 18.5 credits of chemistry in his preparation."[1] The analytic survey may employ a chemistry achievement test which shows that there is significant difference in pupil achievement when scores for students taught by teachers with greater preparation in science are compared with pupil scores obtained when taught by teachers with little preparation. It should be noted that analytic surveys do not discern causal relationships; they do identify possible causal factors that can be studied subsequently through experimentation. Moreover, properly designed analytic surveys can be useful probes into areas where obvious experiments are not apparent on the basis of extant theory. In this sense, the analytic survey may be of value in bridging gaps in theory, gaps that might be filled later through the experimentation-theory development process of science.

3. Experimental Studies

The most powerful tool available to man for obtaining new knowledge is the scientific experiment; it is through measurement of changes in some variables (independent) when other variables are held constant (or randomized) that we obtain new information with the greatest veracity. Failure of experimentation in education, in this respect, has been due largely to failure in adequate measurement, though frequently the design of experiments is also faulty. A large difference produced by a variable might be detected with a crude instrument (*e.g.*, the typical achievement examination). Smaller changes produced by a given variable (which may be of large practical significance) will require tests with greater precision. Unfortunately, the kinds of educational objectives difficult to measure (*e.g.*, gains in analytic thinking) are also least affected by changes in the usual experimental variables used in educational research.

Consider again the important relationship that exists between theory and experimentation. It is almost impossible to develop appropriate measuring devices and successful experiments without adequate theory. The fact that we know little about the nature of the problem-solving process results in frequent frustration in our experiments and measurement attempts.

Though much experimental research is needed in science education, we suffer from a lack of research workers who have continuous, basic research programs in progress. As Cooley pointed out, most studies in science education are one-shot efforts of graduate students pursuing doctorates.[5] It is difficult if not impossible to develop a theoretical framework on which to base experimental studies under these circumstances. It is impossible to construct, refine, and utilize necessary measuring instruments and techniques without organized, on-going research projects. This is a major area where much needs to be done, but little is in progress.

4. Curriculum Research

There is a type of scholarship in science education which does not fall into the categories given above. For lack of a better term we may class these studies as *curriculum* research. Barnes[2] pointed out that a variety of approaches have been used to study the question of what to teach in science and at what grade level. Some of these questions become philosophical in character; others may require some survey research as a beginning, *e.g.*, curriculum analysis based on students' interests *vis-a-vis* what is being taught.

One point needs repeated emphasis; science classes should teach *science;* not isolated facts, not the history of technology, but an understanding of the major ideas of science and the process by which these ideas are advanced.

One of the plagues of the established research worker in science education is the continuous flow of requests by doctoral candidates to "indicate in the blanks provided which principles you feel are important." The issue of what constitutes contemporary science is not resolved by a kind of "popularity poll" of science principles. Any competent scientist will agree that there are aspects of his science which are of greater or lesser significance; he will hasten to point out that the specific aspects selected for emphasis may be neither essential nor nonessential. A decision on the relative importance of some fact or principle can be made only with reference to the broader objectives of the instruction. For example, it is not of significance that a chemistry class include discussion on the noble elements, unless we wish to emphasize how obsolete notions on the nature of chemical bonding blocked for decades any attempt to perform the relatively simple synthesis of xenon tetrafluoride!

All curriculum studies must grow from a base of exceptional comprehension of science. Oddly enough, many studies in this area have been conducted by graduate students with a relatively poor knowledge of science. It is doubtful if any student should attempt a curriculum study in an area of science in which he has had less than 30 or 40 semester hours of graduate training, including some work in philosophy of science. By this standard alone, many of the poor studies in this area could have been avoided. Curriculum research is not an area for the student who wishes to expand his meager understanding of science.

An important area of science education is that of curriculum program development. The activities of PSSC, CBA, CHEMS, BSCS, and other groups are illustrative of this type of endeavor. Program development and program evaluation can be exceedingly complex, requiring the best talents of research scientists, teachers, and science education specialists. Evaluation of curriculum programs is greatly facilitated when objectives to be achieved by the program are specifically defined; unfortunately, the kinds of objectives the new major curriculum programs seek to obtain are not easily defined; consequently, evaluation has been superficial at best. Program development and program analysis should involve research of the four types listed above. But the critical factor is that research activity relative to a specific curriculum program must be carefully planned, coordinated, and executed. It is doubtful if sufficient science education research talent is available in the United States to staff the major curriculum programs, if adequate funds for such evaluation were available. There is a critical shortage of well trained, talented science education personnel.

Reporting Research Studies

A tremendous range in the quality of papers reporting research studies exists. For the most part, poor research reports are a consequence of poor research; it is difficult to present a clear statement of the problem or objective of the study if none was clearly defined at the outset. We have already pointed out that research (in the broad sense) must be based on some kind of theoretical framework. A research report should make clear the nature of the theoretical setting for a specific study, following this with a careful delineation of the purpose of the present study.

In reporting research of the four kinds listed above, considerable difference in key points to be emphasized exists. The following suggestions are proposed for the various kinds of research listed.

1. Surveys

Most survey findings are based on samples drawn from a larger population. In sample surveys it is essential to point out why the sample is not significantly biased. If the sample was selected at random from a population, it should be representative of that population, though the precision with which population parameters will be estimated is a function of sample size (and other factors[13]). Most sample surveys obtain a non-response group; the investigator should show that this non-response group has not introduced a bias.

A greater sophistication in the survey forms is needed. Frequently, information desired is not evident from the questions asked or space provided. Whenever possible, a survey form should be used with a pilot sample, requesting respondents to indicate any ambiguities, etc. A pilot sample from the population to be studied is more likely to be of value than the remarks of science education specialists, the currently most widely used reference.

Summary data should be presented in the best possible form. Though tables are valuable, graphic presentation may aid appreciably in some studies. Careful data reduction to a simple format which illustrates the population characteristics under study is seldom seen in science education literature.

2. Analytic Surveys

All of the points mentioned for surveys also apply to analytic surveys. There is the added concern here that measurements obtained for use in correlation analysis must be reliable and valid. We should not attempt to survey the pupil gains in biology under "traditional" and BSCS programs with a criterion test which is not valid for both groups. Some criteria, *e.g.*, size of school, may have different meanings for different samples; a high school with 300 students may be small, but this would not be true for an elementary school with 300 students.

The most important caution to be observed in reporting analytic studies is to differentiate between significant correlations and cause-and-effect relationships. A significantly higher achievement by pupils in large high schools may be entirely the consequence of better facilities and teachers available in the large schools studied. Though it is possible to parcel out variation from school size, facilities, teacher preparation, etc., through the use of analysis of covariance, there is always the chance that the significant variables were not studied and effects observed are only a consequence of the unstudied significant, but correlated, variables. Moreover, table values for analysis of variance and covariance statistics are computed on the assumption that the sample under study was a random sample from a normally distributed population. Statistical comparison of effects due to differences in sample treatments are not valid, since these treatments were not applied at random to the material studied but were in existence, and we have no way to determine what systematic bias existed in various groups. Albeit, the analytic survey may suggest what some significant variables might be.

3. Experimental Studies

Modern experimental design permits study of one or more dependent and independent variables simultaneously.[3] It is now possible to obtain answers to several related questions in one experimental study. However, application of modern experimental designs and the analysis of data obtained usually requires counsel from capable statisticians and other resource people. Simple studies could frequently be much improved if the experimenter sought such counsel.

A major difficulty (apart from design and analysis) in experimental studies derives from the unreliability of measuring instruments used. Though this has been true in other sciences, educational research suffers most acutely. The limitation imposed by the sensitivity of our test instrument is often critical. Suppose, for example, that we have a test capable of measuring critical thinking ability with only 10% error (this would be an excellent test of this type!). Now suppose we use this test before and after a two-month period of instruction where we believe we are improving an individual's ability to think critically (not to answer the type of questions on the criterion test, however, as is more usually the case). A real gain of 5% would probably result in the statistical finding that no significant gain occurred. But if a real gain of 5% occurred within two months, there would be general agreement that this is an enormous gain in ability to think critically and the practical significance of such a method, extended over several years, perhaps, would be exceedingly great!

Several points emerge from the example above. First, the duration of an experimental treatment should be considered in the light of probable change to be expected with time and the sensitivity of the measuring instrument. Second, accumulated small gains may have large practical significance, while quickly obtained large gains (in ability to define biological terms, for example) may be of little consequence.

Experimentation is the approach by which we establish causal relationships between variables. For this reason, experimentation proceeds best when the relationships submitted to test are suggested by some theory. Moreover, the results of the experiment then support or cast doubt on the theory, and thus lead to further experimentation. Reported experimental research in science education seldom specifies how the experiment derives from theory and how the results do or do not support the theory. This is a critical area that needs development if science education is to advance.

4. Curriculum Studies

The character of curriculum research is so variable that specific suggestions for reporting studies are not likely to apply to more than a minority of the studies. However, a few general considerations are worth noting.

There is a need for better definition of the purpose of the study, including careful description of the group, problem, or area of science under consideration. Every effort should be made to pursue curriculum studies which have more than local significance for a specific teaching problem, *e.g.*, grade placement of "space science" material in a local junior high school program. Curriculum "studies" that have little or no value to a larger science education community are probably not worth reporting, unless some promising new technique was utilized in the study.

We have noted above that curriculum studies should probably be done only by students with the most thorough training in science and philosophy of science. Curriculum studies can easily be found in the literature where the value of the specific project is not only questionable, but recommendations presented are in direct contradiction to the objectives in the field. Also, studies frequently dwell on untrue or irrelevant objectives for the science program proposed.[11]

By application of the above criteria alone, perhaps two-thirds or more curriculum studies would not be selected for publication. This is a "popular" type of research study which has dissipated much energy that might have been devoted to some other kind of science education research.

The road ahead for research workers in science education bristles with temptations and ramifies into myriads of blind alleys. We need talented research workers with the intellectual and moral integrity sufficient to survive the travails; hopefully, we may advance to a better understanding of how students learn *science*, and this could contribute to the advance of the welfare of mankind in general.

References

1. Anderson, Kenneth E., "Analytic Surveys," *Science Education*, **45** (5), 412–417 (December, 1961).

2. Barnes, Cyrus W., "A Definition of Science Education: Curriculum Research," *Science Education*, **45**(5), 394–396 (December 1961).

3. Cochran, Wm. G., and Gertrude M. Cox, *Experimental Designs*, Wiley, New York, 1957.

4. Conant, James B., *On Understanding Science*, Yale Univ. Press, New Haven, 1947.

5. Cooley, William W., "Challenges to the Improvement of Science Education Research," *Science Education*, **45**(5), 383–387 (December, 1961).

6. Crombie, A. C., *Medieval and Early Modern Science*, Doubleday, New York, 1959.

7. Nagel, Ernest, *The Structure of Science*, Harcourt, Brace, & World, New York, 1961.

8. NARST Panel Discussion, "A Definition of Science Education," NARST Annual Meetings, Washington, D.C., 1961, unpublished.

9. Obourn, Ellsworth, S., "Surveys and Status Studies," *Science Education*, **45**(5), 391–393 (December, 1961).

10. Pella, Milton O., "Criteria for Good Experimental Research in the Teaching of Science," *Science Education*, **45**(5), 396–399 (December, 1961).

11. Shamos, Morris H., "Science and Common Sense," *The Science Teacher*, **29**(5), 6–11 (September, 1962).

12. Woolf, Harry (Ed.), *Quantification*, Bobbs-Merrill, New York, 1961.

13. Yates, Frank, *Sampling Methods for Censuses and Surveys*, Griffin & Co., London, 1960.

JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 40, SUPPLEMENT, PP. S8–S18 (2003)

PART I

Cognitive Development in Children: Piaget

Development and Learning

JEAN PIAGET

Center for Genetic Epistemology, Geneva, Switzerland

My dear colleagues, I am very concerned about what to say to you, because I do not know if I shall accomplish the end that has been assigned to me. But I have been told that the important thing is not what you say, but the discussion which follows and the answers to questions you are asked. So this morning I shall simply give a general introduction of a few ideas which seem to me to be important for the subject of this conference.

First I would like to make clear the difference between two problems: the problem of *development* in general and the problem of *learning*. I think these problems are very different, although some people do not make this distinction.

The development of knowledge is a spontaneous process, tied to the whole process of embryogenesis. Embryogenesis concerns the development of the body, but it concerns as well the development of the nervous system and the development of mental functions. In the case of the development of knowledge in children, embryogenesis ends only in adulthood. It is a total developmental process which we must re-situate in its general biological and psychological context. In other words, development is a process which concerns the totality of the structures of knowledge.

Learning presents the opposite case. In general, learning is provoked by situations—provoked by a psychological experimenter; or by a teacher, with respect to some didactic point; or by an external situation. It is provoked, in general, as opposed to spontaneous. In addition, it is a limited process—limited to a single problem, or to a single structure.

So I think that development explains learning, and this opinion is contrary to the widely held opinion that development is a sum of discrete learning experiences. For some psychologists development is reduced to a series of specific learned items, and development is thus the sum, the cumulation of this series of specific items. I think this is an atomistic view which deforms the real state of things. In reality, development is the essential process and each element of learning occurs as a function of total development, rather than being an element which explains development. I shall begin, then, with a first part dealing with development, and I shall talk about learning in the second part.

To understand the development of knowledge, we must start with an idea which seems central to me—the idea of an *operation*. Knowledge is not a copy of reality. To know an object, to know an event, is not simply to look at it and make a mental copy or image of it. To know an object is to act on it. To know is to modify, to transform the object, and to understand the process of this transformation, and as a consequence to understand the way the object is constructed. An operation is thus the essence of knowledge; it is an interiorized action which modifies the object of knowledge. For instance an operation would consist of joining objects in a class to construct a

Originally published in Volume 2, Number 3, pp. 176-186 (1964).

classification. Or an operation would consist of ordering, or putting things in a series. Or an operation would consist of counting, or of measuring. In other words, it is a set of actions modifying the object, and enabling the knower to get at the structures of the transformation.

An operation is an interiorized action. But, in addition, it is a reversible action; that is, it can take place in both directions, for instance, adding or subtracting, joining or separating. So it is a particular type of action which makes up logical structures.

Above all, an operation is never isolated. It is always linked to other operations, and as a result it is always a part of a total structure. For instance, a logical class does not exist in isolation; what exists is the total structure of classification. An asymmetrical relation does not exist in isolation. Seriation is the natural, basic operational structure. A number does not exist in isolation. What exists is the series of numbers which constitute a structure, an exceedingly rich structure whose various properties have been revealed by mathematicians.

These operational structures are what seem to me to constitute the basis of knowledge, the natural psychological reality, in terms of which we must understand the development of knowledge. And the central problem of development is to understand the formation, elaboration, organization, and functioning of these structures.

I should like to review the stages of development of these structures, not in any detail, but simply as a reminder. I shall distinguish four main stages. The first is a sensory-motor, pre-verbal stage, lasting approximately the first 18 months of life. During this stage is developed the practical knowledge which constitutes the substructure of later representational knowledge. An example is the construction of the schema of the permanent object. For an infant, during the first months, an object has no permanence. When it disappears from the perceptual field it no longer exists. No attempt is made to find it again. Later, the infant will try to find it, and he will find it by localizing it spatially. Consequently, along with the construction of the permanent object there comes the construction of practical or sensory-motor space. There is similarly the construction of temporal succession, and of elementary sensory-motor causality. In other words, there is a series of structures which are indispensable for the structures of later representational thought.

In a second stage, we have pre-operational representation—the beginnings of language, of the symbolic function, and therefore of thought, or representation. But at the level of representational thought, there must now be a reconstruction of all that was developed on the sensory-motor level. That is, the sensory-motor actions are not immediately translated into operations. In fact, during all this second period of pre-operational representations, there are as yet no operations as I defined this term a moment ago. Specifically, there is as yet no conservation which is the psychological criterion of the presence of reversible operations. For example, if we pour liquid from one glass to another of a different shape, the pre-operational child will think there is more in one than in the other. In the absence of operational reversibility, there is no conservation of quantity.

In a third stage the first operations appear, but I call these concrete operations because they operate on objects, and not yet on verbally expressed hypotheses. For example, there are the operations of classification, ordering, the construction of the idea of number, spatial and temporal operations, and all the fundamental operations of elementary logic of classes and relations, of elementary mathematics, of elementary geometry, and even of elementary physics.

Finally, in the fourth stage, these operations are surpassed as the child reaches the level of what I call formal or hypothetic-deductive operations; that is, he can now reason on hypotheses, and not only on objects. He constructs new operations, operations of propositional logic, and not simply the operations of classes, relations, and numbers. He attains new structures which are on the one hand combinatorial, corresponding to what mathematicians call lattices; on the other hand, more complicated group structures. At the level of concrete operations, the operations apply within an immediate neighborhood: for instance, classification by successive inclusions. At the level of the combinatorial, however, the groups are much more mobile.

These, then, are the four stages which we identify, whose formation we shall now attempt to explain.

What factors can be called upon to explain the development from one set of structures to another? It seems to me that there are four main factors: first of all, *maturation*, in the sense of Gesell, since this development is a continuation of the embryogenesis; second, the role of *experience* of the effects of the physical environment on the structures of intelligence; third, *social transmission* in the broad sense (linguistic transmission, education, etc.); and fourth, a factor which is too often neglected but one which seems to me fundamental and even the principal factor. I shall call this the factor of *equilibration* or if you prefer it, of self-regulation.

Let us start with the first factor, maturation. One might think that these stages are simply a reflection of an interior maturation of the nervous system, following the hypotheses of Gesell, for example. Well, maturation certainly does play an indispensable role and must not be ignored. It certainly takes part in every transformation that takes place during a child's development. However, this first factor is insufficient in itself. First of all, we know practically nothing about the maturation of the nervous system beyond the first months of the child's existence. We know a little bit about it during the first two years but we know very little following this time. But above all, maturation doesn't explain everything, because the average ages at which these stages appear (the average chronological ages) vary a great deal from one society to another. The ordering of these stages is constant and has been found in all the societies studied. It has been found in various countries where psychologists in universities have redone the experiments but it has also been found in African peoples for example, in the children of the Bushmen, and in Iran, both in the villages and in the cities. However, although the order of succession is constant, the chronological ages of these stages varies a great deal. For instance, the ages which we have found in Geneva are not necessarily the ages which you would find in the United States. In Iran, furthermore, in the city of Teheran, they found approximately the same ages as we found in Geneva, but there is a systematic delay of two years in the children in the country. Canadian psychologists who redid our experiments, Monique Laurendeau and Father Adrien Pinard, found once again about the same ages in Montreal. But when they redid the experiments in Martinique, they found a delay of four years in all the experiments and this in spite of the fact that the children in Martinique go to a school set up according to the French system and the French curriculum and attain at the end of this elementary school a certificate of higher primary education. There is then a delay of four years, that is, there are the same stages, but systematically delayed. So you see that these age variations show that maturation does not explain everything.

I shall go on now to the role played by experience. Experience of objects, of physical reality, is obviously a basic factor in the development of cognitive structures. But once again this factor does not explain everything. I can give two reasons for this. The first reason is that some of the concepts which appear at the beginning of the stage of concrete operations are such that I cannot see how they could be drawn from experience. As an example, let us take the conservation of the substance in the case of changing the shape of a ball of plasticene. We give this ball of plasticene to a child who changes its shape into a sausage form and we ask him if there is the same amount of matter, that is, the same amount of substance as there was before. We also ask him if it now has the same weight and thirdly if it now has the same volume. The volume is measured by the displacement of water when we put the ball or the sausage into a glass of water. The findings, which have been the same every time this experiment has been done, show us that first of all there is conservation of the amount of substance. At about eight years old a child will say, "There is the same amount of plasticene." Only later does the child assert that the weight is conserved and still later that the volume is conserved. So I would ask you where the idea of the conservation of substance can come from. What is a constant and invariant substance when it doesn't yet have a constant weight or a constant volume? Through perception you can get at the weight of the ball or the volume of the ball but perception cannot give you an idea of the amount of substance. No experiment, no experience can show the child that there is the same amount of substance. He can weigh the ball and that would lead to the conservation of weight. He can immerse it in water and that would lead to the conservation of volume. But the notion of substance is attained before either weight or volume. This conservation of substance is simply a logical necessity. The child now understands that when there is a transformation something must be conserved because by reversing the transformation you can come back to the point of departure and once again have the ball. He knows that something is conserved but he doesn't know what. It is not yet the weight, it is not yet the volume; it is simply a logical form—a logical necessity. There, it seems to me, is an example of a progress in knowledge, a logical necessity for something to be conserved even though no experience can have lead to this notion.

My second objection to the sufficiency of experience as an explanatory factor is that this notion of experience is a very equivocal one. There are, in fact, two kinds of experience which are psychologically very different and this difference is very important from the pedagogical point of view. It is because of the pedagogical importance that I emphasize this distinction. First of all, there is what I shall call physical experience, and, secondly, what I shall call logical–mathematical experience.

Physical experience consists of acting upon objects and drawing some knowledge about the objects by abstraction from the objects. For example, to discover that this pipe is heavier than this watch, the child will weigh them both and find the difference in the objects themselves. This is experience in the usual sense of the term—in the sense used by empiricists. But there is a second type of experience which I shall call logical mathematical experience where the knowledge is not drawn from the objects, but it is drawn by the actions effected upon the objects. This is not the same thing. When one acts upon objects, the objects are indeed there, but there is also the set of actions which modify the objects.

I shall give you an example of this type of experience. It is a nice example because we have verified it many times in small children under seven years of age, but it is also an example which one of my mathematician friends has related to me about his own childhood, and he dates his mathematical career from this experience. When he was four or five years old—I don't know exactly how old, but a small child—he was seated on the ground in his garden and he was counting pebbles. Now to count these pebbles he put them in a row and he counted them one, two, three, up to ten. Then he finished counting them and started to count them in the other direction. He began by the end and once again he found ten. He found this marvelous that there were ten in one direction and ten in the other direction. So he put them in a circle and counted them that way and found ten once again. Then he counted them in the other direction and found ten once more. So he put them in some other arrangement and kept counting them and kept finding ten. There was the discovery that he made.

Now what indeed did he discover? He did not discover a property of pebbles; he discovered a property of the action of ordering. The pebbles had no order. It was his action which introduced a linear order or a cyclical order, or any kind of an order. He discovered that the sum was independent of the order. The order was the action which he introduced among the pebbles. For the sum the same principle applied. The pebbles had no sum; they were simply in a pile. To make a sum, action was necessary—the operation of putting together and counting. He found that the sum was independent of the order, in other words, that the action of putting together is independent of the action of ordering. He discovered a property of actions and not a property of pebbles. You may say that it is in the nature of pebbles to let this be done to them and this is true. But it could have been drops of water, and drops of water would not have let this be done to them because two drops of water and two drops of water do not make four drops of water as you know very well. Drops of water then would not let this be done to them, we agree to that.

So it is not the physical property of pebbles which the experience uncovered. It is the properties of the actions carried out on the pebbles, and this is quite another form of experience. It is the point of departure of mathematical deduction. The subsequent deduction will consist of interiorizing these actions and then of combining them without needing any pebbles. The mathematician no longer needs his pebbles. He can combine his operations simply with symbols, and the point of departure of this mathematical deduction is logical–mathematical experience, and this is not at all experience in the sense of the empiricists. It is the beginning of the coordination of actions, but this coordination of actions before the stage of operations needs to be supported by concrete material. Later, this coordination of actions leads to the logical–mathematical structures. I believe that logic is not a derivative of language. The source of logic is much more profound. It is the total coordination of actions, actions of joining things together, or ordering things, etc. This is what logical–mathematical experience is. It is an experience of the actions of the subject, and not an experience of objects themselves. It is an experience which is necessary before there can be operations. Once the operations have been attained this experience is no longer needed and the coordinations of actions can take place by themselves in the form of deduction and construction for abstract structures.

The third factor is social transmission–linguistic transmission or educational transmission. This factor, once again, is fundamental. I do not deny the role of any one of these factors; they all play a part. But this factor is insufficient because the child can receive valuable information via language or via education directed by an adult only if he is in a state where he can understand this information. That is, to receive the information he must have a structure which enables him to assimilate this information. This is why you cannot teach higher mathematics to a five-year-old. He does not yet have structures which enable him to understand.

I shall take a much simpler example, an example of linguistic transmission. As my very first work in the realm of child psychology, I spent a long time studying the relation between a part and a whole in concrete experience and in language. For example, I used Burt's test employing the sentence, "Some of my flowers are buttercups." The child knows that all buttercups are yellow, so there are three possible conclusions: the whole bouquet is yellow, or part of the bouquet is yellow, or none of the flowers in the bouquet are yellow. I found that up until nine years of age (and this was in Paris, so the children certainly did understand the French language) they replied, "The whole bouquet is yellow or some of my flowers are yellow." Both of those mean the same thing. They did not understand the expression, "some *of* my flowers." They did not understand this *of* as a partitive genitive, as the inclusion of some flowers in my flowers. They understood some of my flowers to be my several flowers as if the several flowers and the flowers were confused as one and the same class. So there you have children who until nine years of age heard every day a linguistic structure which implied the inclusion of a subclass in a class and yet did not understand this structure. It is only when they themselves are in firm possession of this logical structure, when they have constructed it for themselves according to the developmental laws which we shall discuss, that they succeed in understanding correctly the linguistic expression.

I come now to the fourth factor which is added to the three preceding ones but which seems to me to be the fundamental one. This is what I call the factor of equilibration. Since there are already three factors, they must somehow be equilibrated among themselves. That is one reason for bringing in the factor of equilibration. There is a second reason, however, which seems to me to be fundamental. It is that in the act of knowing, the subject is active, and consequently, faced with an external disturbance, he will react in order to compensate and consequently he will tend towards equilibrium. Equilibrium, defined by active compensation, leads to reversibility. Operational reversibility is a model of an equilibrated system where a transformation in one direction is compensated by a transformation in the other direction. Equilibration, as I understand it, is thus an active process. It is a process of self-regulation. I think that this self-regulation is a fundamental factor in development. I use this term in the sense in which it is used in cybernetics, that is, in the sense of processes with feedback and with feedforward, of processes which regulate themselves by a progressive compensation of systems. This process of equilibration takes the form of a succession of levels of equilibrium, of levels which have a certain probability which I shall call a sequential probability, that is, the probabilities are not established *a priori*. There is a sequence of levels. It is not possible to reach the second level unless equilibrium has been reached at the first level, and the equilibrium of the third level only becomes possible when the equilibrium of the second level has been reached, and so forth. That is, each level is determined as the most probable given that the preceding level has been reached. It is not the most probable at the beginning, but it is the most probable once the preceding level has been reached.

As an example, let us take the development of the idea of conservation in the transformation of the ball of plasticene into the sausage shape. Here you can discern four levels. The most probable at the beginning is for the child to think of only one dimension. Suppose that there is a probability of 0.8, for instance, that the child will focus on the length, and that the width has a probability of 0.2. This would mean that of ten children, eight will focus on the length alone without paying any attention to the width, and two will focus on the width without paying any attention to the length. They will focus only on one dimension or the other. Since the two dimensions are independent at this stage, focusing on both at once would have a probability of only 0.16. That is less than either one of the two. In other words, the most probable in the beginning is to focus only on one dimension and in fact the child will say, "It's longer, so there's more in the sausage." Once he has reached this first level, if you continue to elongate the sausage, there comes a moment when he will say, "No, now it's too thin, so there's less." Now he is thinking about the width, but he forgets the length, so you have come to a second level which becomes the most probable after the first level, but which is not the most probable at the point of departure. Once he has focused on the width, he will come back sooner or later to focus on the length. Here you will have a third level where he will oscillate between width and length and where he will discover that the two are related. When you elongate you make it thinner, and when you make it shorter, you make it thicker. He discovers that the two are solidly related and in discovering this relationship, he will start to think in terms of transformation and not only in terms of the final configuration. Now he will say that when it gets longer it gets thinner, so it's the same thing. There is more of it in length but less of it in width. When you make it shorter it gets thicker; there's less in length and more in width, so there is compensation—compensation which defines equilibrium in the sense in which I defined it a moment ago. Consequently, you have operations and conservation. In other words, in the course of these developments you will always find a process of self-regulation which I call equilibration and which seems to me the fundamental factor in the acquisition of logical–mathematical knowledge.

I shall go on now to the second part of my lecture, that is, to deal with the topic of learning. Classically, learning is based on the stimulus–response schema. I think the stimulus–response schema, while I won't say it is false, is in any case entirely incapable of explaining cognitive learning. Why? Because when you think of a stimulus–response schema, you think usually that first of all there is a stimulus and then a response is set off by this stimulus. For my part, I am convinced that the response was there first, if I can express myself in this way. A stimulus is a stimulus only to the extent that it is significant, and it becomes significant only to the extent that there is a structure which permits its assimilation, a structure which can integrate this stimulus but which at the same time sets off the response. In other words, I would propose that the stimulus–response schema be written in the circular form—in the form of a schema or of a structure which is not simply one way. I would propose that above all, between the stimulus and the response, there is the organism, the organism and its structures. The stimulus is really a stimulus only when it is assimilated into a structure and it is this structure which sets off the response. Consequently, it is not an exaggeration to say that the response is there first, or if you wish at the beginning there is the structure. Of course we would want to understand how this structure comes to be. I tried to do this earlier by presenting a model of equilibration or self-regulation. Once there is a structure, the stimulus will set off a response, but only by the intermediary of this structure.

I should like to present some facts. We have facts in great number. I shall choose only one or two and I shall choose some facts which our colleague, Smedslund, has gathered. (Smedslund is currently at the Harvard Center for Cognitive Studies.) Smedslund arrived in Geneva a few years ago convinced (he had published this in one of his papers) that the development of the ideas of conservation could be indefinitely accelerated through learning of a stimulus–response type. I invited Smedslund to come to spend a year in Geneva to show us this, to show us that he could accelerate the development of operational conservation. I shall relate only one of his experiments.

During the year that he spent in Geneva he chose to work on the conservation of weight. The conservation of weight is, in fact, easy to study since there is a possible external reinforcement, that is, simply weighing the ball and the sausage on a balance. Then you can study the child's reactions to these external results. Smedslund studied the conservation of weight on the one hand, and on the other hand he studied the transitivity of weights, that is, the transitivity of equalities if A = B and B = C, then A = C, or the transitivity of the inequalities if A is less than B, and B is less than C, then A is less than C.

As far as conservation is concerned,

Smedslund succeeded very easily with five- and six-year-old children in getting them to generalize that weight is conserved when the ball is transformed into a different shape. The child sees the ball transformed into a sausage or into little pieces or into a pancake or into any other form, he weighs it, and he sees that it is always the same thing. He will affirm it will be the same thing, no matter what you do to it; it will come out to be the same weight. Thus Smedslund very easily achieved the conservation of weight by this sort of external reinforcement.

In contrast to this, however, the same method did not succeed in teaching transitivity. The children resisted the notion of transitivity. A child would predict correctly in certain cases but he would make his prediction as a possibility or a probability and not as a certainty. There was never this generalized certainty in the case of transitivity.

So there is the first example, which seems to me very instructive, because in this problem in the conservation of weight there are two aspects. There is the physical aspect and there is the logical–mathematical aspect. Note that Smedslund started his study by establishing that there was a correlation between conservation and transitivity. He began by making a statistical study on the relationships between the spontaneous responses to the questions about conservation and the spontaneous responses to the questions about transitivity, and he found a very significant correlation. But in the learning experiment, he obtained a learning of conservation and not of transitivity. Consequently, he successfully obtained a learning of what I called earlier physical experience (which is not surprising since it is simply a question of noting facts about objects), but he did not successfully obtain a learning in the construction of the logical structure. This doesn't surprise me either, since the logical structure is not the result of physical experience. It cannot be obtained by external reinforcement. The logical structure is reached only through internal equilibration, by self-regulation, and the external reinforcement of seeing that the balance did not suffice to establish this logical structure of transitivity.

I could give many other comparable examples, but it seems useless to me to insist upon these negative examples. Now I should like to show that learning is possible in the case of these logical–mathematical structures, but on one condition—that is, that the structure which you want to teach to the subjects can be supported by simpler, more elementary, logical–mathematical structures. I shall give you an example. It is the example of the conservation of number in the case of one-to-one correspondence. If you give a child seven blue tokens and ask him to put down as many red tokens, there is a preoperational stage where he will put one red one opposite each blue one. But when you spread out the red ones, making them into a longer row, he will say to you, "Now, there are more red ones than there are blue ones."

Now how can we accelerate, if you want to accelerate, the acquisition of this conservation of number? Well, you can imagine an analogous structure but in a simpler, more elementary situation. For example, with Mlle. Inhelder, we have been studying recently the notion of one-to-one correspondence by giving the child two glasses of the same shape and a big pile of beads. The child puts a bead into one glass with one hand and at the same time a bead into the other glass with the other hand. Time after time he repeats this action, a bead into one glass with one hand and at the same time a bead into the other glass with the other hand and he sees that there is always the same amount on each side. Then you hide one of the glasses. You cover it up. He no longer sees this glass but he continues to put one bead into it while at the same time putting one bead into the other glass which he can see. Then you ask him whether the equality has been conserved, whether there is still the same amount in one glass as in the other. Now you will find that very small children, about four years old, don't want to make a prediction. They will say, "So far, it has been the same amount, but now I don't know. I can't see any more, so I don't know." They do not want to generalize. But the generalization is made from the age of about five and one-half years.

This is in contrast to the case of the red and blue tokens with one row spread out, where it isn't until seven or eight years of age that children will say there are the same number in the two rows. As one example of this generalization, I recall a little boy of five years and nine months who had been adding the beads to the glasses for a little while. Then we asked him whether, if he continued to do this all day and all night and all the next day, there would always be the same amount in the two glasses. The little boy gave this admirable reply. "Once you know, you know for always." In other words, this was recursive reasoning. So here the child does acquire the structure in this specific case. The number is a synthesis of class inclusion and ordering. This synthesis is being favored by the child's own actions. You have set up a situation where there is an iteration of one same action which continues and which is therefore ordered while at the same time being inclusive. You have, so to speak, a localized synthesis of inclusion and ordering which facilitates the construction of the idea of number in this specific case, and there you can find, in effect, an influence of this experience on the other experience. However, this influence is not immediate. We study the generalization from this recursive situation to the other situation where the tokens are laid on the table in rows, and it is not an immediate generalization but it is made possible through intermediaries. In other words, you can find some learning of this structure if you base the learning on simpler structures.

In this same area of the development of numerical structures, the psychologist Joachim Wohlwill, who spent a year at our Institute at Geneva, has also shown that this acquisition can be accelerated through introducing additive operations, which is what we introduced also in the experiment which I just described. Wohlwill introduced them in a different way but he too was able to obtain a certain learning effect. In other words, learning is possible if you base the more complex structure on simpler structures, that is, when there is a natural relationship and development of structures and not simply an external reinforcement.

Now I would like to take a few minutes to conclude what I was saying. My first conclusion is that learning of structures seems to obey the same laws as the natural development of these structures. In other words, learning is subordinated to development and not vice-versa as I said in the introduction. No doubt you will object that some investigators have succeeded in teaching operational structures. But, when I am faced with these facts, I always have three questions which I want to have answered before I am convinced.

The first question is: "Is this learning lasting? What remains two weeks or a month later?" If a structure develops spontaneously, once it has reached a state of equilibrium, it is lasting, it will continue throughout the child's entire life. When you achieve the learning by external reinforcement, is the result lasting or not and what are the conditions necessary for it to be lasting?

The second question is: "How much generalization is possible?" What makes learning interesting is the possibility of transfer of a generalization. When you have brought about some learning, you can always ask whether this is an isolated piece in the midst of the child's mental life, or if it is really a dynamic structure which can lead to generalizations.

Then there is the third question: "In the case of each learning experience what was the operational level of the subject before the experience and what more complex structures has this learning succeeded in achieving?" In other words, we must look at each specific learning experience from the point of view of the spontaneous operations which were present at the outset and the operational level which has been achieved after the learning experience.

My second conclusion is that the fundamental relation involved in all development and all learning is not the relation of association. In the stimulus–response schema, the relation between the response and the stimulus is understood to be one of association. In contrast to this, I think that the fundamental relation is one of assimilation. Assimilation is not the same as association. I shall define assimilation as the integration of any sort of reality into a structure, and it is this assimilation which seems to me to be fundamental in learning, and which seems to me to be the fundamental relation from the point of view of pedagogical or didactic applications. All of my remarks today represent the child and the learning subject as active. An operation is an activity. Learning is possible only when there is active assimilation. It is this activity on the part of the subject which seems to me to be underplayed in the stimulus–response schema. The presentation which I propose puts the emphasis on the idea of self-regulation, on assimilation. All the emphasis is placed on the activity of the subject himself, and I think that without this activity there is no possible didactic or pedagogy which significantly transforms the subject.

Finally, and this will be my last concluding remark, I would like to comment on an excellent publication by the psychologist Berlyne. Berlyne spent a year with us in Geneva during which he intended to translate our results on the development of operations into stimulus–response language, specifically into Hull's learning theory. Berlyne published in our series of studies of genetic epistomology a very good article on this comparison between the results obtained in Geneva and Hull's theory. In the same volume, I published a commentary on Berlyne's results. The essence of Berlyne's results is this: Our findings can very well be translated into Hullian language, but only on condition that two modifications are introduced. Berlyne himself found these modifications quite considerable, but they seemed to him to concern more the conceptualization than the Hullian theory itself. I am not so sure about that. The two modifications are these. First of all, Berlyne wants to distinguish two sorts of response in the S-R schema: (*a*) responses in the ordinary, classical sense, which I shall call "copy responses;" (*b*) responses which Berlyne calls "transformation responses." Transformation responses consist of transforming one response of the first type into another response of the first type. These transformation responses are what I call operations, and you can see right away that this is a rather serious modification of Hull's conceptualization because here you are introducing an element of transformation and thus of assimilation and no longer the simple association of stimulus–response theory.

The second modification which Berlyne introduces into the stimulus–response language is the introduction of what he calls internal reinforcements. What are these internal reinforcements? They are what I call equilibration or self-regulation. The internal reinforcements are what enable the subject to eliminate contradictions, incompatibilities, and conflicts. All development is composed of momentary conflicts and incompatibilities which must be overcome to reach a higher level of equilibrium. Berlyne calls this elimination of incompatibilities internal reinforcements.

So you see that it is indeed a stimulus–response theory, if you will, but first you add operations and then you add equilibration. That's all we want!

Editor's note: A brief question and answer period followed Professor Piaget's presentation. The first question related to the fact that the eight-year-old child acquires conservation of weight and volume. The question asked if this didn't contradict the order of emergence of the pre-operational and operational stages. Piaget's response follows:

The conservation of weight and the conservation of volume are not due only to experience. There is also involved a logical framework which is characterized by reversibility and the system of compensations. I am only saying that in the case of weight and volume, weight corresponds to a perception. There is an empirical contact. The same is true of volume. But in the case of substance, I don't see how there can be any perception of substance independent of weight or volume. The strange thing is that this notion of substance comes before the two other notions. Note that in the history of thought we have the same thing. The first Greek physicists, the pre-socratic philosophers, discovered conservation of substance independently of any experience. I do not believe this is contradictory to the theory of operations. This conservation of substance is simply the affirmation that something must be conserved. The children do not know specifically what is conserved. They know that since the sausage can become a ball again there must be something which is conserved, and saying "substance" is simply a way of translating this logical necessity for conservation. But this logical necessity results directly from the discovery of operations. I do not think that this is contradictory with the theory of development.

Editor's note: The second question was whether or not the development of stages in children's thinking could be accelerated by practice, training, and exercise in perception and memory. Piaget's response follows:

I am not very sure that exercise of perception and memory would be sufficient. I think that we must distinguish within the cognitive function two very different aspects which I shall call the figurative aspect and the operative aspect. The figurative aspect deals with static configurations. In physical reality there are states, and in addition to these there are transformations which lead from one state to another. In cognitive functioning one has the figurative aspects—for example, perception, imitation, mental imagery, etc.

The operative aspect includes operations and the actions which lead from one state to another. In children of the higher stages and in adults, the figurative aspects are subordinated to the operative aspects. Any given state is understood to be the result of some transformation and the point of departure for another transformation. But the pre-operational child does not understand transformations. He does not have the operations necessary to understand them so he puts all the emphasis on the static quality of the states. It is because of this, for example, that in the conservation experiments he simply compares the initial state and the final state without being concerned with the transformation.

In exercising perception and memory, I feel that you will reinforce the figurative aspect without touching the operative aspect. Consequently, I'm not sure that this will accelerate the development of cognitive structures. What needs to be reinforced is the operative aspect—not the analysis of states, but the understanding of transformations.

JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 40, SUPPLEMENT, PP. S19–S32 (2003)

RESEARCH PAPERS

WAIT-TIME AND REWARDS AS INSTRUCTIONAL VARIABLES, THEIR INFLUENCE ON LANGUAGE, LOGIC, AND FATE CONTROL: PART ONE–WAIT-TIME

MARY BUDD ROWE

University of Florida, Gainesville, Florida 32601

> The paradigm observer is not the man who sees and reports what all normal observers see and report, but the man who sees in familiar objects what no one else has seen before.
>
> –Norwood Russell Hanson
> in "Patterns of Discovery"

Synopsis

Part One of this paper summarizes six years of investigations devoted to the study of the influence that a variable called teacher "wait-time" has on development of language and logic in children taking part in elementary science programs. Part Two, which will be published in vol. 58, no. 4 of this journal, describes the relation of overt verbal rewards to the outcome variables, especially the possible relation between pauses, rewards, and fate control.

Analysis of a first set of more than 300 tape recordings showed mean wait-time to be on the order of one second. After a teacher asks a question students must begin a response within an average time of one second. If they do not the teacher repeats, rephrases, or asks a different question or calls on another student. A second potential wait-time is involved. When a student makes a response, the teacher normally reacts or asks another question within an average time of 0.9 seconds.

When mean wait-times of three to five seconds are achieved through training, analysis of more than 900 tapes shows changed values on ten student variables: 1. The length of response increases. 2. The number of unsolicited but appropriate responses increases. 3. Failures to respond decrease. 4. Confidence as reflected in decrease of inflected responses increases. 5. Incidence of speculative responses increases. 6. Incidence of child-child comparisons of data increases. 7. Incidence of evidence-inference statements increases. 8. The frequency of student questions increases. 9. Incidence of responses from students rated by teachers as relatively slow increases. 10. The variety in type moves made by students increases.

Servo-chart plots of recordings show that students discussing science phenomena tend to speak in bursts with intervals of three to five seconds between bursts being fairly common. The average post-student response wait-time of 0.9 seconds apparently intervenes between bursts to prevent completion of thought.

The classroom is conceptualized as a two-player game in which the quality of inquiry will tend to improve when there is better equity in the distribution of moves between the players.

Originally published in Volume 11, Number 2, pp. 81-94 (1974).

Over time a classroom on the prolonged wait-time schedule takes on other properties. Three teacher variables change: 1. Response flexibility scores increase. 2. Teacher questioning patterns becomes managable. 3. There is some indication that teacher expectations for performance of students rated as relatively slow improves.

A model which involves the relation of wait-time and reward as input variables to language, logic, and fate control as complex outcome variables is discussed.*

Introduction

Virtually all of the major elementary science programs extant today were designed to provoke children to inquire about relationships among natural phenomena. All of them provide situations meant to be suitable for the development in children of certain skills and a viable knowledge structure. In spite of provocative stimuli, the people who prepared the programs frequently admitted that the amount and quality of inquiry actually occurring fell well below expectations. While some people connected with the projects blamed the situation on teachers' lack of science knowledge, certain observations made by me and those working in my group made us think that this generally held explanation was too superficial.

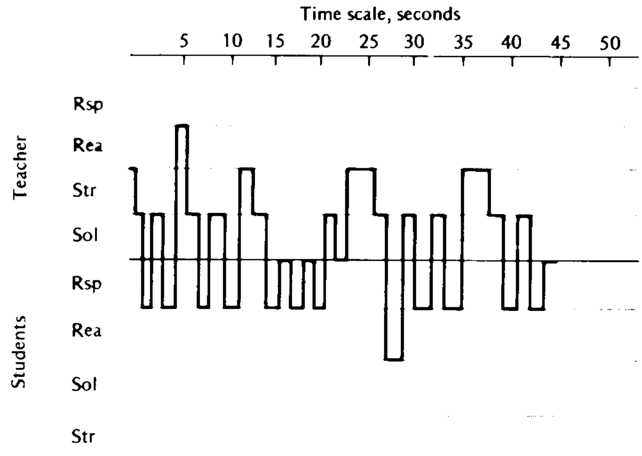
We found, for example, that children taught by teachers with considerable training in one of the programs did not exhibit substantially different patterns of inquiry from those taught by teachers with less exposure to the program. Neither were we able to distinguish different patterns of inquiry in one program as compared with another. With a few marked exceptions, which will be discussed shortly, the quality of discourse tended to stay at a low level and the pattern of interchange between teachers and children still more closely resembled an inquisition than a joint investigation or a reasonable conversation.

That we could further discount the "lack of knowledge" argument as a primary explanatory factor seemed to be demonstrated in data from two conferences funded by the National Science Foundation in which we had the opportunity to compare the instruction of children as carried out by a total of 54 scientists and science educators with instruction as conducted by a sample of classroom teachers. It was clear that some factors other than knowledge differences must be at work because the patterns of questions and responses were remarkably alike. To make that determination, we conceptualized the classroom as a two-player system consisting of a teacher and the collection of students (treated as the other player). Each player had four kinds of available verbal moves–structuring, soliciting, responding, and reacting or evaluating. By simply categorizing the sequence of moves and plotting them approximately on a time line, the patterns of interaction could be clearly displayed. In theory, to maximize outcomes there should be some equity in the number and distribution of type moves between players.[10] In practice three of the moves tended to be concentrated in the hands of the teacher while one of the moves, responding, fell largely to the student players. Figure 1(a), for example, shows what the "inquisition" look like. It is characterized by a rapid question-answer sequence with the solicitation coming usually from the teacher. Figure 1(b), on the other hand, shows

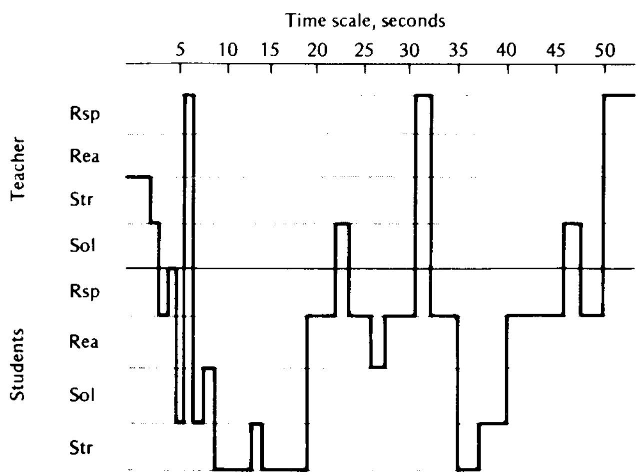
*The theoretical learning model that undergirds this research can be found in.[19]

A distinction could be made between what might be labeled as silent pauses and those which communication scientists refer to as filled pauses (words separated by ums, ahs, ers).[15] In these investigations filled and silent pauses were treated as equivalent.

In the two years which have passed since the original presentation of this data at the National Association for Research in Science Teaching annual meeting, 1972, two other researchers have reported work which complements the data presented here. At the NARST annual meeting in the spring of 1973, Campbell showed that lower I.Q. Junior High School students receive less wait-time and, indeed, the incidence of three-second pauses is generally very rare in junior high school science classrooms. Blosser trained people to ask questions and found that wait-time one pauses are between one and two seconds in duration.[17]



(a)



(b)

Fig. 1. (a) The "inquisition." The heavy line separates the two players. The moves, soliciting (SOL), structuring (STR), reacting (REA), and responding (RSP) are listed above and below the line according to their relative frequency of occurrence in the usual classroom pattern. Teachers do most of the structuring and soliciting and the students do the responding; (b) The "conversation." In this pattern both players engage in all of the kinds of moves. Students begin to suggest experiments (structuring) or they converse and react to each other's statements (responding and reacting). In contrast to the inquisition pattern, more of the weight of the moves falls below the center line and the over all pattern exhibits more variety.

what an inquiry pattern or a conversational pattern looks like. Both players employ all of the available moves.

In addition, we found that whatever pattern the teacher exhibited when working with four children closely resembled the pattern displayed when carrying on a discussion with a whole class. The fact that the form of the two-player graphs made from recordings of microteaching sessions that involved four students mirrored the form found for each teacher when working with a whole class suggested that size of groups could not be a major factor in determining whether an inquiry pattern would or would not develop. Thus we were left with none of the usual remedies offered for improving inquiry. Within rather broad limits we could not blame its low incidence on lack of materials since these had been supplied; on lack of science knowledge since professors and teachers generated similar player profiles; on group size since the profiles for small groups resembled profiles for large groups; on types of program since tape samples came from a variety of science curricula and lessons; on age of students since grades one through five were represented in the sample; or on pacing characteristics of various geographic areas since recordings included examples from many parts of the country.

Wait-Time, Rewards, and Expectations

After visiting and recording examples of science instruction carried out in classrooms located in suburban, urban, and rural areas, it finally became clear that while different curricula served as the vehicle of instruction, almost all of the discourse had one stable property. With the exception of a few teachers (three in two hundred recordings) the pace of instruction was very fast. Teachers allowed students an "average" time of only one second to start an answer to a question. If a response did not commence within one second, teachers usually either repeated the question or called on others to respond. When students did respond, teachers usually waited slightly less than a second (average of 0.9 seconds) before commenting on the response (reacting) asking another question (soliciting) or moving to a new topic (structuring).

In the few classrooms where the discourse was marked by the appearance in the speech of the children of speculation, sustained conversational sequences, alternative explanations, and arguments over the interpretation of data, the average wait-time hovered around three seconds. With reference to the two-player game model, there was also greater equity in the distribution of moves made by the two players. It appeared that more of the desired inquiry behaviors occurred in classrooms where teachers had longer wait-times.

One other observation connected with wait-time merits attention. We asked each teacher to designate the five best and five poorest students. When we examined the amount of wait-time given, on the average, to each group we found that the top five got nearly two seconds to begin an answer while the bottom five got slightly less than one second (0.9 seconds).

This last small piece of inquiry repeated many times since, alerted us to another variable, reward frequency. The amount of sanctioning behavior directed toward the two groups differed. The pupils ranked at the bottom actually received more overt verbal praise than did those ranked at the top, but it was difficult to know with certainty what was being rewarded. Top ranked pupils received relatively less evaluative comment from their teachers but the rewards were usually more pertinent to the responses made. Those at the bottom gathered more praise but its intent was far more ambiguous. It appeared that teachers rewarded top groups for correct responses but they rewarded the bottom groups for both correct and incorrect responses. Presumably the intent of some of this reward behavior must be to encourage effort.

We surmised that a clear teacher expectation pattern develops early in the history of each classroom. Differences in the wait-time and reward patterns administered to children ranked at the top as compared with those at the bottom suggest that teachers unconsciously acted in such a way as to confirm their expectations. (Expectations and rewards will be discussed in more detail in Part Two of this paper, especially their postulated relationship to the development of a sense of fate control.) So we came to concentrate attention on two input variables, wait-time and rewards. Figure 2 shows the heuristic which governed the research. Analysis of tape transcripts led us to characterize the usual classroom as one in which exchange rates between players is rapid, sanctioning rates of teachers are high, and teachers constantly mimic student responses, i.e., repeat portions of student speech as later can be seen illustrated in Figure 3.

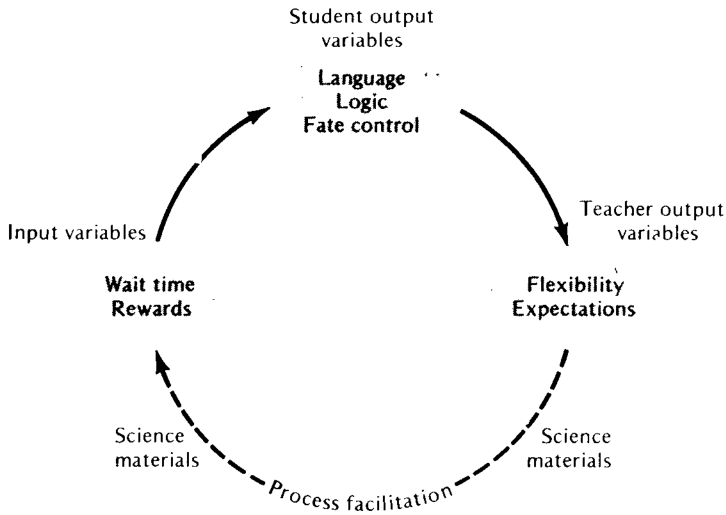


Fig. 2. Postulated relationships among variables.

Item	Repeat	Reward		Dialogue	Time	STUDENT Str	STUDENT Rea	STUDENT Sol	STUDENT Rsp	Wait time	TEACHER Sol	TEACHER Str	TEACHER Rea	TEACHER Rsp	
1			C1	Ice is melting.				1			1		1		Rsp
2	1		T	The ice is melting. What do we mean by melting?							1		1		Rea/Sol
3			C1,C2	Getting smaller				1							Rsp
4	1	1	T	Getting smaller, very good.									2		Rea/Rea
5			C2	Shrinking				1							Rsp
6	1		T	Shrinking									1		Rea
7			C2	The ice is reducing.				1							Rsp
8	2	1	T	The ice is reducing, that's a good word, reducing.									3		Rea/Rea/Rea
9			C1	And water is dripping and the ice. (Inaudible.)	15 sec			1		1+					Rsp
10			C2	Oh, I know what happens. Well, when the ice is dripping it causes water vapor.				1							Rsp
11	2	2	T	Water vapor, good, that's very, very good. Water vapor, what else? You both are looking at the hot bottle, the one with the warm water in it. What about . . . (interrupted.)	30 sec / 30 sec						2	1	3		Rea/Rea/Rea/Sol/Str/Sol
12			C1,C2	Hot air, hot air, hot air.				1							Rsp
13	1		T	Hot air. Why is there hot air?							1		1		Rea/So
14			C1	Because there is no space for the new clean air to get in.				1							Rsp
15			C2	No room for oxygen.				1							Rsp
16	1	1	T	For oxygen, that's right. You see the hot air here. (Points to A.) What about here? (Points to B.)							1	1	2		Rea/Rea/Str/Sol
17			C2	You don't see any	45 sec			1							Rsp
TOTALS	9	5		45 sec of transcript; average wait time = 0.4 sec		0	0	0	10	1 sec	5	2	13	0	
						Student moves = 10					Teacher moves = 20				Total moves – 30

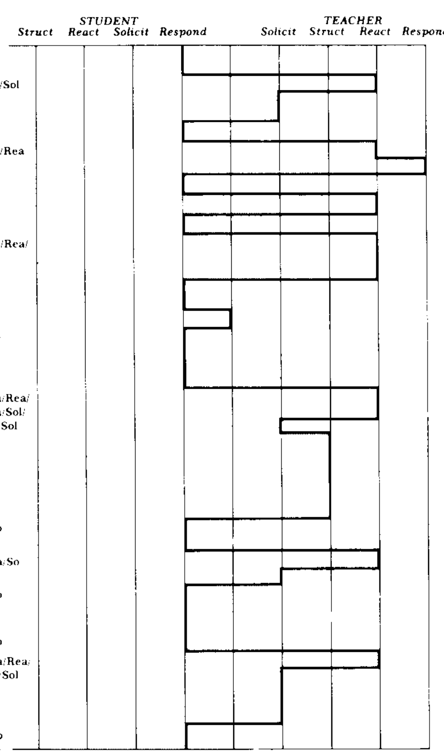


Fig. 3. A typical analysis of a transcript illustrates the high incidence of mimicry and rewarding. Experiment: observing changes in a thermal system.

Detection and Measurement of Wait-Time

Measurement of wait-time is based on the two-player model. As shown in Figure 4 two distinct species are identified.

a. "*Wait-time of the species one type*" may appear in two varieties. Normally it begins when the teacher stops speaking and terminates when a student responds or the teacher speaks again. If, as sometimes happens, a teacher asks a question, pauses, calls on a student, and pauses again, the two pauses are summed. Together they constitute an instance of the first species of wait-time.

b. "*Wait-time of the species two variety*" is calculated by taking the sum of all pauses occurring on the student player side and terminates when the teacher speaks. The more common varieties include the case where a student speaks, stops, and the teacher speaks again; a student speaks, pauses, speaks again, and the teacher rejoins the play. (Note: Here the term student is used generically, i.e., refers to the two-player model. The pauses may occur within the speech of a single pupil or they may occur between the speech of a succession of pupils. In either case, the collection of pauses are summed and constitute a single instance of the post-student response wait-time.) Correlations between the two species of wait-time vary somewhat but tend to be on the order of 0.17.

Computation of Mean Wait-Time

How mean wait-time is calculated depends on the purpose to be served.

a. An unweighted mean is calculated. The sum of seconds for all pauses is divided by the number of between player exchanges. This method is the one most easily understood by teachers. Measurements were made to the nearest half second. (In practice, however, this is more refinement than is necessary for general use.)

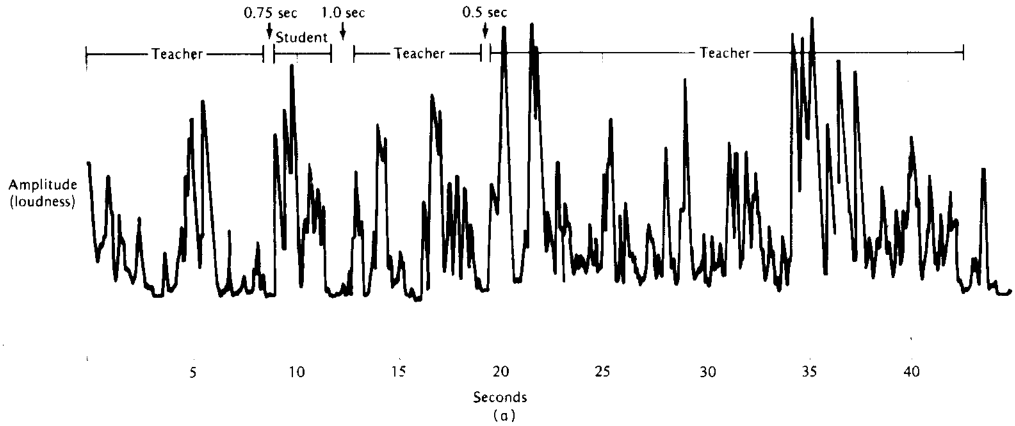
b. A weighted mean is calculated. The sum of seconds for each species of pause is divided by the total number of between-player exchanges for that species. This has the effect of weighting each specie according to its frequency of occurrence. The mean of these two means is the average wait-time.

Servo-Chart Plots. It soon became clear that original estimates of wait-time were too high. By the time the stop-watch was punched the discourse had moved on. We needed to find another way to measure wait-time that would not be hampered by the fact that reaction times, the motor responses necessary to actuate the clock, are very slow in comparison to mental responses. This problem was partially solved by delivering the sound from tape recordings into a servo-chart plotter. With a diode inserted between the tape recorder and the plotter, the needle could be made to track horizontally when there were silences. The paper for the plotter is calibrated. By running the plotter as fast as it will go, each calibrated interval equals one second. Figures 5 (a) and (b) show examples of plots. The horizontal axis is a time axis. The height of the peaks represents the variations in amplitude of the sound generated.

When peaks cluster closely words are being generated rapidly. Students, especially those below fourth grade, frequently speak at a slower rate than do their teachers. This will be reflected in the plots by the width of the curves somewhat below the peaks and by the distance the needle tracks horizontally along the time axis. Longer pauses produce horizontal tracks.

	Specie #1		Specie #2	
	P		P	
	A	TEACHER	A	TEACHER
TEACHER	U		U	
	S	STUDENT	S	STUDENT
	E		E	

Fig. 4. The location of potential pauses in a sequence of moves. As specie #2 increases in length, the probability of S-S-S-S sequences increases.



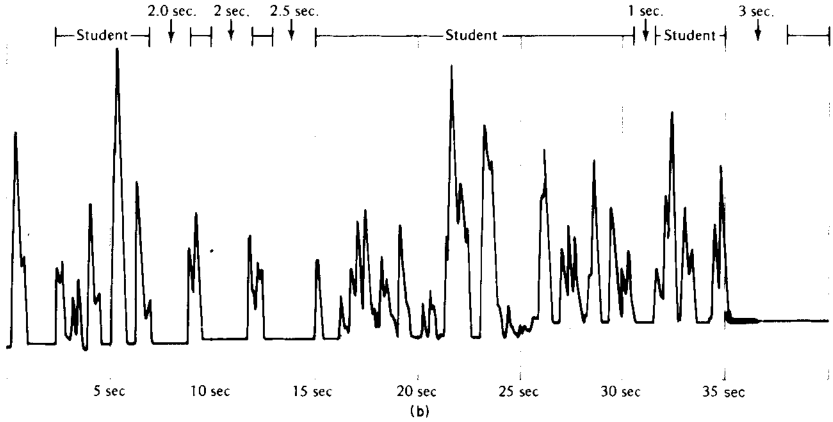


Fig. 5. (a) Servo-chart plot of discourse in an average classroom shows how rapid exchanges are, and the needle tracks horizontally when there are pauses; (b) Servo-chart plot of student speech when an explanation is being attempted. Notice the pauses in the body of the discourse. When the teacher reacts too quickly one only hears a fragment of an idea.

Examination of chart plots made when students are trying to explain some phenomenon suggests the part that species two, the post-student response wait-time plays in determining the quality of explanation. As Figure 5 (b) illustrates, when uninterrupted, student talk tends to come in bursts separated by pauses that often equal or exceed three seconds. Apparently the reason for the high incidence of phrase-line rather than propositional responding that marks classrooms with short average wait-times is brought about by the intervention of the teacher between the speech bursts. When this wait-time is increased, the mean length of student response increases and concomitantly the probability that inferences will be connected to evidence increases. It is as though the mapping of experience and thought into language proceeds in pieces.[1] Intrusion between the bursts by another player prevents the expression of a complete sequence. The servo-chart plots shown in Figure 5 (a) is typical of exchange rates prior to training on wait-time. As indicated by the horizontal tracking, the pauses are few, the duration of student speech is brief. Students as well as teachers frequently interrupt each other, the rate of interchange being so fast that the plotter pen often does not reach the base line. (Also see Figure 3.)

Types of Wait-Time Investigations Pursued

Over the last six years the investigations of wait-time have been quite varied in form as befits applied research where the primary goal is to produce desirable outcomes in operating systems.

"In vitro studies." The wait-time variable was identified first through regular observation of 36 primary grade classrooms in six schools in which the Science Curriculum Improvement Study (SCIS) was being taught. Six tape recordings were made in each of the rooms during the year (a total of 103 tapes). Once the fact of short wait-times, high question frequency[4] and rewards was recognized it became a matter of interest to determine how general the short pause phenomenon was. It could be the case, after all, that something about SCIS promoted the inquisitional pattern or that the pattern was unique to primary grades or that speech in the region of the six schools (New York and New Jersey) was always fast-paced. In the following year 84 tapes were made in classrooms scattered around the country where SCIS as well as other science curricula were being taught. Wait-time typically fell below three seconds. Another sample of 34 tapes made of fourth grade science classrooms of different kinds showed mean wait-time of 1.3 seconds. In addition, there gradually accumulated a miscellaneous collection of tapes, in excess of one hundred, sent by teachers, project directors, or collected by staff members on visits to different parts of the country. We also acquired 22 high school and college seminar tapes in which wait-times ranged from 1 second to 2.8 seconds with a mean of 1.8 seconds. (Borg has obtained similar results for high school populations.[2]) And 73 recordings made in Louisiana, Alabama, and Tennessee yielded similar results. Thus it seemed safe to infer that short wait-times were not localized in first and second grade classrooms, nor were they specific to a curriculum or geographic area.

"Microstudies." In order to study the influence of prolonged wait-times which did not seem to be occurring with sufficient frequency in natural settings, a series of microstudies was begun in which duration of pauses at two locations was systematically varied.

a. The staff worked with small groups of students, finally settling on four as producing a reasonable facsimile of a classroom. Here the effort was to increase wait-time one and two separately and then jointly.

Both audio and television tapes were used. Students came from different grade levels and the lessons were selected from various curricula. In these investigations there was an attempt to identify the relative influence of the two species of wait-time, the pause occurring after a teacher move and the pause following a student move. We attempted to manipulate these pauses independently. Thus, for example the same lesson might be taught to different groups of children but the sequence of treatments would be as follows.

Treatment I	Standard Wait-Time (1 second)
Treatment II	Wait-Time #1 long, #2 short
Treatment III	Wait-Time #1 short, #2 long
Treatment IV	Both wait-times long (3 seconds)

The reward schedule (overt verbal rewards) would be set at as near neutral as possible.

b. A pool of six lessons was prepared. Ninty-six teachers engaged in a series of teach-twice cycles designed to get them to produce criterion wait-times of three seconds. They each taught the same four students in each cycle. This allowed a study of the cumulative impact of wait-time on intact groups.

In order to minimize the influence of sequence of lessons, the six lessons were grouped into three sequences. One third of the group did each sequence. This decision would make it safer to infer that differences on the outcome variables, if there were any, could be attributed to the influence of the wait-time changes rather than to characteristics of content. The treatment sequence went as follows.

1. Base line tape. No instruction on wait-time prior to taping. Portions of the tape were transcribed by the teacher.

2. Tape 2. Wait-times and outcomes variables discussed following the taping. Portions of the tape were transcribed by the teacher and the wait-times were measured.

3. Tape 3. The wait-times were measured and techniques for getting control of both wait-times discussed. Teachers were asked to drop mimicry. (See Figure 3.)

4. Tapes 4 through 6. Analyzed and discussed in the same way. A subset of each group instructed to drop overt verbal rewards.

All tapes were transcribed and coded according to the two-player model. Wait-times were measured, the outcome variables identified, and their values determined. Criterion wait-time was set at three seconds or longer. This was an empirically based decision stemming from three observations: (1) the few classrooms which exhibited acceptable values on the outcome variables averaged about three seconds; (2) student pauses between speech bursts often equaled or exceeded three seconds; (3) microstudies showed detectable changes in the outcome variables beginning at a threshhold of approximately three seconds. Table I contrasts values of 5 outcome variables for tapes 1 and 6. Twenty percent of this particular group failed to achieve criterion.

Return to the Classroom. Twelve teachers with criterion wait-times in the teach-twice cycles were studied and given help in the classroom. For a period of one year, observations and tape recordings were made at approximately two-week intervals (once a week for the first four weeks and then at longer intervals). In addition four other teachers elected to study the influence of wait-time in their own classes and to supply us with tapes and transcripts. A total of 74 tapes were accumulated in this phase. Figure 6 shows shifts in values on some of the outcome variables for one classroom. The pattern was fairly typical of the 12 classrooms.

Similar shifts occurred for 76 intact four-person microgroups which achieved criterion wait-times.

Recently Garigliano,[3] who was investigating the influence of wait-time on five student outcome variables, taped both classroom and microsessions taught by the same teachers. He found mean wait-times to be on the order of one second. With the limited training he was able to give teachers, he was unable to raise mean wait-time to a three-second criterion. Although the difference in mean wait-time between experimental and controls was signficant, functionally the difference meant little.

Student and Teacher Outcome Variables

Ten student outcome variables and three teacher outcome variables were identified and their relationship to wait-time investigated.

A. Student Outcome Variables

1. *The Length of Student Responses Increased.* Under a fast schedule, responses tend to consist of short phrases and rarely exhibit explanation of any complexity. Data from the chart plots suggest that the second wait-time, when it is prolonged, contributes measurably to the appearance of longer statements. The average shift was from seven words (S.D. = 3) to 28 words (S.D. = 6).

2. *The Number of Unsolicited but Appropriate Student Responses Increased.* This outcome is more responsive to the second than the first wait-time, but is influenced by both. The average shift was from a mean of three (S.D. = 2) to a mean of 37 (S.D. =11).

3. *Failures to Respond Decreased.* "I don't know" or no responses were often as high as 30% in normal classrooms, i.e., in classrooms where the mean wait-time fell at one second or less. This outcome is more susceptible to manipulation of the first wait-time, the pause which the teacher allows before calling on another student or repeating a question. (It also happens to be responsive to reward incidence.)

TABLE 1

Student Outcome Variables: Contrasts Between Tape 1 and Tape 6 of the Training Sequence for 76 out of 95 Teachers Who Achieved Criterion Wait-Times of Three Seconds or Longer Prior to Tape 6

Student Variables		Tape 1	Tape 6
Mean length of response		8	27
	range	(3-12)	(14-39 words)
Mean number of unsolicited but appropriate responses		5	17
	range	(0-17)	(12-28)
Mean number of failures to respond		7	1
	range	(1-15)	(0-3)
Number of evidence-inference statements		6	14
	range	(0-11)	(6-21)
Number of solicitation, structuring, and reacting moves		5	32
	range	(1-6)	(11-46)

15 minutes of tapescript; Tape 1: mean wait time in this sample was 1.2 seconds; range 0.8–2.4 seconds. Tape 2: mean wait time = 3.3 seconds; range = 3.0–5.6 seconds.

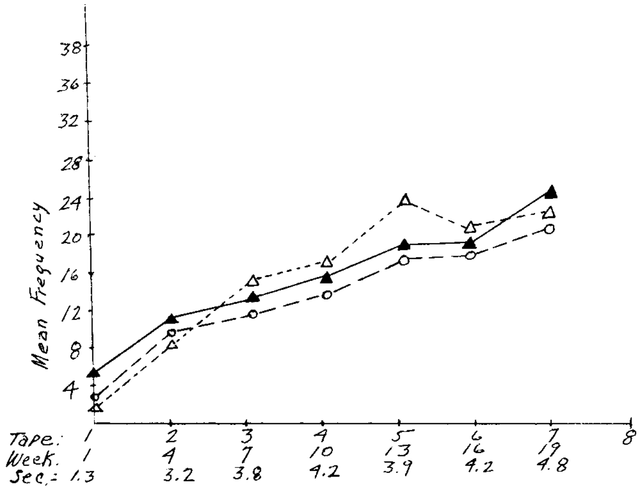


Fig. 6. Typical shifts in values of three student outcome variables for a class on a criterion wait-time schedule (third grade): (▲) mean number of solicitations, structuring, and reacting moves; (△) mean number of evidence-inference statements; (○) mean length of response.

4. *Confidence as Reflected in Fewer Inflected Responses Increased.* Under a fast schedule, responses tend to be phrased as though the child were saying, "Is that what you want?" (In the middle of a prolonged fast sequence one can ask a child for any fact which he knows well, and it will not be unusual to have him respond with a question mark in his tone.) This confidence indicator, inflected responding, is also susceptible to the reward variable. As reward increases so does the incidence of inflected responses.

5. *The Incidence of Speculative Thinking Increased.* This is influenced by both species of wait-times. The average shift was from a mean of two events to a mean of eleven events.

6. *Teacher-Centered Show and Tell Decreases and Student-Student Comparing Increases.* Under a fast schedule and a high reward or sanctioning schedule, children "stack up" waiting to tell the teacher. There is very little indication that they listen to each other. This variable is not examined in detail in this paper since it seems to be influenced as much by the reward pattern as by the pacing.

7. *More Evidence Followed by or Preceded by Inference Statements Occurred.* Under a fast schedule, the incidence of qualified inferences is extremely low. When wait-time is lengthened, this outcome variable changes in a desired direction. It is more susceptable to the species two pause.

8. *The Number of Questions asked by Children Increased and the Number of Experiments They Proposed Increased, i.e., the Number of Structuring Moves Increased.* It is a well established fact from classroom interaction studies that students do not ask questions very often. When they do, the questions are usually for clarification of procedures and are rarely ever directed at other students. This outcome variable seems to be susceptible to both classes of wait-times. Structuring and soliciting moves by students shifted from a mean of four (S.D. = 3) to a mean of 18 (S.D. = 5).

9. *Slow Student Contributions Increased.* Under a fast schedule most responses came from a particular faction of the class. When wait-times were increased, the sources of response became more varied. (Interestingly, this outcome appears apparently to influence teacher expectations. Although we have not had time to investigate it in any detail yet, it seems to be both surprising and rewarding to the teachers that students who do not usually respond as readily begin to do so. The possible indirect influence of wait-time on teacher expectations needs systematic investigation.)

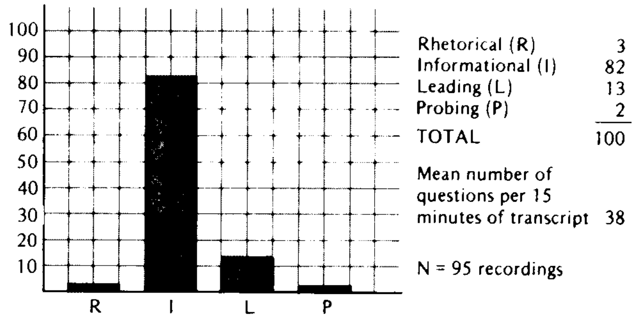
10. *The Variety in Type Moves made by Students Increases.* Structuring, soliciting, and reacting moves increase. Reacting moves increased from a mean of four to a mean of nine.

B. Teacher Outcome Variables

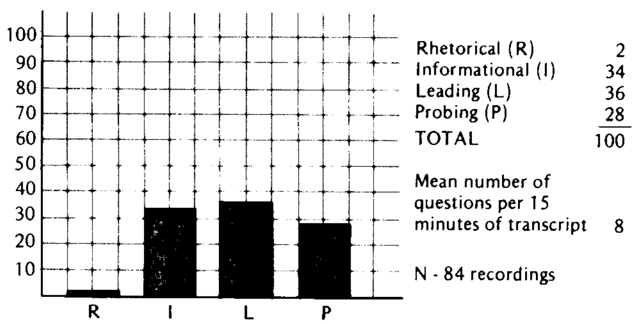
Once wait-time is changed and the behavior is stabilized for a period, certain characteristics of teacher input variables change. They are regarded here as outcome variables because they seem to be influenced by the wait-time factor.

1. *Teachers Exhibit Greater Response Flexibility as Indicated by the Occurrence of Fewer Discourse Errors.* Under a rapid schedule, the normal situation obtaining in classrooms, the probability that a detectable discontinuity in the discourse occurs increases. Conversation does not build into structural propositions. Instead the sequence of discourse resembles a smorgasbord in which everyone goes along commenting on what he passes and picks up but nobody pays any attention to or gives any indication that he has heard the comments of others. If a teaching machine were to ask a question and a student responded with something that was not in storage, the machine would either continue to the next question as though nothing happened or it would cycle back and repeat, perhaps with progressive cueing. In either case a discontinuity is scored against the discourse. Our tapes suggest that frequently the teacher on a fast schedule achieves a less favorable flexibility score than does a moderately good computer program. At least the computer program has the advantage of leaving the response time up to the student. The flexibility score increases with increases in wait-time. It is computed by simply counting the mismatches between a student statement and a teacher response or reaction.

2. *Number and Kind of Teacher Questions Change.* a. The total number of questions decreases per a 15 minute interval. Prior to wait-time training it was not unusual to find as many as seven to ten questions per minute asked by a teacher. The mean number of questions averaged between two and three per minute. (Inner city rates tend to be slightly higher than suburban rates, 3.4 questions per minute as compared with 2.8. Samples of tape recordings made in the Cumberland Mountains in Tennessee, in Louisiana, and in North Carolina show mean rates of approximately 2.2. questions per minute.) As wait-times increase the rate of questioning drops. For teachers who have achieved criterion wait-times of three seconds or longer mean question rates tend to approach 0.4 per minute. This follows from the fact that student responses become



(a)



(b)

Fig. 7. (a) Typical distribution of question types asked by teachers prior to wait-time training; tapes #1 and #2; (b) Typical change in distribution of question types once criterion wait times of three seconds or more are attained and sustained; tapes #5 and #6.

longer; unsolicited student responses increase; there are more pauses between speakers as well as within the speech of speakers.

b. The net variability in teacher questions increases as teachers achieve criterion wait-time of three seconds. Figures 7(a) and (b) show how the pattern of questions changed for a sample of 74 teachers who achieved criterion wait-times of three seconds or longer.

longer; unsolicited student responses increase; there are more pauses between speakers as well as within the speech of speakers.

Teacher questions were categorized according to two systems, one modified from Ashner and Gallagher and the system developed by T. W. Parsons (1971). The Parsons categories are more coarse and so proved more practical since they could be correctly identified more easily by teachers with less training. Coding required less interpretation of intent. Intercoder agreement for Ashner and Gallagher varied from 72 to 84%. For Parsons' system agreement varied from 76 to 94%. The primary objective was not to study questions as an input variable as is the usual procedure, but rather to regard teacher questions as an outcome variable. We were asking whether the pattern of question asking spontaneously changes as a result of increases in wait-time. Figure 7(a) and (b) show the results for Parsons' category system. The pattern of questioning seems to be responsive to changes in wait-time. This is not to suggest, however, that giving explicit training in how to ask questions is not desirable. But it might not be totally surprising to find that people who receive such training are inadvertently slowed down, i.e., have longer wait-time.[3b,7a]

3. *Teacher Expectations for Performance of Certain Children Seem to Change.* The fact that an initial sample of 26 teachers who identified their five best and five poorest students gave the best students more time to reply to questions leads one to suspect that the relationship of wait-time to expectations should be investigated in more detail. At the moment all we have accumulated over the last five years are unsolicited comments from teachers on longer wait-time schedules that members of the bottom group perform in new and surprising ways. That teachers may modify expectations gradually is indicated by comments about a student such as, "He has not done anything like that before. Maybe he has a special aptitude for science." It may be useful to study the influence of wait-time and rewards on teacher expectations in a more systematic way. We have not had an opportunity, however, to discover whether the effect on expectations is general or

how long it persists. Neither do we know how it may change the real performance of the students rated at the bottom of the class, given that the pattern of responding could be sustained. The matter will be discussed again in Part Two of this paper.

Summary

In Part One of this two part presentation the relation of wait-time to ten outcome variables is discussed. Wait-time at two locations was studied in natural settings and in experimental settings to determine how location and duration of pauses influence the outcome variables. It was noted, in addition, that there is an interaction between wait-time and rewards and that students rated at the top or at the bottom of a class receive differential treatment with respect to these variables. A heuristic which relates wait-time and rewards to language, logic, and fate control is introduced. It will be discussed in more detail in Part Two of this paper. As far as can be determined this is the first series of studies in which pauses in two locations have systematically been varied to determine how language and logic are affected. The work is based on the view that pausing and complex cognitive processes may be related in the context of inquiry.[1] The theoretical implication of pause location and the functional significance of pause duration must be considered in terms of cognitive, affective, and social interaction variables. The implications for affective and social interaction variables will be discussed in Part Two of this paper.

It would also be of some interest to determine how the quality of discussion in student groups would be altered as a result of giving wait-time and listening training to students.

Work described in this paper has been supported by grants from the Alcoa Foundation, International Nickel Company of Canada, Shell Companies Foundation, The Xerox Corporation, The Mary Duke Biddle Foundation, and the Hebrew Technical Institute. A post-doctoral year at New York University (USOE) provided an opportunity to draw the data together.

Much credit is due Dr. Francis X. Lawlor who has taken part in these studies since their inception more than six years ago. Graduate assistants and classroom teachers who participated must also be credited. This kind of research is something done with people, not something done to them.

References

1. Aschner, M. J. and J. J. Gallagher, "Aschner-Gallagher System," In *Mirrors for Behavior*, II, vol. A, A. Simon and E. G. Boyer, Eds., "Classroom Interaction Newsletter," Research for Better Schools, Inc., 1970.

2. Borg, Walter, "The Minicourse as a Vehicle for Changing Behavior, The Research Evidence," Far West Laboratory for Educational Research and Development, 1969. (Borg worked with older students but found pauses to be below a 3-second criterion.)

3. Garigliano, L. J., "The Relation of Wait-Time to Student Behavior in Science Curriculum Improvement Study Lessons," unpublished Ed. D. dissertation, Teachers College, Columbia University, New York, 1972.

4. Kondo, Alan K., "A Study of the Questioning Behavior of Teachers in the Science Curriculum Improvement Study Teaching the Unit on Material Objects," unpublished Ed. D. dissertation, Teachers College, Columbia University, New York, 1968.

5. Lawlor, Francis X., "The Effects of Verbal Reward on the Behavior of Children in the Primary Grades at a Cognitive Task Typical of the New Elementary Science Curricula," *Journal of Research in Science Teaching*, 7, 327, 1970.

6. Lefcourt, Herbert., "Internal Versus External Control of Reinforcement: A Review," *Psychological Bulletin*, **65**, 206, 1966.

7. Locke, E. A., "Toward a Theory of Task Motivation and Incentives," *Organizational Behavior and Human Performance*, **3**, 157, 1968.

8. Ogunyemi, Ebenezer Olanrewaju, "The Effects of Different Sources of Verbalized Information on Performance at a Science Related Cognitive Task," Doctor's thesis, Teachers College, Columbia University, New York, 1970.

9. Parsons, Theodore W., *Guided Self-Analysis System for Professional Development Education Series*, Berkeley, California, 1971.

10. Pritchard, R. D., "Equity Theory: A Review and Critique," *Organizational Behavior and Human Performance*, **4**, 197, 1969.

11. Rotter, Julian B., "Generalized Expectancies for Internal Versus External Control of Reinforcement," Psychological Monographs: General and Applied, Whole No. 609, **80**, No. 1 (1966).

12. Schwab, D. P. and L. L. Cummings, "Theories of Performance and Satisfaction: A Review," *Industrial Relations*, **9**, 408, 1970.

13. Sieber, Joan E., "Overcoming Secondary Ignorance: Learning To Be Uncertain," *ERIC Edrs.*, Ed #050-035, 1971.

14. Wiethake, Henry., "The Use of Group Dynamics Training to Foster Verbal Interaction in the Elementary Science Classroom," paper presented at the National Science Teachers Association Annual meeting, Washington, D. C., 1971.

15. Rochester, S. R., "The Significance of Pauses in Spontaneous Speech," *Journal of Psycholinguistic Research,* 2, No. 1, 1973.

16. Rowe, M. B., *Teaching Science as Continuous Inquiry*. McGraw-Hill, New York, 1973.

17. Blosser, P. E., "The Development of the Skill of Questioning by Prospective Secondary School Science Teachers: An Extension," paper presented at the National Association for Research in Science Teaching Annual Meeting, Chicago, 1973.

18. Campbell, J. R., "Pattern Analysis—A Macroscopic Development for Interaction Analysis," paper presented at the National Association for Research in Science Teaching Annual Meeting, Chicago, 1973.

19. M. R. Rowe, "Conflict and Inquiry," in *Teaching Science as Continuous Inquiry*, McGraw-Hill, New York, 1973, ch. 7, also see, "Language in Use," ch. 12.

JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 40, SUPPLEMENT, PP. S33–S50 (2003)

ENCOURAGING THE TRANSITION FROM CONCRETE TO FORMAL COGNITIVE FUNCTIONING–AN EXPERIMENT

ANTON E. LAWSON and WARREN T. WOLLMAN

AESOP, Lawrence Hall of Science,
University of California, Berkeley, California 94720

Introduction

Flavell (1963) describes Piaget's stage of formal thought as:

> . . . not so much this or that specific behavior as it is a generalized *orientation*, sometimes explicit and sometimes implicit, towards problem solving: an orientation towards organizing data (combinatorial analysis), towards isolation and control of variables, towards the hypothetical, and towards logical justification and proof. (p. 211)

Encouraging this type of orientation towards problem solving is of utmost importance to educators. As stated by the Educational Policies Commission (1961), the central purpose of American education is the development of problem-solving processes called rational powers.* The parallel between the rational powers and the problem-solving processes utilized by formal thinkers is clear.**

One must not be misled to interpret Piaget's theory as implying that maturation of the nervous system is sufficient for the development of formal thought. If this were the case, the job of our educational system would be small indeed. Rather, maturation determines only the totality of possibilities and impossibilities at a given stage. As Inhelder and Piaget (1958) state:

> A particular social environment remains indispensable for the realization of these possibilities. It follows that their realization can be accelerated or retarded as a function of cultural and educational conditions. (p. 237)

The finding that formal thought is normally demonstrated by only about 50% of the subjects in most adolescent and even adult samples which have been studied in this country suggests a very real educational problem (e.g., Friot, 1970; Higgens-Trenk & Gaite, 1971; Kohlberg & Gilligan, 1971; Lawson & Renner, 1974; McKinnon & Renner, 1971; Wollman & Karplus, 1974).

Little is presently known about specific factors and how they interact to contribute to and affect the transition from concrete to formal cognitive functioning. Piagetian theory itself only provides a general view. At best, Piaget's factors of neurological development, social transmission, experience with things, and internal cognitive reorganization (equilibration) are

*The rational powers are listed as the processes of recalling and imagining, classifying and generalizing, comparing and evaluating, analyzing and synthesizing, and deducing and inferring.

**For an in-depth discussion of intellectual and moral development as the aim of education, see Kohlberg and Mayer (1972).

Originally published in Volume 13, Number 5, pp. 413-430 (1976).

suggestive but not sufficient to provide sound instructional theory. How can these factors be interrelated to facilitate formal or abstract thought?

In a recent survey, Wollman (1975) observed that students as young as second-graders have at least an intuitive notion of the formal idea of controlling variables. Piaget and Inhelder (1958) claim this idea is fully developed only during adolescence. Ausubel (1964) suggests that it might be possible to utilize prior intuitive understanding to facilitate learning and stabilization of more formal (abstract) conceptualizations during adolescence. In the same vein, Bruner and Kenney (1970) argue that intellectual development begins with instrumental activity, a kind of development by doing. They suggest that these activities become represented and summarized by individuals in the form of particular images. Further, with the aid of symbolic notation that remains invariant across transformations in imagery, the learner gradually becomes aware of the formal or abstract properties of the things he is experiencing. These findings and statements suggest that it might be possible to design instructional sessions which will enable students who are in Piaget's stage of concrete operations to function at the formal-operational level with respect to specific concepts. However, as Ausubel (1964) states:

> If stages of development have any true meaning, the answer to this question [can training enable children to acquire genuine appreciation of concepts which are normally acquired only at advanced stages of development?] can only be that although some acceleration is possible it is necessarily limited in extent. (p. 264.)

Problem Statement

This investigation, then, addressed itself to the following questions. (1) Can instructional procedures incorporating the above ideas be designed and employed to successfully affect the transition from concrete to formal cognitive functioning in fifth and seventh-grade students with regard to one aspect of formal thought, i.e., the ability to isolate and control variables? (2) If training can enable concrete students to perform at a formal level on tasks which were used in the training phase, will this training transfer to tasks also involving the control of variables but using novel materials? This is termed specific transfer (Brainerd & Allen, 1971). (3) If training can enable concrete students to perform at a formal level on tasks requiring the control of variables, will this training transfer to tasks involving different concepts but ones which also involve formal thought (nonspecific transfer)? In other words, if the training was effective, was it limited to the specific concepts involved or did it affect a more general shift from concrete to formal cognitive functioning? (4) What is the relationship between a student's level of intellectual development and his ability to profit from the training?

Method

Subjects

Thirty-two fifth-grade students and 32 seventh-grade students enrolled in an elementary school and a junior high school in Lafayette, California, served as subjects (Ss). The fifth-grade Ss (14 males and 18 females) ranged in age from 9.5 years to 12.1 years; mean age = 10.5 years. No IQ data was available for these Ss. The seventh-grade students (16 males and 16 females) ranged in age from 11.9 years to 13.6 years; mean age = 12.6 years. Only Ss whose IQ's were below 115 were chosen for the study. The IQ's of those selected ranged from 100-115; mean IQ = 109. Lafayette is an upper-middle-class suburban community in the San Francisco Bay region.

Many studies, such as those mentioned previously, suggest that formal thought often does not develop during the general age guidelines of 11-15 years given by Piaget. It was, nevertheless, decided to work with students slightly younger or slightly older than the figure suggested as the age of onset of formal thought. If instructional procedures can successfully affect a shift from concrete- to formal-operational levels of thinking, the ages of 10 to 13 may be the optimum time for such instruction.

Experimental Design and Procedure

The Fifth-Grade Study

The experimental design employed in this investigation is referred to by Campbell and Stanley (1966) as the pretest-posttest control-group design. In effect, two separate investigations were conducted. In the first experiment, the 32 fifth-grade Ss were randomly placed into two groups of 16 Ss each–an experimental group which received training concerning the concept of controlling variables, and a control group which did not receive training. Both groups were pretested in individual interviews with a battery of Piagetian tasks (conservation of weight, conservation of volume, and volume displacement) which allowed classification of Ss into early concrete-IIA, late concrete-IIB, postconcrete, and early formal-IIIA levels of intellectual development. The experimental-group Ss then participated in four sessions of individual training. Each session lasted about 30 minutes. The control-group Ss attended their regularly scheduled classes during this time. The training sessions, which will be described in more detail below, involved the presentation of problems involving the determination of cause-and-effect relationships. The first session involved bouncing tennis balls, the second session involved bending rods (materials used during the posttest), the third session involved an apparatus called a "Whirly Bird" (SCIS, 1970), and the fourth session involved two biology experiments presented in a pencil-and-paper format.

Posttesting of all 32 Ss followed the training sessions and was conducted in two phases. The first phase consisted of individual interviews conducted by two trained examiners who had no prior knowledge of which Ss were members of the experimental and control groups. A male examiner interviewed the male Ss and a female examiner interviewed the female Ss.* Three Piagetian manipulative tasks (bending rods, the pendulum, and the balance beam) were administered. Each task allowed classification of Ss into concrete- or formal-operational levels of intellectual development. The bending rods task was used to determine whether or not the training was effective in facilitating the ability to control variables with materials identical to those used during the training. The pendulum task was used to determine whether or not the training was generalizable to a problem also involving controlling variables but using novel materials (specific transfer). The balance beam task was also used to determine the extent to which the training encouraged formal thinking. This task, however, did not involve controlling variables; rather, it involved proportional reasoning. Piaget claims that this reasoning ability develops concomitantly with the ability to control variables. This task, therefore, could be considered as a measure of nonspecific transfer of training. In addition to the three Piagetian tasks, students responded orally to a written question involving value judgment adapted from Peel (1971). Responses were tape recorded and later classified into developmental categories.

*Research such as that conducted by Brekke and Williams (1973) has shown that male Ss perform better on Piagetian tasks when the examiner is a male; likewise female Ss perform better when the examiner is a female.

During the second phase of the posttesting, all 32 Ss were grouped together and two pencil-and-paper examinations were administered. Ss responded to a spheres task involving the control of variables, a logic question involving the logical fallacy known as affirming the consequent, and one combinatorial question. These problems, like the balance beam task and the Peel question, were administered to determine the extent to which the training encouraged a general shift from concrete- to formal-operational thinking rather than a specific shift limited to the control of variables. The spheres task, like the pendulum task, tested for specific transfer of the trained concept. It was, however, an additional step removed from the training in that it did not involve the manipulation of materials.

The Seventh-Grade Study

The experimental design employed and procedures described above were repeated using the sample of seventh-grade Ss. The only changes made were on the posttests. It was decided to use a shortened version of the Longeot examination (Longeot, 1962, 1965) that incorporated problems of class logic, propositional logic, proportional reasoning, and combinatorial reasoning as the measure of nonspecific transfer of training. The Longeot examination has been shown to be an effective instrument to measure general levels of concrete- and formal-operational thought (Lawson & Blake, 1976; Sheehan, 1970)

Pretests

Four Piagetian-styled tasks were administered to experimental and control Ss in individual interviews. All questions on the tasks were asked in a counterbalanced order. Following all responses, Ss were asked why they responded as they did. Since each task has been employed by previous investigators, only brief descriptions of the tasks and materials used are included.

Conservation of weight (Elkind, 1961). Two balls of clay of approximately 50 grams each were presented to S. One ball was then transformed into a pancake shape and S was asked about the relative weights of the balls.

Conservation of volume using clay (Elkind, 1961). The two balls of clay from the previous task were used. S agreed that two beakers (400 ml) contained the same amount of water and was asked about the relative amount of water displaced by the two pieces of clay.

Displacement volume (Karplus & Lavatelli, 1969). Two metal cylinders of equal volume but different weight (18 g and 55 g) were handed to S. The equal height and thickness of the metal cylinders were pointed out. The examiner then took the cylinders and lowered the lighter one into one of two test tubes (30 ml) which were partially filled with equal amounts of water. The rise in water level was noted and S was asked to predict the rise in water level when the heavier cylinder was lowered into the other test tube.

On the basis of responses given to these three tasks, Ss were classified into developmental levels as follows:

Concrete-IIA Nonconservation responses on all three tasks.
Concrete-IIB Conservation of weight and nonconservation of volume and volume displacement.
Postconcrete Conservation of weight and conservation of volume or volume displacement.
Formal-IIIA Conservation responses on all three tasks.

Bending rods (Inhelder & Piaget, 1958). This task tested S's ability to identify and control variables. Given six flexible metal rods of varying length, diameter, shape, and materials which were fastened to a stationary block of wood, S was asked to identify variables and demonstrate proof of the affect of each variable on the amount of bending of the rods.*

Reliability of the conservation of volume using clay task and the displacement volume task, as well as the pendulum and balance beam tasks used on the posttests, have been established by Lawson, Nordland, and DeVito (1974). The tasks, with the exception of the conservation of volume using clay task, were found to have acceptable reliability. Test-retest correlation coefficients ranged between .48 and .78. Lawson, Nordland, and Kahle (1975) found the reliability of a series of Piagetian tasks, including those used in this study, to be high. Cronbach's Alpha coefficient, a modification of the KR-20 formula for scalable items, was .86. Validity of the Piagetian tasks has been determined by numerous investigators (e.g., Goldschmid & Bentler, 1968; Lawson & Blake, 1976; Lawson & Renner, 1975; Lawson, Nordland, & DeVito, 1975; Lawson, Nordland, & Kahle, 1975).

Training

Each S in the experimental group met with an experimenter (E) for four 30-minute individual training sessions. The sessions were conducted over a period of approximately two weeks and took place in as reasonably quiet and private places as the schools could furnish. During the sessions, S and E were seated at a table with the instructional materials in front of them.

Session 1

The first session began by giving S a brief introduction to the intent and format of the training. S was told that a number of different kinds of materials would be used to try to teach him how to perform "fair tests." This, coupled with the initial use of this term in the context of bouncing tennis balls, was done to provide S with an intuitive feel for what the training was all about, in a sense to provide S with a "ball park" in which to work. In psychological terms it amounts to the "whole" which will later become differentiated (Lawson, 1967). The materials used in this session were materials very familiar to children, three tennis balls (two which were relatively bouncy and one which was considerably less bouncy) two square pieces of cardboard, two square pieces of foam rubber, and a table. S was told that the first problem was to find out which of the tennis balls was the bounciest. To do this, S would instruct E in how to perform the experiment and E would carry out S's instructions. Although each session varied somewhat, in general S would begin by telling E to take two balls and drop them to see which bounced higher (height of bounce then became the dependent variable). E would then drop the two balls but drop them from different heights (an uncontrolled experiment). S would then respond by saying: "That isn't fair. Drop them from the same height." On the next trial, the height would be equalized; however, one ball would be dropped so that it hit the table top while the other ball hit the floor (again an uncontrolled experiment). This procedure was followed by continually trying to intervene with new uncontrolled variables (spin one ball, push one ball, let one ball hit cardboard or foam rubber). Ss were then told that a test was called a "fair test" if all the things (variables) that might make a difference were the same in both balls (except, of course, for the difference in the balls themselves). Each time a test was made in which these variables were not the same was called an "unfair test." Following introduction of those more general statements and terms, several additional examples were given and talked through.

*For a more detailed explanation of the scoring procedures used for this task as well as for the pendulum and the balance beam tasks, see Lawson, Nordland, and DeVito (1974).

The overall intent of this first session was to start with an undifferentiated whole (as suggested by Lawson, 1967), capitalize on the S's intuitive understanding (as suggested by Ausubel, 1964), provide numerous particular images (as suggested by Bruner & Kenney, 1970), provide contradictions (as suggested by Piaget, 1967), and provide symbolic notation (the phrases "fair" and "unfair tests") which remained invariant across changes in imagery–provided by the materials used in the subsequent sessions (also suggested by Bruner & Kenney, 1970; and by Lawson, Blake, & Nordland, 1975).

Session 2

The second session began by reminding Ss of the intent of the training and by pointing out the new materials. The materials were those used for the bending rods task administered during the posttesting. Six metal rods of varying size, shape, and material were placed on the table and S was asked to classify them in as many ways as possible. This was done to determine S's ability to form the classes of size, shape, and material, and to insure that these differences in the rods were noted. The rods were then placed into a stationary block of wood and all the factors (variables) which might affect the amount of bending of the rods (the dependent variable) were discussed. S was then asked to perform "fair tests" to find out if the variables of length, thickness, shape, and material of the rods, as well as the amount of weight hung on the end of the rods affects the amount the rods will bend. Whenever S performed a test, he was asked: "Is this a fair test?" "Why is it a fair test?" "Can you be sure that this rod bends more than that one *only* because it is thinner?" "Is there any other reason (an uncontrolled variable) why it might be bending more?" These questions and others were used to focus S's attention on all the relevant variables and recognize unambiguous and ambiguous experiments in an attempt to lead than to understand the necessity for keeping "all factors the same" except the one being tested to determine causal relationships. A number of examples and counterexamples were discussed at length. The concepts involved in this session were identical to that of the first; the material (the context), however, was different.

Session 3

At the outset of the third session, Ss were asked to experiment with an apparatus called a Whirly Bird (SCIS, 1970). The Whirly Bird consists of a base which holds a post. An arm is attached to the end of the post. When pushed or propelled by a wound rubber band, the arm will spin around like the rotor on a helicopter. Metal weights can be placed at various positions along the arm. Ss were briefly shown how the Wirly Bird worked and were given the task of finding out all the things (variables) which they thought might make a difference in the number of times the arm would spin before it came to rest (the dependent variable). Possible variables included the number of times the rubber band was wound, the number of rubber bands, the number of weights placed on the arm, the position of the weights, how tight the arm and post were fastened together, the angle of the base, etc. Following Ss' exploration with the apparatus, they were asked to perform "fair tests" to prove that the independent variables mentioned actually did make a difference in the number of times the arm would spin. Again, whenever a test was performed Ss were asked questions such as these: "Is this a fair test?" "Why is it a fair test?" "Does it prove that it makes a difference?" "Why else might the arm spin more times?" (i.e., were all other independent variables held constant?)

The general intent of this session was similar to that of the second session and the fourth and final session. The concepts underlying the questions and materials were identical in all sessions. The symbolic notation (the language used) remained invariant, while transformations in imagery were gained by using materials extending from the familiar to the unfamiliar. Ss were given a variety of concrete experiences so they could learn by doing and at each opportunity they were challenged to transform that doing into language.

Session 4

In this session, the use of concrete materials as the source of activity and discussion was replaced by the use of written problems. Problems posed only in a written fashion were considered to represent an additional step away from the concrete and towards the abstract or formal level. Probing questions relative to Ss' understanding of the written situations were asked as was done in the previous sessions. In a sense, learning by doing was replaced by learning by discussion (language alone). The following two written problems were presented and discussed at length.

Written problem 1. Fifty pieces of various parts of plants were placed in each of five sealed jars of equal size under different conditions of color of light and temperature. At the start of the experiment each jar contained 250 units of carbon dioxide. The amount of carbon dioxide in each jar at the end of the experiment is shown in the table.

Which two jars would you select to make a fair comparison to find out if temperature makes a difference in the amount of carbon dioxide used?

Jar	Plant Type	Plant Part	Color of Light	Temp. (°C)	CO_2
1.	Willow	Leaf	Blue	10	200
2.	Maple	Leaf	Purple	23	50
3.	Willow	Root	Red	18	300
4.	Maple	Stem	Red	23	400
5.	Willow	Leaf	Blue	23	150

Written problem 2. An experimenter wanted to test the response of mealworms to light and moisture. To do this he set up four boxes as shown in the diagram below. He used lamps for light sources and watered pieces of paper in the boxes for moisture. In the center of each box he placed 20 mealworms. One day later he returned to count the number of mealworms that had crawled to the different ends of the boxes.

The diagrams show that mealworms respond (respond means move to or away from) to:*

A. light but not moisture
B. moisture but not light
C. both light and moisture
D. neither light nor moisture

*During the training some Ss noticed that temperature might also play a role in determining the mealworm's response. For these Ss a brief discussion of this variable and its influence on the experiment's results was undertaken. For Ss who did not isolate temperature as a significant variable no mention of it was made by E.

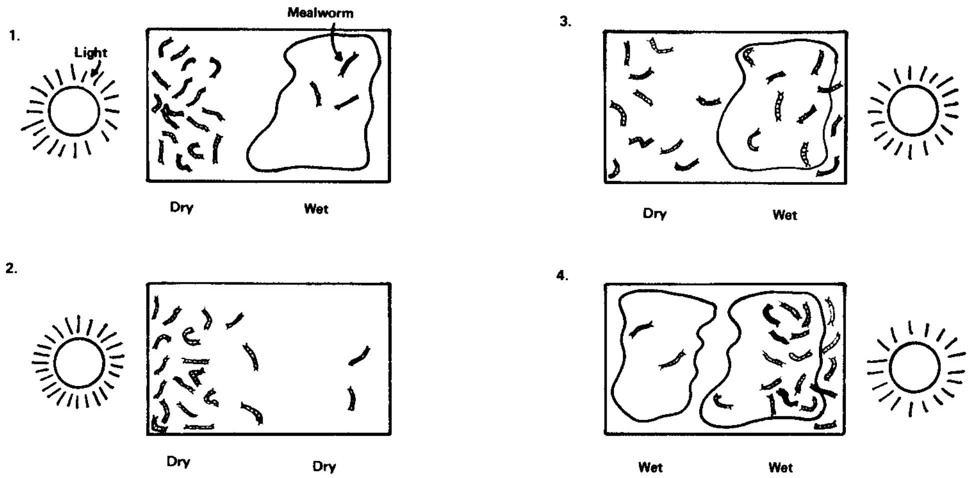


Fig. 1. Diagram showing experimental conditions.

Posttests

The bending rods task administered during the pretesting plus the following tasks were individually administered to Ss in phase one of the posttesting.

The pendulum (Inhelder & Piaget, 1958). The pendulum task tested S's ability to control and exclude irrelevant variables using a simple pendulum. First, E pointed out the independent and dependent variables. S was then given the problem of determining what variables affect the period. Since the only causal variable was length of the string, the variables of bob, angle of drop, and force or push must be excluded. This demonstration required understanding of concept "all other things being equal"–the trained concept.

The balance beam (Inhelder & Piaget, 1958). Using a balance beam and hanging weights, this task tested S's ability to balance various combinations of weights at various locations along the beam, e.g., given a 10-unit weight 5 units of length from the fulcrum, S was asked to predict the proper location of a 5-unit weight to achieve a balance. Successful completion of this task implied understanding of inverse proportionality.

Piagetian level of performance on each of these three tasks was assessed on the basis of the quality of S's verbal responses and their ability to exhibit the appropriate behavior. Performances were categorized into the following levels.

Piagetian Level	Points
Preoperational	1
Early concrete-operational-IIA	2
Middle concrete-operational	3
Fully concrete-operational-IIB	4
Post concrete operational	5
Early formal-operational-IIIA	6
Middle formal operational	7
Fully formal-operational-IIIB	8

The Peel question (Peel, 1971). A paragraph and question written by Peel which involved the floods of Florence, Italy was modified to read as follows:

All countries have art museums and Mexico is very rich in art treasures. Many people travel to Mexico just to see these old paintings and statues. Floods in Mexico recently damaged many of these great works of art. Old paintings are rare, valuable and beautiful and should be kept in a safe place.

Question: Are the people of Mexico to blame for the damage to the paintings and art treasures?

Responses were tape recorded and classified into one of the following categories:

1. Restricted–1 point. Irrelevancy, tautology, and inconsistency may dominate. S may deny premises or other conditions of the problem.
2. Circumstantial–2 points. Thinking is dominated by the content of the material. S is unable to look outside it.
3. Imaginative–3 points. The subject realizes that he has to go beyond the content of the passage to evoke possible hypothesis from his own experience–extenuating possibilities are evoked.

Although no reliability figures are available for this measure, Peel (1971) reports a high degree of consistency of level of judgment among a series of similar passages.

The following measures were administered as group tests during the second phase of the posttesting.

The spheres task (Wollman, 1975). The spheres task consisted of three written questions requiring understanding of the necessity for the control of variables in the context of rolling spheres down inclined planes. The independent variables were the position on the inclined plane from which the spheres were released, the weight of the spheres, and the weight of target spheres which were positioned at the base of the inclined plane. The dependent variable was the distance traveled by the target sphere after it was hit by the rolling sphere.

Questions required Ss to select the proper release positions, weight of rolling spheres, and weight of target spheres to determine the affect of each variable on the distance the target sphere would travel. All selections were followed by written explanations. Responses on each of the three questions were classified into the following four categories.

1. Incorrect selection, or a correct selection but irrelevant explanation–1 point.
2. Correct selection followed by a simple description of what would occur when the experiment was performed–2 points.
3. Correct selection followed by an explanation stating that the selection was made to insure that the test be fair–3 points.
4. Correct selection followed by an explanation stating that *if* the comparison was not made in this fashion, *then* it would not be fair; *therefore*, an unambiguous solution would not be obtained–4 points.

The Longeot examination (Longeot, 1962, 1965). The original Longeot examination is a subject matter-free examination, consisting of 28 problems requiring either concrete-, transitional-, or formal-operational thinking for successful solution. Since the time available did not allow administration of all 28 problems, a shortened version of the examination consisting of eight problems was administered. Two problems involving each of the following reasoning processes were selected: class inclusion operations, proportional reasoning, propositional logic, and combinatorial analysis. The examination required approximately 20 minutes for completion. Total scores were obtained for each S. Also, following procedures adopted by Longeot, Ss were categorized into concrete-operational-IIA, concrete-operational-IIB, postconcrete operational, formal-operational-IIIA, and formal-operational-IIIB levels of intellectual development.*

*Details of the examination and the scoring procedures can be obtained from the authors.

The KR-20 reliability of this shorter version was .63. Validity of the examination was originally established by Longeot (1965).

Combinatorial question. One question involving combinatorial analysis was given to the sample of fifth-grade Ss. The question read as follows:

> The Rapid Rover Bicycle Club has license plates for its members. On each license plate is a combination of letters. How many different license plates can you make up using only three letters A, B, and C, and not using a letter more than once in the same license plate?

One point was awarded for each original license plate S was able to generate.

Logic question. The question involving propositional logic which was presented to the fifth-grade sample was:

> When big dogs bark the mailman runs.
> You see the mailman running.
> Did a big dog bark?

Subjects were instructed to answer "yes" or "no" and justify their answer. A "yes" answer constitutes committing the logical fallacy known as affirming the consequent. One point was awarded for a correct answer and justification; no points were awarded for an incorrect answer. No reliability or validity data was available for the combinatorial question or the logic question prior to this study.

Results

Pretest-Posttest Gains in Intellectual Level

Table I shows the mean pretest and posttest levels of intellectual development for both the fifth- and seventh-grade samples as measured by the bending rods task. The fifth-grade experimental group showed a gain in level from 3.93 (slightly less than fully concrete-operational-IIB) to 7.06 (between early formal-IIIA and full formal-operational-IIIB). The Wilcoxon T test (Siegel, 1956) was used to test for significant differences in pre- to posttest performance. The calculated Wilcoxon T value of 0.0 for experimental group's gain was highly significant ($p < .001$). The fifth-grade control group's slight gain from 4.00 to 4.42 was not significant ($T = 9.0$, $p > .05$). The seventh-grade experimental group showed a gain in level from 4.50 to 7.37. This gain was highly significant ($T = 0.0, p < .001$). The seventh-grade control group's gain from 4.75 to 5.43 was not significant ($T = 11.5, p > .05$).

Experimental and Control Group Posttest Comparison–Fifth Grade

Means and standard deviations for the trained task (bending rods), the specific transfer tasks involving the concept of controlling variables (pendulum task, spheres task), and the nonspecific transfer tasks of general intellectual level (balance beam task, Peel question, combinatorial question) for the fifth-grade experimental and control groups are shown in Table II. The Mann-Whitney μ test was used to test for significant differences in the experimental and control group performance. The values in Table II show that the experimental group performed significantly better than the control group on the bending rods task ($p < .001$), on the pendulum task ($p < .10$), and on the spheres task ($p < .10$). However, on the remaining measures, group differences did not reach significance ($p > .10$).

TABLE I

Mean Pretest and Posttest Levels of Intellectual Development as Measured by the Bending Rods Task

Group	Pretest	Posttest
Fifth Grade		
Experimental	3.93	*7.06
Control	4.00	4.42
Mean	3.96	5.75
Seventh Grade		
Experimental	4.50	*7.37
Control	4.75	5.43
Mean	4.62	6.41

*$p < .001$.

TABLE II

Means, Standard Deviations, and Mann-Whitney μ Values for Experimental and Control Group Posttest Measures—Fifth-Grade Sample

Posttest Measure	Experimental		Control		μ
	M	SD	M	SD	
Trained Task					
Bending Rods	7.06	.99	4.43	1.02	**10.5
Specific Transfer Tasks					
Pendulum	5.38	2.13	4.14	.95	*83.5
Spheres	4.75	2.77	3.00	1.18	*78
Nonspecific Transfer Tasks					
Balance Beam	4.63	1.15	4.14	.54	89
Peel	2.00	.37	2.00	.39	105.5
Combinatorial	5.74	.45	5.29	1.07	88

*$p < .01$.
**$p < .001$.

TABLE III

Means, Standard Deviations, and Mann-Whitney μ Values for Experimental and Control Group Posttest Measures–Seventh-Grade Sample

Posttest Measure	Experimental M	Experimental SD	Control M	Control SD	µ
Trained Task					
Bending Rods	7.37	.81	5.43	1.50	***24
Specific Transfer Tasks					
Pendulum	7.56	.73	5.13	1.54	***20
Spheres	6.88	2.78	3.94	2.02	**53
Nonspecific Transfer Tasks					
Balance Beam	4.63	1.15	4.63	.96	121
Peel	2.31	.60	2.38	.72	181
Longeot	4.13	1.45	3.44	1.36	*69

*$p < .05$.
**$p < .01$.
***$p < .001$.

Experimental and Control Group Posttest Comparison–Seventh Grade

Means and standard deviations for the six posttest measures for the seventh-grade experimental and control groups are shown in Table III. Inspection of the table shows that the experimental group performed significantly better than the control group on the bending rods task ($p < .001$) and on the specific transfer tasks–the pendulum task ($p < .001$) and the spheres task ($p < .01$). The experimental group also performed significantly better on one measure of nonspecific transfer–the Longeot examination ($p < .05$). However, on the balance beam task and on the Peel question, group differences failed to reach significance ($p > .10$).

Comparison of Pretest Level of Intellectual Development and Posttest Performance–Combined Experimental Groups

Table IV shows a comparison of Ss' level of intellectual development as determined by performance on three pretest tasks with their performance on the posttest measures. The 32 experimental group Ss from the fifth- and seventh-grade samples were judged to be at either the concrete-IIA, concrete-IIB, postconcrete, or formal-IIIA level of intellectual development on the basis of combined responses on the conservation of weight, conservation of volume, and volume displacement tasks. The number of Ss which were categorized into each level is shown in Table IV as is the mean posttest score for each group on each posttest measure. F-ratios and their associated probability values are also shown. For the trained task (bending rods), the more formal subjects (postconcrete and formal-IIIA) demonstrated slightly higher posttest scores

TABLE IV

Comparison of Pretest Level of Intellectual Development with Mean Scores on Posttest Measures for Combined Fifth- and Seventh-Grade Experimental Groups

Posttest Measure	Pretest Level of Intellectual Development: Concrete-IIA (n = 7)	Concrete-IIB (n = 11)	Postconcrete (n = 9)	Formal-IIIA (n = 5)	F-ratio	Prob.
Trained Task						
Bending Rods	6.85	7.09	7.44	7.60	.92	.45
Specific Transfer						
Pendulum	5.00	6.72	6.55	7.80	2.59	.07
Spheres	4.00	5.27	6.56	8.20	2.69	.06
Nonspecific Transfer						
Balance Beam	4.00	4.54	5.11	4.80	1.38	.27
Peel	2.14	2.18	2.00	2.40	.64	.60
Longeot	2.00 *(n = 2)	3.83 (n = 6)	4.33 (n = 3)	5.20	3.79	.04

*Since the Longeot examination was not adminstered to both samples, the number of subjects in the first three levels was reduced.

(6.85 for the concrete-IIA Ss to 7.60 for the formal-IIIA Ss). The group differences, however, were not significant ($F_{3,28} = .92$, $p = .45$). Mean scores for the specific transfer tasks (pendulum and spheres) were also slightly higher for the more formal Ss. The obtained F-ratios and probability levels for the mean differences for the pendulum and spheres tasks were $F_{3,28} = 2.59$; $p = .07$, and $F_{3,28} = 2.69$; $p = .06$, respectively. Significant group differences were found on the Longeot examination ($F_{3,12} = 3.79$; $p < .05$) but not on the balance beam task and Peel question ($p > .10$ for both measures).

Discussion

The answer to the first question posed at the outset of this investigation is yes. Instruction incorporating the described procedures can affect the transition from concrete to formal cognitive functioning in these fifth- and seventh-grade students with respect to the ability to control variables. The results in Table I indicated both fifth- and seventh-grade experimental group Ss performed at the formal level on the posttests. Not surprisingly, the seventh-grade Ss performed at a slightly more formal level than the fifth-grade Ss. The finding that the experimental groups also performed significantly better than control groups on the specific transfer tasks (the pendulum task and the spheres tasks) indicates that the training was generalizable to tasks involving novel materials (see Tables II and III). The answer to the second question, therefore, is also yes.

On tasks designed to measure nonspecific transfer of training (i.e., tasks involving concepts other than the trained concept but still involving concrete and formal reasoning), differences between the fifth-grade experimental and control groups were not significant ($p > .10$). This indicates that, although the training was effective in promoting formal thought with regard to one aspect of formal reasoning, it was limited in extent. This is precisely the result predicted by Ausubel (1964). In an extensive review of studies attempting to train conservation reasoning, Brainerd and Allen (1971) found a similar result. They reported: "In short, those studies that looked for specific transfer of induced first-order conservations, without exception, have found it." (p. 137.) Conversely, in several studies which looked for nonspecific transfer of induced conservation, none was found. These results were taken to be supportive of Piaget's position with regard to the existence of general levels of development.

In light of these findings, the fact that the seventh-grade experimental group performed at a significantly ($p < .05$) more formal level on one measure of nonspecific transfer (the Longeot examination) seems indeed surprising. Did the training facilitate a general shift in the experimental group's ability to reason formally? At first glance it appears that this may be the case since the Longeot examination has been shown to be a valid measure of concrete- and formal-reasoning abilities (Lawson & Blake, 1976; Sheehan, 1970). An alternative explanation however, can be advanced. That is, the difference is not due to the fact that the experimental group performed more formally because of a general advance in reasoning but that the control group performed *below* their capabilities. Possibly a personal rapport established during the training sessions among the examiners and the experimental Ss did not develop with the control group Ss. For this reason, the control group simply did not try as hard as the experimental group did on the written examination. The likelihood of this occurring on the Longeot examination was great because it, unlike the interview tasks, did not involve personal contact with the examiners. Which of these interpretations is correct we cannot say. Indeed, other factors may be operating. However, on the average, the control group did perform at a somewhat more concrete level on the Longeot examination than they did on the bending rods, pendulum, and balance beam tasks.

Data on Table IV can be used to answer the fourth question addressed by this investigation. What is the relationship between a student's level of intellectual development and his ability to profit from the training? The data indicate that the more formal Ss (those in the postconcrete and formal-IIIA levels) were somewhat more receptive to training than the more concrete Ss (those in the concrete-IIA and concrete-IIB levels). Nevertheless, the concrete-IIA and concrete-IIB Ss did perform at the formal level on the bending rods posttest task. (Recall that a score of 6 was equivalent to the formal-IIIA level and a score of 8 was equivalent to the formal-IIIB level.) This finding is somewhat contrary to the Piagetian position that training can be effective only for persons in a transition period. The fact that specific transfer of training (a better indication of actual comprehension of the trained concept) was significantly related ($p <$ .10) to pretest level of intellectual development is a result more closely aligned with the Piagetian position.

What, then, can be said about the transition from concrete- to formal-operational thought in light of these experimental results and our conservations with children? It appears to us that formal operations develop initially as intuitions during the concrete stage, perhaps as early as six to seven years of age. At the onset of this investigation, virtually all the experimental Ss insisted that to determine which tennis ball was the bouncier the balls must be dropped from the same height and hit the same surface on the floor. In each instance, Ss demonstrated an intuitive feeling that the tests were "not fair" and would respond by saying: drop them from the same height, make them both hit the floor, don't spin one, etc. After the comparisons with the tennis balls were made, they were able to accept or reject them as fair or unfair but they were unable to state a general rule or procedure for performing fair tests prior to the test itself (i.e., to perform a fair test, keep all the factors equal except that which you are testing). Not even the most articulate Ss were able to spontaneously repond by telling E to have "every thing the same" for both balls. Even when they were asked to summarize their instructions without mentioning specific factors, they were intially at a loss for words. Apparently they had a feeling for eveness, fairness, and symmetry, but not a general rule to act as a guide for behavior–i.e., they lacked the ability to use language to structure their thinking. This phenomenon appears very much akin to the experience we all have had when we "know" something is true but just cannot seem to find the words to explain it.

Our belief is that the extension of this intuitive understanding to the point where this intuition can be expressed clearly through the use of language and applied successfully to solve problems constitutes the acquisition of formal thought. This process which enables persons to overcome the pull of the perceptual field (they are no longer "object-bound") we believe is the fundamental process in intellectual growth. For intuitions to manifest themselves in the form of useful linguistic rules (formal operations) we presumed (and the experimental results supported) that children need (1) a variety of concrete experiences involving a conceptualization and instrumental activities and (2) a useful symbolic notation which remains invariant across transformations in images in this instance, the symbolic notation was language, the key words being "fair test" and "unfair test." This, of course, is essentially the position taken by Bruner and Kenney (1970) cited above. It incorporates key points suggested by Raven (1974) as well. Raven, in designing instructional strategy to facilitate the acquisition of logical operations suggests three necessary factors: (1) the task organization must correspond to the child's levels of reasoning, (2) the instructional strategy must incorporate the active engagement of the student in using his logical operations in the construction of rules and concepts, and (3) concrete referents must, whenever possible, be provided.

What then about the formal-operational stage in general? Does it exist? To that question we have no clear answer. However, a number of studies have shown that the development of formal ideas such as proportional reasoning and the control of variables do develop roughly in a parallel fashion (Kohlberg & DeVries, 1971; Lawson, 1973; Lawson & Blake, 1976). For example, Lawson (1973) found correlations of .40-.48 ($p < .01$) between the bending rods at the balance beam tasks in samples of high school biology, chemistry, and physics students. Further, the mean level of performance on the two tasks was about identical for the three samples. In other words, if a student performs at a formal-IIIA level on the bending rods task, it is likely (but by no means certain) that they will also perform at the formal-IIIA level on the balance beam task. The correlation between the same two tasks in the sample of fifth- and seventh-grade Ss in this study following training was predictively less (rho = .33 and .23, respectively). The training, in effect, increased the decalage (separation) between these two aspects of formal reasoning.

The posttest performance of one student on the bending rods task was particularly interesting and demonstrative of this artificial separation. He performed almost perfectly. He systematically and unhestitatingly isolated and proved the effect of each variable until he attempted to prove that the shape of the rod made a difference in the amount it would bend. He chosen two rods of different shape but of equal thickness, equal length, same material, and hung the same amount of weight on each rod. Then he checked to see which rod was bending more and could not notice a difference. What was extraordinary was that he had extended the rods from the block of wood in which they were fastened only about 3 cm. When no difference in bending was observed, he was unable to proceed. This is a phenomenon that we have never observed in nontrained Ss who spontaneously and unhesitatingly controlled variables (i.e., a spontaneous formal-IIIB performance). When differences in bending are likely to be small, these Ss will spontaneously pull the rods out as far as possible in order to magnify any differences which may exist. The difference in bending between the square and the round rods varies directly with the length of the two rods. Understanding of this proportional relationship was lacking in this trained S and other trained Ss, as evidenced by their lower performance on the balance beam task. Therefore, it appears that the training manifested itself in one aspect of formal thought but not in what could be interpreted as formal reasoning in general.

Kohlberg and DeVries (1971) claim such a concept as formal reasoning in general is meaningful because they and others have isolated a general Piagetian cognitive-level factor distinct from psychometric intelligence.* One might ask then was the training of any value to the students? The answer, we believe, is yes. It does represent a meaningful but limited advance toward an abstract quality of thought in the student. Inhelder and Matalon (1960) apparently would agree. In a discussion of the acquisition of trained conservation concepts they stated, "This process of acquisition which can, of course be accelerated by training, corresponds to a general progess toward an 'operational' quality in the thought of the child." (p. 446.)

It should, however, be pointed out that the aim of efforts such as those reported in this study, if used on a wide scale in classrooms, should not be to accelerate intellectual development as mentioned by Inhelder and Matalon (1960), but to avoid what might be called "stage-retardation." Ample evidence exists, as mentioned previously, that this phenomenon of stage-retardation is indeed widespread.

Implications for Teaching

The position taken by Elkind (1972) that instruction in controlled experimentation should generally not be introduced until adolescence seems unfortunate. Rather, what seems called for is a very gradual introduction and continued reintroduction of lessons involving concrete materials and activities which enable students to make comparisons and to make judgments on the basis of their comparisons. The elementary school science program recently developed by the Science Curriculum Improvement Study (SCIS) at the University of California, Berkeley, appears to do just this. SCIS first-graders begin to carry out experiments on living organisms and physical materials. At first, these experiments are conducted with much teacher guidance. However, as the students continue to perform experiments, they gain experience and are gradually able to design and conduct these experiments on their own. The SCIS first-grade students (with much guidance from the teacher and much peer interaction) can carry out and discuss scientific experiments based on their intuitive feel for evenness and fairness. It is likely that only with the aid of useful words and phrases and a rich variety of concrete experiences that progress from the intuitive to the abstract, and that allow students to make mistakes and then critically reflect upon those mistakes, will these intuitive feelings become explicit and powerful conceptualizations during adolescence.

The authors wish to express sincere appreciation to Eugene Ballock and Kenneth Garrett of the Lafayette Public Schools for assistance in conducting this research. They also would like to thank Jane Bowyer and Dennis Schatz for interviewing students. AESOP (Advancing Education Through Science-Oriented Programs) is supported by a grant from the National Science Foundation.

References

Ausubel, D. P. The transition from concrete to abstract cognitive functioning: Theoretical issues and implications for education. *Journal of Research in Science Teaching*, 1964, **2**, 261-266.

*The unity of formal thought has been brought into serious question by a number of Piagetian scholars. Perhaps the most notable is Professor Eric Lunzer. In his 1973 address to the Jean Piaget Society, he reviewed some of the pertinent literature and recent findings relevant to this issue (Lunzer, 1973). Piaget himself has modified his earlier position somewhat in light of recent findings (Piaget, 1972).

Brainerd, D. J., & Allen, T. W. Experimental inductions of "first-order" quantitative invariants. *Psychological Bulletin*, 1971, **75**(2), 128-144.

Brekke, B., & Williams, J. D. Conservation and sex. *Perceptual and Motor Skills*, 1973, **37**, **14.**

Bruner, J. S., & Kenney, H. J. Representation and mathematics learning. In W. Kessen and C. Kuhlman (Eds.), *Cognitive development in children*. Chicago: University of Chicago Press, 1970.

Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1966.

Education Policies Commission. *The central purpose of American education*. Washington, D.C.: National Education Association of the United States, 1961.

Elkind, D. Children's discovery of the conservation of mass, weight, and volume: Piaget replication study II. *Journal of Genetic Psychology*, 1961, **98**, 219-227.

Elkind, D. Piaget and science education. *Washington Monthly*, 1972, **10**(3), 9-12.

Flavell, J. H. *The developmental psychology of Jean Piaget*. New York: D. Van Nostrand, 1963.

Friot, F. E. The relationship between an inquiry teaching approach and intellectual development. Unpublished doctoral dissertation, University of Oklahoma, 1970.

Goldschmid, M. L., & Bentler, P. M. *Manual: Concept assessment kit–conservation*. San Diego, Calif.: Educational and Industrial Testing Service, 1968.

Higgens-Trenk, A., & Gaite, A. J. H. Elusiveness of formal-operational thought in adolescents. *Proceedings 79th Annual Convention*, APA, 1971, pp. 201-202.

Inhelder, B., & Matalon, B. The study of problem solving and thinking. In P. Mussen (Ed.), *Handbook of research methods in child development*. New York: Wiley, 1960.

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

Karplus, R., & Lavatelli, C. *The developmental theory of Piaget: Conservation*. San Francisco: Davidson Films, 1969.

Kohlberg, L., & DeVries, R. Relations between Piaget and psychometric assessments of intelligence. In C. Lavatelli (Ed.), *The natural curriculum*. Urbana, Ill.: ERIC, 1971.

Kohlberg, L., & Gilligan, C. The adolescent as a philosopher: The discovery of the self in a postconventional world. *Daedalus*, 1971, **100**, 1051-1084.

Kohlberg, L., & Mayer, R. Development as the aim of education. *Harvard Educational Review*, 1972, **42**(4), 449-496.

Lawson, A. E. Relationships between concrete- and formal-operational science subject matter and the intellectual level of the learner. Unpublished doctoral dissertation, University of Oklahoma, 1973.

Lawson, A. E., & Blake, A. J. D. Concrete and formal thinking abilities in high school biology students as measured by three separate instruments. *Journal of Research in Science Teaching*, 1976, **13**(3), 227-235.

Lawson, A. E., & Renner, J. W. A quantitative analysis of responses to Piagetian tasks and its implications for curriculum. *Science Education*, 1974, **58**(4), 545-559.

Lawson, A. E., & Renner, J. W. Relationship of science subject matter and developmental levels of learners. *Journal of Research in Science Teaching*, 1975, **12**(4), 347-358.

Lawson, A. E., Blake, A. J. D., & Nordland, F. H. Training effects and generalization of the ability to control variables in high school biology students. *Science Education*, 1975, **59**(3), 387-396.

Lawson, A. E., Nordland, F. H., & DeVito, A. Piagetian formal operational tasks: A crossover study of learning effect and reliability. *Science Education*, 1974, **58**(2), 267-276.

Lawson, A. E., Nordland, F. H., & DeVito, A. Relationships of formal reasoning to achievement, aptitudes, and attitudes in preservice teachers. *Journal of Research in Science Teaching*, 1975, **12**(4), 423-431.

Lawson, A. E., Nordland, F. H., & Kahle, J. B. Levels of intellectual development and reading ability in disadvantaged students and the teaching of science. *Science Education*, 1975, **59**(1), 113-125.

Lawson, C. A. *Brain mechanisms and human learning*. Boston: Houghton Mifflin, 1967.

Longeot, F. Un essai d'application de la psychologie genetique a la psychologie differentielle. *Bulletin de l'Institut National D'Etude*, 1962, **18**(3), 153-162.

Longeot, F. Analyse statistique de trois tests genetiques collectifs. *Bulletin de l'Institut National D'Etude*, 1965, **20**(4), 219-237.

Lunzer, E. A. Formal reasoning: A re-appraisal. Address to the Jean Piaget Society Annual Meeting, May, 1973.

McKinnon, J. W., & Renner, J. W. Are colleges concerned with intellectual development? *American Journal of Physics*, 1971, **39**(9), 1047-1052.

Peel, E. A. *The nature of adolescent judgment*. New York: Wiley-Interscience, 1971.

Piaget, J. *Biology and knowledge*. Chicago: University of Chicago Press, 1967.

Piaget, J. Intellectual evolution from adolescence to adulthood. *Human Development*, 1972, **15**, 1-12.

Raven, R. J. Programming Piaget's logical operations for science inquiry and concept attainment. *Journal of Research in Science Teaching*, 1974, **11**(3), 251-261.

Science Curriculum Improvement Study. *Subsystems and variables teacher's guide*. Chicago: Rand McNally, 1970.

Siegel, S. *Nonparametric statistics for the behavioral science*. New York: McGraw-Hill, 1956.

Sheehan, D. J. The effectiveness of concrete and formal instructional procedures. Unpublished doctoral dissertation, State University of New York, Albany, 1970.

Wollman, W., & Karplus, R. Intellectual development beyond elementary school V: Using ratio in differing tasks. *School Science and Mathematics*, 1974, **74**(7), 593-611.

Wollman, W. Intellectual development beyond elementary school VI: Controlling variables: A survey. *School Science and Mathematics*, 1975, submitted.