DOCUMENT RESUME

ED 104 859 SP 009 097

AUTHOR Bart, William M., Ed.; Wong, Martin R., Ed.
TITLE Psychology of School Learning: Views of the Learner. Volume I: Environmentalism.
INSTITUTION MSS Information Corp., New York, N.Y.
PUB DATE 74
NOTE 249p.
AVAILABLE FROM MSS Information 'orporation, 655 Madison Avenue, New York, New York 10021 ($6.25 paper, $12.00 cloth, 20 percent discount on orders of $200.00 or more)

EDRS PRICE MF-$0.76 HC-$12.05 PLUS POSTAGE
DESCRIPTORS Behavioral Objectives; Computer Assisted Instruction; Educational Philosophy; *Educational Psychology; *Environmental Influences; Higher Education; *Learning; *Learning Specialists; Reinforcement; Teacher Educatica

ABSTRACT

This document is the first of three volumes presenting essays from three schools of thought regarding learning. Volume one consists of readings from psychologists, philosophers, and learning theorists concerning the view that the learner is a product primarily of environmental factors. The list of essays includes the following: (a) "Ideas and Their Origin," (b) "The Free and Happy Student," (c) "The Technology of Teaching," (d) "Treatment of Nonreading in a Culturally Deprived Juvenile Delinquent: An Application of Reinforcement Principles," (e) "Production and Elimination of Disruptive Classroom Behavior by Systematically Varying Teacher's Behavior," (f) "Learning Theory Approaches to Classroom Management: Rationale and Intervention Techniques," (g) "A Token Reinforcement Program in a Public School: A Replication and Systematic Analysis," (h) "Educational Technology: New Myths and Old Realities," (i) "Computerized Instruction and the Learning Process," (j) "Teaching Machines: A Review," (k) "Instruction and the Conditions of Learning," (l) "Mastery Learning and Mastery Testing," (m) "Student Evaluation and Promotion of Learning," (n) "Behavioral Objectives: A Close Look," (o) "Why Behavioral Objectives?," (p) "Hereditary, Environment, and the Question 'How?'" (q) "Reflections on a Decade of Teaching Machines," (r) "The Classroom as a System," and (d) "The Behaviorally Engineered Classroom: A Learner-Sensitive Environment." Authors include B. F. Skinner, John Locke, Samuel Mayo, Anne Anastasi, and others. (JS)

PSYCHOLOGY OF SCHOOL LEARNING:

Views of the Learner

VOLUME I: Environmentalism
(First of a Two-Volume Series)

Edited by
William M. Bart, Ph.D.
Martin R. Wong, Ph.D.
University of Minnesota

MSS Information Corporation
655 Madison Avenue, New York, N.Y. 10021

This is a custom-made book of readings prepared for the courses taught by the editors, as well as for related courses and for college and university libraries. For information about our program, please write to:

MSS INFORMATION CORPORATION
655 Madison Avenue
New York, New York 10021

MSS wishes to express its appreciation to the authors of the articles in this collection for their cooperation in making their work available in this format.

Library of Congress Cataloging in Publication Data

Bart, William M comp.
Psychology of school learning. Volume I.

1. Learning, Psychology of — Addresses, essays, lectures. I. Wong, Martin R., joint comp. II. Title.
LB1055.B33 370.15'2 74-13235
ISBN 0-8422-5200-2
ISBN 0-8422-0447-4 (pbk.)

CONTENTS

ENVIRONMENTALISM

INTRODUCTION

Educational systems emanate from the way the builder of the system views the learner. The actual behaviors of each teacher in each classroom stem directly from her/his assumptions about what learners are like and how learning occurs--assumptions that one makes based on an overall view of what kind of being the human animal is.

Educational learning theories are applications of psychological theories. Each theory advocates a different kind of treatment for the facilitation of learning in people. The roots of psychological theories in turn, stem from basic philosophical views of the human animal that are traceable back to the original thinkers about the condition of and makeup of the human animal.

It is possible to place most of these positions roughly on a continuum that would stretch from a belief that the human animal is a perfect bud that will blossom into a perfect being if left alone to grow uninjured and unimpeded by the vissicitudes of the environment, to a belief on the other end of the continuum, that everything an adult human was, is, or ever will be, is determined by the kinds of interactions that person has with his environment.

Obviously, there are few philosophers, psychologists, or learning theorists whose beliefs would put them on the extremes of either end of this imaginary continuum. Nevertheless, it is possible to break the continuum roughly into three different sections and to place different theorists into one of the three without doing too much injustice to most.

This set of readings is based on that assumption. The three sections that follow each contain theorists and practitioners that in our judgment belong for the most part to that segment of the theoretical continuum.

It is our belief that many problems in the schools may not emanate from questionable starting views of the learner. We have assembled these readings in order to give the student at least a superficial view of each of the positions. It is the hope that this superficial view will stimulate a closer examination of the reader's own philosophical viewpoint and that this will lead to a greater understanding of self and of actions directed toward learners.

We have titled the two ends of the continuum Nativism -- which stresses the innate capacities of the learner -- and Environmentalism -- which stresses the part played by environmental factors in learning. The middle segment of the continuum has been titled Interactionism.

Brief synopses covering each of the three sections and introductions to some of the theorists and practitioners whose works are contained herein follow, as a guide to the structure of the rest of the book.

I. Nativism

A nativist orientation to school learning expresses the view that there are structures innate to the individual which precede experience and which allow us to adapt to reality. Nativists would accept the view that a variety of behaviors (e.g., language competency, perceptual capabilities, intellectual processes) are under the control of genetic components.

A. Arnold Gesell

Gesell, who is a medical doctor, has posited a series of stages of development that humans pass through. He posits that the stages alternate with a stage manifesting behavioral equilibrium following a stage manifesting behavioral disequilibrium followed by a stage manifesting behavioral equilibrium and so on. For example, he posits that a 4 year old will tend to be vigorous and expansive in his actions, whereas a 4-1/2 year old will tend to be troubled, seemingly neurotic and introspective and a 5 year old will tend to be smooth, consolidated, and balanced in his behavior.

Understandably, Gesell views IQ as an important variable. In his contention that there are seven stages of development marked by certain age intervals, there is allowed room for certain age deviations but the sequence is complied with by all youngsters. To discipline a youngster, he suggests that teachers be aware of the behaviors that characterize the stage of development of the child and discipline the child accordingly. For example, to discipline a 4 year old, one may want to use tricks, numbers, or whisper commands as well as suggest new ways of doing things and incorporating imaginary companions in activities of the child. Gesellian theory is based on many child observations and is quite useful as a guide to handling children.

B. Gestalt psychology

Gestalt psychology, which refers to the psychology of perceptual and cognitive wholes, developed in Germany at the early decades of the 20th century. Gestaltists included men such as Max Wertheimer, Christian von Ehrenfels, Edgar Rubin, Wolfgang Kohler, and Kurt Koffka. They, in general, contended that there were cognitive and

9

and perceptual relationships and unities such as melody, figure and ground, phi phenomenon, and insight that cannot be meaningfully decomposed into their respective elements. In other words, whereas associationists and environmentalists tend to hold that a behavioral whole is the sum of its parts, gestaltists contend that a behavioral whole is more than the sum of its parts and often different also. They felt that wholes would have to be understood before parts can be understood. They posited principles such as the principle of proximity according to which things close together tend to be grouped together and the principle of closure according to which incomplete figures tend to be viewed as completed figures and they accepted concepts such as insight, consciousness, and introspection.

Kohler contended that insight is crucial in resolving problems as it relates to the seeing of relations unseen before. Wertheimer contends that for productive thinking children and adolescents should have practice in designating different figure and ground pairs, in developing alternative mental sets or ways of looking at things, and in structured problems that make available wholes or Gestalten that aid in making solutions. Rule-learning can be viewed as a topic for Gestaltists. Through processes such as insight can be triggered by the environment, they relate to innate structures according to the Gestaltists.

C. Noam Chomsky

The tenets of Chomsky's position are adequately discussed in his Review of Skinner's Verbal Behavior and further discussed in K. MacCorquodale's reply in the Journal of the Experimental Analysis of Behavior, 1970, 13, 83-99. Two aspects of his thought that may be pointed out are 1) that there are linguistic complexes such as grammar that should not be viewed as mere sums of consituent elements and 2) that there is a system of linguistic transformations (deep-structure grammar) that each person is born with that allows each person to learn any language and say a myriad of meaningful sentences in any language learned. For education, one could recommend that the teacher of language teach for phrase structure and surface structure transformations and give opportunities to the youngster to generate sentences with similar or dissimilar meanings -- i.e., opportunities to actualize latent deep-structure transformations. Teachers of language would have to learn the language of transformational linguistics to richly use Chomsky's ideas in the schools.

D. Arthur Jensen

Articles by Arthur Jensen and Sir Cyril Burt provide a reasonable introduction to the nativist orientation to intelligence. The Bart article on intelligence measurement gives an alternative view. Nativists such as Jensen and Burt hold that the genetic makeup of an individual has a much greater determining influence on intelligence of a person than the environment.

In fact, from an examination of correlations for sets of pairs of youngsters with differing environmental and genetic relationships, it was determined that about 80% of the variation in IQ scores can be attributed to heredity, 12% of the variance in IQ scores can be attributed to environmental influences between families, and 8% of the IQ score variance can be attributed to within family environmental influences.

It is important to note that their statistical analysis is quite accurate but their starting tenets are questionable. Do IQ test items favor middle-class white youngsters? Is the language in IQ tests biased? Do IQ tests really test for intellectual processes?

These questions are still being debated.

Nativists tend to view IQ test performance as a predictive measure for future scholastic success and a selection measure that aids in the classification of youngsters. IQ is viewed as basically a genetic given and education is viewed as having no effect on IQ. Thus IQ tests are not used to any rich extent, as diagnostic and prescriptive measures. In general, youngsters with high IQs should enter academic programs that require usage of cognitive and abstract thinking skills and youngsters with low IQs should have school experiences that allow them to employ associative and psycho-motor skills that can be habituated. Given that the belief the IQ can be substantially changed is held, nativists contend that a child's school experiences should be compatible with his IQ.

II. Environmentalism

This view is that the learner is a product primarily of environmental factors. Innate tendencies and structures are deemphasized in importance. The learner is primarily a reactive organism, learning by responding to environmental stimuli.

Environmentalists use the language of stimulus and response to describe human behavior. They are concerned primarily about overt behavior witnessable by two or more people. Rich use of statistics and graphs is made as linear variables are often examined. Functional relationships among variables are sought and parsimony in explanation is a goal. An additive reductionism is followed as human behavior is conceived to be reducible to stimuli and responses (the atoms in psychology to an environmentalist) and complex behaviors are conceived to be chains of stimuli and responses.

A. B. F. Skinner

To Skinner the behavior of an individual is primarily under the control of stimuli. As opposed to Pavlov who was interested in elicited behaviors terms respondents, Skinner is interested in emitted behaviors termed operants. Operant conditioning relates to the study of conditions under which a response enter under the control of reinforcing stimuli.

The key dependent variable to be examined is the strength of an operant - defined in terms of rate of operant occurrence. If an operant is followed by a reinforcing stimulus, then the strength of the operant increases. Extinction relates to the diminuation of a response through nonreinforcement. A positive reinforcer is any stimulus which increases operant strength. A negative reinforcer is any stimulus when removed that increases operant strength. Definitions of other important operant conditioning terms may be obtained from any text on operant conditioning.

A few comments on reinforcement schedules are appropriate. The smaller the interval used (within limits) in a fixed interval schedule, the larger the total number of operants will tend to be. Variable ratio schedules relate to operant development highly resistant to extinction. Also, punishment is best used to inhibit an operant occurring when administered intermittently. In employing operant methods in the schools, one may follow the following steps: 1) decide on some goal behaviors (or terminal behaviors) for student; 2) determine a viable reinforcer to be used for the student; 3) assess the behaviors of the student related to the goal; 4) plan a sequence of behavioral steps leading from the initial behaviors of the student and the goal behaviors; 5) decide on a reinforcement schedule; 6) proceed to interact with the student. Rich use of games with tokens and ood as secondary and primary reinforcers would be compatible with Skinner's method. Behavior modification techniques are quite useful in animal training and many human learning situations where unique behavioral sequences can be posited.

The intent of the study of learning to Skinner is to determine those stimuli and environmental agents that determine human behavior and to obtain data showing orderly changes characteristic of the learning process. Skinner feels that we do not know enough about learning to formulate a theory of learning.

There are substantial ethical problems associated with the Skinnerian approach. The primary one is that who decides on what goal behaviors will be used for a given human and on what basis is the decision made. Skinner does not accept such notions as insight, self-actualization, becoming, self-determination, innate capacity to imitate, etc. There is reinforcement and the innate tendency of humans to have their behavior shaped through reinforcement; everything else follows with respect to human behavior development.

B. Educational technology

Environmentalists would encourage the use of films, radio, television, and other information-dissemination devices to facilitate school learning. Programmed instruction and computer-based instruction (C.A.I.) are other methods of educational technology.

In learning a skill, there are usually a cognitive phase followed by a fixation phase followed by an automation phase. Overpractice and training in subroutines or parts of a task facilitate skill learning. Probably the careful description of a task is more important than the employment of psychological principles in teaching a task.

To enhance learning with the use of films, one should: 1) increase the redundancy of the film content, 2) allow for audience participation in the film, and 3) provide attention directive devices. To enhance learning with programmed instruction, one should: 1) begin where the subject is, 2) hint at answers and use small steps in the programmed text, 3) allow for over-learning of responses, and 4) consider carefully organization of knowledge. In some learning cases, branched programming which is similar to an irregular reading of a book is preferable to linear programming in which a linear sequence of behavioral steps is posited. For teaching machines of Skinner's design, a subject writes a response to a question cited and a response is reinforced when it matches the actual answer of the question cited on the machine.

Youngsters do learn with programmed instruction. The programmed method is best adapted to complex but structured materials with the chain property. Often superior students do better with programmed instruction but complain more. A wider range of rewards needs to be formulated as well as needs of students to be satisfied with programmed instruction. Inconclusive and conflicting evidence has been provided as to whether 1) the ordered sequences of steps are essential, 2) short steps are to be preferred to larger behavioral steps, 3) free response is the ideal response type, 4) learner should set his own pace, 5) immediate knowledge of results maximally facilitates learning. To Skinner, programmed instruction is like individual tutoring in that 1) it allows one to begin where the subject is, 2) the subject learns at his own rate, and 3) false answers are always corrected.

C. Benjamin Bloom

Much of Bloom's views are articulated in works such as "Learning for Mastery". To Bloom, school learning must be successful and rewarding in order for learning to occur through life. If certain environmental factors are manipulated, then 95% of any class should get A's and should manifest criterion behaviors. The variables that Bloom attends to include one that is not very mutable, aptitude, and four that are: 1) quality of instruction, 2) ability of the student to understand instruction, 3) perseverance or the time the learner is willing to spend on learning, and 4) time allowed for learning. Bloom favors rich use of 1) diagnostic testing procedures, 2) formative evaluation devices, and 3) small discussion groups to facilitate mastery learning. Mastery for students should be a goal for educators as it relates to better self-concept, mental health, and positive affect for a subject.

D. Robert Gagne

Gagne posits seven types of learning to be considered by educators: 1) signal learning which relates to the Pavlovian link in which a general diffuse response is tied to a signal; 2) S-R learning in which a precise response is related to a discriminated stimulus; 3) chaining or sequencing of S-R bonds; 4) verbal association or chains of words; 5) multiple discrimination which involves relating different responses to different stimuli; 6) concept learning in which a response is related to a variety of stimuli; and 7) principle learning which involves chaining of concepts.

Gagne has also encouraged the use of task analysis and learning hierarchies in learning situations. To designate learning hierarchies, one must determine the array of subordinate skills for the task in question. To validate a hypothesized hierarchy, one may use Guttman scaling procedures to examine response patterns. In most cases, more than one learning hierarchy could be determined for a given goal task.

E. Additional Comments on Environmentalism

In environmentalism, terms such as consciousness, mind, and mental are dismissed. Techniques from animal psychology have been found helpful in describing human behavior. Innate capacities, native intelligence, etc. are usually also dismissed as viable concepts.

Objective methods and scientific techniques are emphasized. Reductionism in which simple entities and S-R bonds as atoms is the prevalent view. Simplification is another key characteristic to environmentalism.

IQ tests are in general not of much use to environmentalists. However, intelligence tests which provide specific information on student performance with respect to given problems is of use as they allow for diagnostic and prescriptive purposes of tests.

III. Interactionism

Interactionism is an approach to school learning which is distinct from Nativism and Environmentalism in that the influence of both maturational factors and environmental processes on mental development is recognized as occurring in an interactive manner. For example, some environmental processes are posited to have an effect on the mental development of an individual only when the person has reached a certain level of maturation.

Interactionists tend to posit an invariant sequence of mental processes and to emphasize either the learning of cognitive structures that lead to levels of cognitive growth. Development which is regulated by certain genetic clocks and environmental forces is richly considered. Biology, logic, and mathematics are fields that feed into this general approach.

A. Jean Piaget

Jean Piaget is a noted Swiss scholar renowned for his work in child and developmental cognitive psychology. His background includes zoology and philosophy. As Freud formulated psychoanalysis as a science of neuroses, Piaget formulated genetic epistemology as a science which inquires as to what are the relationships between the histories of various concepts in time and the development of corresponding concepts in a person. From this perspective it was determined that the historical sequence of certain geometry concepts (parallelism, perspective, inside-outside) is the opposite of the developmental appearance sequence for those same concepts (inside-outside, perspective, parallelism).

Piaget is probably most noted for his theory of cognitive development which posits an invariant sequence of cognitive stages. From partly on examination of why youngsters committed certain errors in taking IQ tests, Piaget formulated his theory which is rich with biological concepts. The two prime concepts are adaptation and organization. Adaptation refers to the state of balance between an individual and his environment. Adaptation involves assimilation which relates to the process of the individual to take in the world and accommodation which relates to the process in which some cognitive mechanism becomes altered so that it is better able to incorporate some situation. Assimilation occurs with the use of schemata (e.g., grasping) which are mechanisms to take in stimuli. Schemata are posited to form systems according to the organization principle. The principle of equilibration indicates that the human organism tends to develop higher and higher levels of cognitive functioning and cognitive equilibria. A cognitive equilibrium relates to a successful level of adaptation. Equilibria have three properties: 1) field of application - the range of stimuli handled by the cognitive processes available; 2) stability - the richness of paths and solutions that a given cognitive system can provide; 3) mobility - the spatio-temporal distance between the organism and stimuli that he can handle. To Piaget higher forms of equilibria are effected with three factors: 1) maturation; 2) social milieu - language, culture, etc.; 3) physical environment and problems subjects are exposed to.

One way of viewing Piaget's system is that there are three main stages of cognitive development: 1) the stage in which children think in terms of overt actions such as grasping and which incorporates the sensorimotor level of development; 2) the stage of concrete operations in which children can perform operations such as classification and seriation on objects; 3) the stage of formal operations in which adolescents can perform operations such as theorizing on concrete on concrete operations. Each succeeding stage is marked by a richer and broader equilibrium than the previous stage. The last stages are of great concern to elementary and secondary school teachers.

The concrete stage is between ages 5 and 12 approximately and the formal stage is between the age of 13 and death. The onset of formal thought marks adolescence.

There is no difference to Piaget between the adolescent stage of thought and the adult stage of thought.

In the concrete stage operations which are internalizable, reversible acts that form systems begin to develop. Also conservation behavior develops. To each concept one can posit a concept of conservation. For example, for weight one can examine whether an individual has the concept of conservation of weight by taking an object and changing it in various ways but not in terms of its weight and asking the subject if it weighs the same as before. If the subject indicates that the weight has not changed (thus conserved) he has that conservation concept. Concrete reasoners develop conservation of concepts such as mass, weight, area, etc.

Formal reasoners develop conservation of concepts which are functions of a variety of variables such as volume. Formal thought is marked by the following: 1) hypothetico-deductive reasoning; 2) combinatorial reasoning - capability to generate all the possible combinations in a situation; 3) ability to make theories; 4) ability to think in probabilistic terms; 5) ability to understand proportionalities and analogies; 6) view that the realm of the real is a subset of all that is possible. It is recommended to employ tasks in schools that activate the highest cognitive processes available to a youngster. Thus in high school and college, it is reasonable to expect formal operations to be manifested.

With respect to motivation, Piaget believes in intrinsic motivation in terms of functional assimilation which refers to the innate tendency of the organism to utilize his cognitive processes - especially his highest developed ones. To J. Mc. V. Hunt there should be a match between the cognitive demands of a school task and the cognitive capabilities of a student to effect school learning. However, in cases where students are well set in a thought stage, new and challenging stimuli and problems may become very attractive.

B. Jerome Bruner

Jerome Bruner is a cognitive psychologist who has contended that all fields of knowledge have underlying knowledge structures and that any discipline structure can be taught to any person in some way. He shares many Piagetian ideas such as 1) there should be considered a readiness for learning for many things to be learned; 2) students must play with materials before they can understand certain concepts - play and intuition being important cognitive activities; 3) understanding of concepts with the use of concrete materials precedes symbolic abstract understanding. Bruner holds that there is a sequence of ways in humans represent the world:

1) enactive representation - marked by rise of actions;

2) ikonic representation - marked by use of images;

3) symbolic representation - marked by use of symbols. Each successive representational level is marked by more independence of responses from the nature of the stimuli. Students should gain practice in manifesting these forms of representation in school learning situations.

C. Other Comments on Interactionism

Some interactionists such as Zoltan Dienes hold that abstraction should precede generalization; abstraction refers to learning of a given concept or rule-system in a variety of contexts and generalizability refers to the learning of a more complicated concept of rule-system than one before but in a similar context. Interactionists would contend that teachers can teach for a given subject matter but also teach for a variety of thinking skills such as classification and theorizing.

To interactionists such as Piaget intelligence is adaptability so that one person is more intelligent than another person if he is more adaptible or equivalently has more cognitive processes to resolve a wider range of problems than the second person. To assess intelligence from a Piagetian stance one may use Piagetian tasks such as the Equilibrium in the Balance task. To Piaget the intelligence of a child can change quite a bit with a positive environmental press and maturation. More on intelligence from an interactionist stance can be gained from the article entitled "Issues in the measurement of intelligence" by William M. Bart.

William M. Bart
Martin R. Wong

June, 1974

IDEAS AND THEIR ORIGIN

John Locke

Idea is the Object of Thinking. Every man being conscious to himself that he thinks, and that which his mind is applied about whilst thinking being the ideas that are there, it is past doubt that men have in their minds several ideas, such as are those expressed by the words, "whiteness, hardness, sweetness, thinking, motion, man, elephant, army, drunkenness," and others. It is in the first place then to be enquired, How he comes by them? I know it is a received doctrine, that men have native ideas and original characters stamped upon their minds in their very first being. This opinion I have at large examined already; and, I suppose, what I have said in the foregoing Book will be much more easily admitted, when I have shown whence the understanding may get all the ideas it has, and by what ways and degrees they may come into the mind; for which I shall appeal to every one's own observation and experience.

AN ESSAY CONCERNING HUMAN UNDERSTANDING, abridged and edited by A. S. Pringle Pattison (Clarendon Press. Oxford) 1924, pp. 42-91.

All Ideas Come From Sensation or Reflection. Let us then suppose the mind to be, as we say, white paper, void of all characters, without any ideas; how comes it to be furnished? Whence comes it by that vast store, which the busy and boundless fancy of man has painted on it with an almost endless variety? Whence has it all the materials of reason and knowledge? To this I answer, in one word, from Experience; in that all our knowledge is founded, and from that it ultimately derives itself. Our observation, employed either about external sensible objects, or about the internal operations of our minds, perceived and reflected on by ourselves, is that which supplies our understandings with all the materials of thinking. These two are the fountains of knowledge, from whence all the ideas we have, or can naturally have, do spring.

The Objects of Sensation One Source of Ideas. First, our senses, conversant about particular sensible objects, do convey into the mind several distinct perceptions of things, according to those various ways wherein those objects do affect them; and thus we come by those ideas we have of yellow, white, heat, cold, soft, hard, bitter, sweet, and all those which we call sensible qualities; which when I say the senses convey into the mind, I mean, they from external objects convey into the mind what produces there those perceptions. This great source of most of the ideas we have, depending wholly upon our senses, and derived by them to the understanding, I call, Sensation.

The Operations of Our Minds the Other Source of Them. Secondly, the other foutain, from which experience furnisheth the understanding with ideas, is the perception of the operations of our own minds within us, as it is employed about the ideas it has got; which operations, when the soul comes to reflect on and consider, do furnish the understanding with another set of ideas which could not be had from things without: and such are perception, thinking, doubting, believing, reasoning, knowing, willing, and all the different actings of our own minds; which we being conscious of, and observing in ourselves, do from these receive into our understanding as distinct ideas, as we do from bodies affecting our senses. This source of ideas every man has wholly in himself: and though it be not sense, as having nothing to do with external objects, yet it is very like it, and might properly enough be called internal sense. But as I call the other Sensation, so I call this Reflection, the ideas it affords being such only as the mind gets by reflecting on its own operations within itself. By Reflection, then, in the following part of this discourse, I would be understood to mean that notice which the mind takes of its own operations, and the manner of them, by reason whereof there come to be ideas of these operations in the understanding. These two, I say, viz., external material things as the objects of Sensation, and the operations of our own minds within as the objects of Reflection, are, to me, the only originals from whence all our ideas take their beginnings. The term operations here, I sue in a large sense, as comprehending not barely the actions of the mind about its ideas, but some sort of passions arising sometimes from them, such as is the satisfaction or uneasiness arising from any thought.

All Our Ideas Are of the One or the Other of These. The understanding seems to me not to have the least glimmering of any ideas which it doth not receive from one of these two. External objects furnish the mind with the ideas of sensible qualities, which are all those different perceptions they produce in us; and the mind furnishes the understanding with ideas of its own operations. These, when we have taken a full survey of them, and their several modes, combinations, and relations, we shall find to contain all our whole stock of ideas; and that we have nothing in our minds which did not come in one of these two ways. Let any one examine his own thoughts, and thoroughly search into his understanding, and then let him tell me, whether all the original ideas he has there, are any other than of the objects of his reflection; and how great a mass of knowledge soever he imagines to be lodged there, he will, upon taking a strict view, see that he has not any idea in his mind but what one of these two have imprinted, though perhaps with infinite variety compounded and enlarged by the understanding, as we shall see hereafter.

Observable in Children. He that attentively considers the state of a child at his first coming into the world, will have little reason to think him stored with plenty of ideas that are to be the matter of his future knowledge. It is by degrees he comes to be furnished with them: and though the ideas of obvious and familiar qualities imprint themselves before the memory begins to keep a register of time and order, yet it is often so late before some unusual qualities come in the way, that there are few men that cannot recollect the beginning of their acquaintance with them: and if it were worth while, no doubt a child might be so ordered as to have but a very few even of the ordinary ideas till he were grown up to a man. But all that are born into the world being surrounded with bodies that perpetually and diversely affect them, variety of ideas, whether care be taken about it or no, are imprinted on the minds of children. Light and colours are busy and at hand everywhere when the eye is but open; sounds and some tangible qualities fail not to solicit their proper senses, and force an entrance to the mind; but yet I think it will be granted easily, that if a child were kept in a place where he never saw any other but black and white till he were a man; he would have no more ideas of scarlet or green, than he that from his childhood never tasted an oyster or a pine-apple has of those particular relishes.

Men are Differently Furnished with These According to the Different Objects they Converse with. Men then come to be furnished with fewer or more simple ideas from without, according as the objects they converse with afford greater or less variety; and from the operations of their minds within, according as they more or less reflect on them. For, though he that contemplates the operations of his mind cannot but have plain and clear ideas of them; yet, unless he turn his thoughts that way, and considers them *attentively*, he will no more have clear and distinct ideas of all the operations of his mind, and all that may be observed therein, than he will have all the particular ideas of any landscape, or of the parts and motions of a clock, who will not turn his eyes to it, and with attention heed all the parts of it. The picture or clock may be so placed, that they may come in his way every day; but yet he will have but a confused idea of all the parts they are made up of, till he applies himself with attention to consider them each in particular.

* * * * * * *

And thus I have given a short and, I think, true history of the first beginnings of human knowledge; whence the mind has its first objects, and by what steps it makes it progress to the laying in and storing up those ideas out of which is to be framed all the knowledge it is capable of; wherein I must appeal to experience and observation whether I am in the right. This is the only way that I can discover whereby the ideas of things are brought into the understanding. If other men have either innate ideas or infused principles, they have reason to enjoy them; and if they are sure of it, it is impossible for others to deny them the privilege that they have above their neighbours. I can speak but of what I find in myself.

B. F. Skinner

THE FREE AND HAPPY STUDENT

The natural, logical outcome of the struggle for personal freedom in education is that the teacher should improve his control of the student rather than abandon it. The free school is no school at all.

His name is Emile. He was born in the middle of the eighteenth century in the first flush of the modern concern for personal freedom. His father was Jean-Jacques Rousseau, but he has had many foster parents, among them Pestalozzi, Froebel, and Montessori, down to A. S. Neill and Ivan Illich. He is an ideal student. Full of goodwill toward his teachers and his peers, he needs no discipline. He studies because he is naturally curious. He learns things because they interest him.

Unfortunately, he is imaginary. He was quite explicitly so with Rousseau, who put his own children in an orphanage and preferred to say how he would teach his fictional hero, but the modern version of the free and happy student to be found in books by Paul Goodman, John Holt, Jonathan Kozol, or Charles Silberman is also imaginary. Occasionally a real example seems to turn up. There are teachers who would be successful in dealing with people anywhere – as statesmen, therapists, businessmen, or friends – and there are students who scarcely need to be taught, and together they sometimes seem to bring Emile to life. And unfortunately they do so just often enough to sustain the old dream. But Emile is a will o' the wisp, who has led many teachers into a conception of their role which could prove disastrous.

The student who has been taught *as if he were Emile* is, however, almost too painfully real. It has taken a long time for him to make his appearance. Children were first made free and happy in kindergarten, where there seemed to be no danger in freedom, and for a long time they were found nowhere else, because the rigid discipline of the grade schools blocked progress. But eventually they broke through – moving from kindergarten into grade school, taking over grade after grade, moving into secondary school and on into college and, very recently, into graduate school. Step by step they have insisted upon their rights, justifying their demands with the slogans that philosophers of education have supplied. If sitting in rows restricts personal freedom, unscrew the seats. If order can be maintained only through coercion, let chaos reign. If one cannot be really free while worrying about examinations and grades, down with examinations and grades! The whole Establishment is now awash with free and happy students.

Dropping Out of School, Dropping Out of Life

If they are what Rousseau's Emile would really have been like, we must confess to some disappointment. The Emile we know doesn't work very hard. "Curiosity" is evidently a moderate sort of thing. Hard work is frowned upon because it implies a "work ethic," which has something to do with discipline.

The Emile we know doesn't learn very much. His "interests" are evidently of limited scope. Subjects that do not appeal to him he calls irrelevant. (We should not be surprised at this, since Rousseau's

PHI DELTA KAPPAN. 1973. Vol. 55, pp. 13-16.

Emile, like the boys in Summerhill, never got past the stage of knowledgeable craftsman.) He may defend himself by questioning the value of knowledge. Knowledge is always in flux, so why bother to acquire any particular stage of it? It will be enough to remain curious and interested. In any case the life of feeling and emotion is to be preferred to the life of intellect, let us be governed by the heart rather than the head.

The Emile we know doesn't think very clearly. He has had little or no chance to learn to think logically or scientifically and is easily taken in by the mystical and the superstitious. Reason is irrelevant to feeling and emotion.

And, alas, the Emile we know doesn't seem particularly happy. He doesn't like his education any more than his predecessors liked theirs. Indeed, he seems to like it less. He is much more inclined to play truant (big cities have given up enforcing truancy laws), and he drops out as soon as he legally can, or a little sooner. If he goes to college, he probably takes a year off at some time in his four-year program. And after that his dissatisfaction takes the form of anti-intellectualism and a refusal to support education.

Are there offsetting advantages? Is the free and happy student less aggressive, kinder, more loving? Certainly not toward the schools and teachers that have set him free, as increasing vandalism and personal attacks on teachers seem to show. Nor is he particularly well disposed toward his peers. He seems perfectly at home in a world of unprecedented domestic violence.

Is he perhaps more creative? Traditional practices were said to suppress individuality, what kind of individuality has now emerged? Free and happy students are certainly different from the students of a generation ago, but they are not very different from each other. Their own culture is a severely regimented one, and their creative works - in art, music, and literature - are confined to primitive and elemental materials. They have very little to be creative with, for they have never taken the trouble to explore the fields in which they are now to be front-runners.

Is the free and happy student at least more effective as a citizen? Is he a better person? The evidence is not very reassuring. Having dropped out of school, he is likely to drop out of life too. It would be unfair to let the hippie culture represent young people today, but it does serve to clarify an extreme. The members of that culture do not accept responsibility for their own lives, they sponge on the contributions of those who have not yet been made free and happy who have gone to medical school and become doctors, or who have become the farmers who raise the food or the workers who produce the goods they consume.

These are no doubt overstatements. Things are not that bad, nor is education to be blamed for all the trouble. Nevertheless, there is a trend in a well-defined direction, and it is particularly clear in education. Our failure to create a truly free and happy student is symptomatic of a more general problem.

The Illusion of Freedom

What we may call the struggle for freedom in the Western world can be analyzed as a struggle to escape from or avoid punitive or coercive treatment. It is characteristic of the human species to act in such a way as to reduce or terminate irritating, painful, or dangerous stimuli, and the struggle for freedom has been directed toward those who would control others with stimuli of that sort. Education has had a long and shameful part in the history of that struggle. The Egyptians, Greeks, and Romans all whipped their students. Medieval sculpture showed the carpenter with his hammer and the schoolmaster with the tool of his trade too, and it was the cane or rod. We are not yet in the clear. Corporal punishment is still used in many schools, and there are calls for its return where it has been abandoned.

A system in which students study primarily to avoid the consequences of not studying is neither humane nor very productive. Its by-products include truancy, vandalism, and apathy. Any effort to eliminate punishment in education is certainly commendable. We ourselves act to escape from aversive control, and our students should escape from it too. They should study because they want to, because they like to, because they are interested in what they are doing. The mistake - a classical mistake in the literature of freedom - is to suppose that they will do so as soon as we stop punishing them. Students are not literally free when they have been freed from their teachers. They then simply come under the control of other conditions, and we must look at those conditions and their effects if we are to improve teaching.

Those who have attacked the "servility" of students, as Montessori called it, have often put their faith in the possibility that young people will learn what they need to know from the "world of things," which includes the world of people who are not teachers. Montessori saw possibly useful behavior being suppressed by schoolroom discipline. Could it not be salvaged? And could the environment of the schoolroom not be changed so that other useful behavior would occur? Could the teacher not simply guide the student's natural development? Or could he not accelerate it by teasing out behavior which would occur naturally but not so quickly if he did not help? In other words, could we not bring the real world into the classroom, as John Dewey put it, or destroy the classroom and turn the student over to the real world, as Ivan Illich has recommended? All these possibilities can be presented in an attractive light, but they neglect two vital points.

1. No one learns very much from the real world without help. The only evidence we have of what can be learned from a nonsocial world has been supplied by those wild boys said to have been raised without contact with other members of their own species. Much more can be learned without formal instruction in a social world, but not without a good deal of teaching, even so. Formal education has made a tremendous difference in the extent of the skills and knowledge which can be acquired by a person in a single lifetime.

2. A much more important principle is that the real world teaches only what is relevant to the present; it makes no explicit preparation for the future. Those who would naturalize teaching have contended that no preparation is needed, that the student will follow a natural line of development and move into the future in the normal course of events. We should be content, as Carl Rogers has put it, to trust

> . . . the insatiable curiosity which drives the adolescent boy to absorb everything he can see or hear or read about gasoline engines in order to improve the efficiency and speed of his "hot rod." I am talking about the student who says, "I am discovering, drawing in from the outside, and making that which is drawn in a real part of me." I am talking about my learning in which the experience of the learner progresses along the line: "No, no, that's not what I want"; "Wait! This is closer to what I'm interested in, what I need." "Ah, here it is! Now I'm grasping and comprehending what I need and what I want to know!"[1]

Rogers is recommending a total commitment to the present moment, or at best to an immediate future.

Formal Education as Preparation for the Future

But it has always been the task of formal education to set up behavior which would prove useful or enjoyable *later* in the student's life. Punitive methods had at least the merit of providing current reasons for learning things that would be rewarding in the future. We object to the punitive reasons, but we should not forget their function in making the future important.

It is not enough to give the student advice — to explain that he will have a future, and that to enjoy himself and be more successful in it, he must acquire certain skills and knowledge now. Mere advice is ineffective because it is not supported by current rewards. The positive consequences that generate a useful behavioral repertoire need not be any more explicitly relevant to the future than were the punitive consequences of the past. The student needs current reasons, positive or negative, but only the educational policy maker who supplies them need take the future into account. It follows that many instructional arrangements seem "contrived," but there is nothing wrong with that. It is the teacher's function to contrive conditions under which students learn. Their relevance to a future usefulness need not be obvious.

It is a difficult assignment. The conditions the teacher arranges must be powerful enough to compete with those under which the student tends to behave in distracting ways. In what has come to be called "contingency management in the classroom," tokens are sometimes used as rewards or reinforcers. They become reinforcing when they are exchanged for reinforcers that are already effective. There is no "natural" relation between what is learned and what is received. The token is simply a reinforcer that can be made clearly contingent upon behavior. To straighten out a wholly disrupted classroom, something as obvious as a token economy may be needed, but less conspicuous contingencies — as in a credit-point system, perhaps, or possibly in the long run merely expressions of approval on the part of teacher or peer — may take over.

The teacher can often make the change from punishment to positive reinforcement in a surprisingly simple way — by responding to the student's success rather than his failures. Teachers have too often supposed that their role is to point out what students are doing wrong, but pointing to what they are doing *right* will often make an enormous difference in the atmosphere of a classroom and in the efficiency of instruction. Programmed materials are helpful in bringing about these changes, because they increase the frequency with which the student enjoys the satisfaction of being right, and they supply a valuable intrinsic reward in providing a clear indication of progress. A good program makes a step in the direction of competence almost as conspicuous as a token.

Programmed instruction is perhaps most successful in attacking punitive methods by allowing the student to move at his own pace. The slow student is released from the punishment which inevitably follows when he is forced to move on to material for which he is not ready, and the fast student escapes the boredom of being forced to go too slow. These principles have recently been extended to college education, with dramatic results, in the Keller system of personalized instruction.[2]

The Responsibility of Setting Educational Policy

There is little doubt that a student can be given nonpunitive reasons for acquiring behavior that will become useful or otherwise reinforcing at some later date. He can be prepared for the future. But what *is* that future? Who is to say what the student should learn? Those who have sponsored the free and happy student have argued that it is the student himself who should say. His current interests should be the source of an effective educational policy. Certainly they will reflect his idiosyncrasies, and that is good, but how much can he know about the world in which he will eventually play a part? The things he is "naturally" curious about are of current and often temporary interest. How many things must he possess besides his "hot rod" to provide the insatiable curiosity relevant to, say, a course in physics?

It must be admitted that the teacher is not always in a better position. Again and again education has gone out of date as teachers have continued to teach subjects which were no longer relevant at any time in the student's life. Teachers often teach simply what they know. (Much of what is taught in private schools is determined by what the available teachers can teach.) Teachers tend to teach what they can teach easily. Their current interests, like those of students, may not be a reliable guide.

Nevertheless, in recognizing the mistakes that have been made in the past in specifying what students are to learn, we do not absolve ourselves from the responsibility of setting educational policy. We should say, we should be *willing* to say, what we believe students will need to know, taking the individual student into account wherever possible, but otherwise making our best prediction with respect to students in general. Value judgments of this sort are not as hard to make as is often argued. Suppose we undertake to prepare the student to produce his share of the goods he will consume and the services he will use, to get on well with his fellows, and to enjoy his life. In doing so are we imposing *our* values on someone else? No, we are merely choosing a set of specifications which, so far as we can tell, will at some time in the future prove valuable to the student and his culture. Who is any more likely to be right?

The natural, logical outcome of the struggle for personal freedom in education is that the teacher should improve his control of the student rather than abandon it. The free school is no school at all. Its philosophy signalizes the abdication of the teacher. The teacher who understands his assignment and is familiar with the behavioral processes needed to fulfill it can have students who not only feel free and happy while they are being taught but who will continue to feel free and happy when their formal education comes to an end. They will do so because they will be successful in their work (having acquired useful productive repertoires), because they will get on well with their fellows (having learned to understand themselves and others), because they will enjoy what they do (having acquired the necessary knowledge and skills), and because they will from time to time make an occasional creative contribution toward an even more effective and enjoyable way of life. Possibly the most important consequence is that the teacher will then feel free and happy too.

We must choose today between Cassandran and Utopian prognostications. Are we to work to avoid disaster or to achieve a better world? Again, it is a question of punishment or reward. Must we act because we are frightened, or are there positive reasons for changing our cultural practices? The issue goes far beyond education, but it is one with respect to which education has much to offer. To escape from or avoid disaster, people are likely to turn to the punitive measures of a police state. To work for a better world, they may turn instead to the positive methods of education. When it finds its most effective methods, education will be almost uniquely relevant to the task of setting up and maintaining a better way of life.

---

1. Carl R. Rogers, *Freedom to Learn* (Columbus, O.: Merrill, 1969).

2. *PSI Newsletter*, October, 1972 (published by Department of Psychology, Georgetown University, J. G. Sherman, ed.).

The technology of teaching

BY B. F. SKINNER

REFERENCES

More than 60 years ago, in his *Talks to teachers on psychology*, William James (1899) said: 'You make a great, a very great mistake, if you think that psychology, being the science of the mind's laws, is something from which you can deduce definite programs and schemes and methods of instruction for immediate schoolroom use. Psychology is a science, and teaching is an art, and sciences never generate arts directly out of themselves. An intermediary inventive mind must make the application, by using its originality.' In the years which followed, educational psychology and the experimental psychology of learning did little to prove him wrong. As late as 1962, an American critic, Jacques Barzun (1962), asserted that James's book still contained 'nearly all that anyone need know of educational method'.

Speaking for the psychology of his time James was probably right, but Barzun was clearly wrong. A special branch of psychology, the so called experimental analysis of behaviour, has produced if not an art at least a technology of teaching from which one can indeed 'deduce programs and schemes and methods of instruction'. The public is aware of this technology through two of its products, teaching machines and programmed instruction. Their rise has been meteoric. Within a single decade hundreds of instructional programmes have been published, many different kinds of teaching machines have been offered for sale, and societies for programmed instruction have been founded in a dozen countries. Unfortunately, much of the technology has lost contact with its basic science.

Teaching machines are widely misunderstood. It is often supposed that they are simply devices which mechanize functions once served by human teachers. Testing is an example. The teacher must discover what the student has learned and can do so with the help of machines, the scoring of multiple choice tests by machine is now common. Nearly 40 years ago Sidney Pressey (1926) pointed out that a student learned something when told whether his answers are right or wrong and that a *self*-scoring machine could therefore teach. Pressey assumed that the student had studied a subject before coming to the testing machine, but some modern versions also present the material on which the student is to be tested. They thus imitate, and could presumably replace, the teacher. But holding a student responsible for assigned material is not teaching, even though it is a large part of modern school

PROCEEDINGS OF THE ROYAL SOCIETY, 1965, Vol. 162, pp. 427-443.

and university practice. It is simply a way of inducing the student to learn without being taught.

Some so-called teaching machines serve another conspicuous function of the teacher: they are designed primarily to attract and hold attention. The television screen is praised for its hypnotic powers. A machine has recently been advertised which holds the student's head between earphones and his face a few inches from a brightly lit text. It is intended that he will read a few lines, then listen to his recorded voice as he reads them over again—all in the name of 'concentration.' Machines also have the energy and patience needed for simple exercise or drill. Many language laboratories take the student over the same material again and again, as only a dedicated private tutor could do, on some theory of 'automaticity.'

These are all functions which should never have been served by teachers in the first place, and mechanizing them is small gain.

The programming of instruction has also been widely misunderstood. The first programmes emerging from an experimental analysis of behaviour were copied only in certain superficial aspects. Educational theorists could assimilate the principles they appeared to exemplify to earlier philosophies. Programmed instruction, for example, has been called Socratic. The archetypal pattern is the famous scene in the *Meno* in which Socrates takes the slave boy through Pythagoras's theorem on doubling the square. It is one of the great frauds in the history of education. Socrates asks the boy a long series of leading questions and, although the boy makes no response which has not been carefully prepared, insists that he has told him nothing. In any case the boy has learned nothing, he could not have gone through the proof by himself afterwards, and Socrates says as much later in the dialogue. Even if the boy had contributed something to the proof by way of a modest original discovery, it would still be wrong to argue that his behaviour in doing so under Socrates's careful guidance resembled Pythagoras's original unguided achievement.

Other supposed principles of programming have been found in the writings of Comenius in the seventeenth century—for example, that the student should not be asked to take a step he cannot take—and in the work of the American psychologist, E. L. Thorndike, who more than 50 years ago pointed to the value of making sure that the student understood one page of a text before moving on to the next. A good programme does lead the student step by step, each step is within his range, and he usually understands it before moving on, but programming is much more than this. What it is, and how it is related to teaching machines, can be made clear only by returning to the experimental analysis of behaviour which gave rise to the movement.

An important process in human behaviour is attributed, none too accurately, to 'reward and punishment.' Thorndike described it in his Law of Effect. It is now commonly referred to as 'operant conditioning'—not to be confused with the conditioned reflexes of Pavlov. The essentials may be seen in a typical experimental arrangement. Figure 1, plate 50, shows a hungry rat in an experimental space which contains a food dispenser. A horizontal bar at the end of a lever projects from one wall. Depression of the lever operates a switch. When the switch is connected with the food dispenser, any behaviour on the part of the rat which depresses the lever is, as we say, 'reinforced with food.' The apparatus simply makes the appearance of food *contingent upon* the occurrence of an arbitrary bit of behaviour. Under such circumstances the probability that a response to the lever will occur again is increased (Skinner 1938).

The basic contingency between an act and its consequences has been studied over a fairly wide range of species. Pigeons have been reinforced for pecking at transilluminated disks (figure 2, plate 51), monkeys for operating toggle switches which were first designed for that more advanced primate, man, and so on. Reinforcers which have been studied include water, sexual contact, the opportunity to act aggressively, and with human subjects approval of one's fellow men and the universal generalized reinforcer, money.

The relation between a response and its consequences may be simple, and the change in probability of the response is not surprising. It may therefore appear that research of this sort is simply proving the obvious. A critic has recently said that King Solomon must have known all about operant conditioning because he used rewards and punishment. In the same sense his archers must have known all about Hooke's Law because they used bows and arrows. What is technologically useful in operant conditioning is our increasing knowledge of the extraordinarily subtle and complex properties of behaviour which may be traced to subtle and complex features of the contingencies of reinforcement which prevail in the environment.

We may arrange matters, for example, so that the rat will receive food only when it depresses the lever with a given force. Weaker responses then disappear, and exceptionally forceful responses begin to occur and can be selected through further differential reinforcement. Reinforcement may also be made contingent upon the presence of stimuli. depression of the lever operates the food dispenser, for example, only when a tone of a given pitch is sounding. As a result the rat is much more likely to respond when a tone of that pitch is sounding. Responses may also be reinforced only intermittently. Some common schedules of reinforcement are the subject of probability theory. Gambling devices often provide for the reinforcement of varying numbers of responses in an unpredictable sequence. Comparable schedules are programmed in the laboratory by interposing counters between the operandum and the reinforcing device. The extensive literature on schedules of reinforcement (see, for example, Ferster & Skinner 1957) also covers intermittent reinforcement arranged by clocks and speedometers.

A more complex experimental space contains two operanda -two levers to be pressed, for example, or two disks to be pecked. Some of the resulting contingencies are the subject of decision-making theory. Responses may also be chained together, so that responding in one way produces the opportunity to respond in another. A still more complex experimental space contains two organisms with their respective operanda and with interlocking schedules of reinforcement. Game theory is concerned with contingencies of this sort. The study of operant behaviour, however, goes beyond an analysis of possible contingencies to the behaviour generated.

The application of operant conditioning to education is simple and direct. Teaching is the arrangement of contingencies of reinforcement under which students learn. They learn without teaching in their natural environments, but teachers arrange special contingencies which expedite learning, hastening the appearance of behaviour which would otherwise be acquired slowly or making sure of the appearance of behaviour which might otherwise never occur.

A teaching machine is simply any device which arranges contingencies of reinforcement. There are as many different kinds of machines as there are different kinds of contingencies. In this sense the apparatuses developed for the experimental analysis of behaviour were the first teaching machines. They remain much more complex and subtle than the devices currently available in education - a state of affairs to be regretted by anyone who is concerned with making education as effective as possible. Both the basic analysis and its technological applications require instrumental aid. Early experimenters manipulated stimuli and reinforcers and recorded responses by hand, but current research without the help of extensive apparatus is unthinkable. The teacher needs similar instrumental support, for it is impossible to arrange many of the contingencies of reinforcement which expedite learning without it. Adequate apparatus has not eliminated the researcher, and teaching machines will not eliminate the teacher. But both teacher and researcher must have such equipment if they are to work effectively.

Programmed instruction also made its first appearance in the laboratory in the form of programmed contingencies of reinforcement. The almost miraculous power to change behaviour which frequently emerges is perhaps the most conspicuous contribution to date of an experimental analysis of behaviour. There are at least four different kinds of programming. One is concerned with generating new and complex patterns or 'topographies' of behaviour. It is in the nature of operant conditioning that a response cannot be reinforced until it has occurred. For experimental purposes a response is chosen which presents no problem (a rat is likely to press a sensitive lever within a short time), but we could easily specify responses which never occur in this way. Can they then never be reinforced?

The programming of a rare topography of response is sometimes demonstrated in the classroom in the following way. A hungry pigeon is placed in an enclosed space where it is visible to the class. A food dispenser can be operated with a handswitch held by the demonstrator. The pigeon has learned to eat from the food dispenser without being disturbed by its operation, but it has not been conditioned in any other way. The class is asked to specify a response which is not part of the current repertoire of the pigeon. Suppose, for example, it is decided that the pigeon is to pace a figure eight. The demonstrator cannot simply wait for this response to occur and then reinforce it. Instead he reinforces any current response which may contribute to the final pattern possibly simply turning the head or taking a step in, say, a clockwise direction. The reinforced response will quickly be repeated (one can actually see learning take place under these circumstances), and reinforcement is then withheld until a more marked movement in the same direction is made. Eventually only a complete turn is reinforced. Similar responses in a counterclockwise direction are then strengthened, the clockwise movement suffering partial extinction.' When a complete counterclockwise movement has thus been 'shaped', the clockwise turn is reinstated, and eventually the pigeon makes both turns in succession and is reinforced. The whole pattern is then quickly repeated.

Q.E.D. The process of 'shaping' a response of this complexity should take no more than five or ten minutes. The demonstrator's only contact with the pigeon is by way of the handswitch, which permits him to determine the exact moment of operation of the food dispenser. By selecting responses to be reinforced he improvises a programme of contingencies, at each stage of which a response is reinforced which makes it possible to move on to a more demanding stage. The contingencies gradually approach those which generate the final specified response.

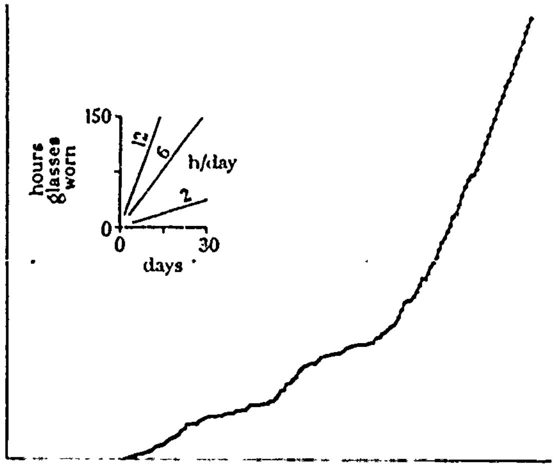


FIGURE 3. Curve showing the number of hours per day during which glasses were worn, plotted cumulatively. The final slope is about twelve hours per day.

This method of shaping a topography of response has been used by Wolf, Mees & Risley (1964) to solve a difficult behaviour problem. A boy was born blind with cataracts. Before he was of an age at which an operation was feasible, he had begun to display severe temper tantrums, and after the operation he remained unmanageable. It was impossible to get him to wear the glasses without which he would soon become permanently blind. His tantrums included serious self destructive behaviour, and he was admitted to a hospital with a diagnosis of 'child schizophrenia.' Two principles of operant conditioning were applied. The temper tantrums were extinguished by making sure that they were never followed by reinforcing consequences. A programme of contingencies of reinforcement was then designed to shape the desired behaviour of wearing glasses. It was necessary to allow the child to go hungry so that food could be used as an effective reinforcer. Empty glasses frames were placed about the room and any response which made contact with them was reinforced with food. Reinforcement was then made contingent on picking up the frames, carrying them about, and so on, in a programmed sequence. Some difficulty was encountered in shaping the response of putting the frames on the face in the proper position. When this was eventually achieved, the prescription lenses were put in the frames. Wolf *et al* publish a cumulative curve (figure 3) showing the number of hours per day the glasses were worn. The final slope represents essentially all the child's waking hours.

Operant techniques were first applied to psychotic subjects in the pioneering work of Lindsley (1960). Azrin and others have programmed contingencies of reinforcement to solve certain management problems in institutions for the psychotic (Ayllon & Azrin 1965). The techniques are not designed to cure psychoses but to generate trouble free behaviour. In one experiment a whole ward was placed on an economic basis. Patients were reinforced with tokens when they behaved in ways which made for simpler management, and in turn paid for services received, such as meals or consultations with psychiatrists. Such an economic system, like any economic system in the world at large, represents a special set of terminal contingencies which in neither system guarantee appropriate behaviour. The contingencies must be made effective by appropriate programmes.

A second kind of programming is used to alter temporal or intensive properties of behaviour. By differentially reinforcing only the more vigorous instances in which a pigeon pecks a disk and by advancing the minimum requirement very slowly, a pigeon can be induced to peck so energetically that the base of its beak becomes inflamed. If one were to begin with this terminal contingency, the behaviour would never develop. There is nothing new about the necessary programming. An athletic coach may train a high jumper simply by moving the bar higher by small increments, each setting permitting some successful jumps to occur. But many intensive and temporal contingencies such as those seen in the arts, crafts, and music are very subtle and must be carefully analysed if they are to be properly programmed.

Another kind of programming is concerned with bringing behaviour under the control of stimuli. We could determine a rat's sensitivity to tones of different pitches by reinforcing responses made when one tone is sounding and extinguishing all responses made when other tones are sounding. We may wish to avoid extinction, however, the organism is to acquire the discrimination without making any 'errors.' An effective procedure has been analysed by Terrace (1963). Suppose we are to condition a pigeon to peck a red disk but not a green. If we simply reinforce it for pecking the red disk, it will almost certainly peck the green as well and these 'errors' must be extinguished. Terrace begins with disks which are as different as possible. One is illuminated by a red light, but the other is dark. Although reinforced for pecking the red disk, the pigeon is not likely to peck the dark disk, at least during a period of a few seconds. When the disk again becomes red, a response is immediately made. It is possible to extend the length of time the disk remains dark. Eventually the pigeon pecks the red disk instantly, but does not peck the dark disk no matter how long it remains dark. The important point is that it has never pecked the dark disk at any time.

A faint green light is then added to the dark disk. Over a period of time the green light becomes brighter and eventually is as bright as the red. The pigeon now responds instantly to the red disk but not to the green *and has never responded to the green.*

A second and more difficult discrimination can then be taught without errors by transferring control from the red and green disks. Let us say that the pigeon is to respond to a white vertical bar projected on a black disk but not to a horizontal. These patterns are first superimposed upon red and green backgrounds, and the pigeon is reinforced when it responds to red vertical but not to green horizontal. The intensity of the colour is then slowly reduced. Eventually the pigeon responds to the black and white vertical bar, does not respond to the black and white horizontal bar, *and has never done so.* The result could perhaps be achieved more rapidly by permitting errors to occur and extinguishing them, but other issues may need to be taken into account. When extinction is used, the pigeon shows powerful emotional responses to the wrong stimulus, when the Terrace technique is used it remains quite indifferent. It is, so to speak, 'not afraid of making a mistake'. The difference is relevant to education, where the anxiety generated by current methods constitutes a serious problem. There are those who would defend a certain amount of anxiety as a good thing, but we may still envy the occasionally happy man who readily responds when the occasion is appropriate but is otherwise both emotionally and intellectually disengaged. The important point is that the terminal contingencies controlling the behaviour of both anxious and nonanxious students are the same; the difference is to be traced to the programme by way of which the terminal behaviour has been reached.

The discriminative capacities of lower organisms have been investigated with methods which require very skilful programming. Blough (1956), for example, has developed a technique in which a pigeon maintains a spot of light at an intensity at which it can just be seen. By using a range of monochromatic lights he has shown that the spectral sensitivity of the pigeon is very close to that of man. Several other techniques are available which make it possible to use lower organisms as sensitive psychophysical observers. They are available, however, only to those who understand the principles of programming.

Some current work by Murray Sidman provides a dramatic example of programming a subtle discrimination in a microcephalic idiot. At the start of the experiment Sidman's subject (figure 4, plate 51) was 40 years old. He was said to have a mental age of about 18 months. He was partially toilet trained and dressed himself with help. To judge from the brain of his sister, now available for post-mortem study, his brain is probably about one third the normal size. Sidman investigated his ability to discriminate circular forms projected on translucent vertical panels. Small pieces of chocolate were used as reinforcers. At first any pressure against a single large vertical panel (figure 5A) operated the device which dropped a bit of chocolate into a cup within reach. Though showing relatively poor motor co ordination, the subject eventually executed the required, rather delicate response. The panel was then subdivivded into a three by three set of smaller panels (to be seen in figure 4, plate 51, and represented schematically in figure 5B), the central panel not being used in what follows. The subject was first reinforced when he pressed any of the eight remaining panels. A single panel was then lit at random, a circle being projected on it (figure 5C). The subject learned to press the lighted panel. Flat ellipses were then projected on the other panels at a low illumination (figure 5D). In subsequent settings the ellipses, now brightly illuminated, progressively approached circles (figure 5 *E* to *G*). Each stage was maintained until the subject had formed the necessary discrimination, all correct responses being reinforced with chocolate. Eventually the subject could successfully select a circle from an array approximately like that shown in figure 5H. Using similar shaping techniques Sidman and his associates have conditioned the subject to pick up and use a pencil appropriately, tracing letters faintly projected on a sheet of paper.

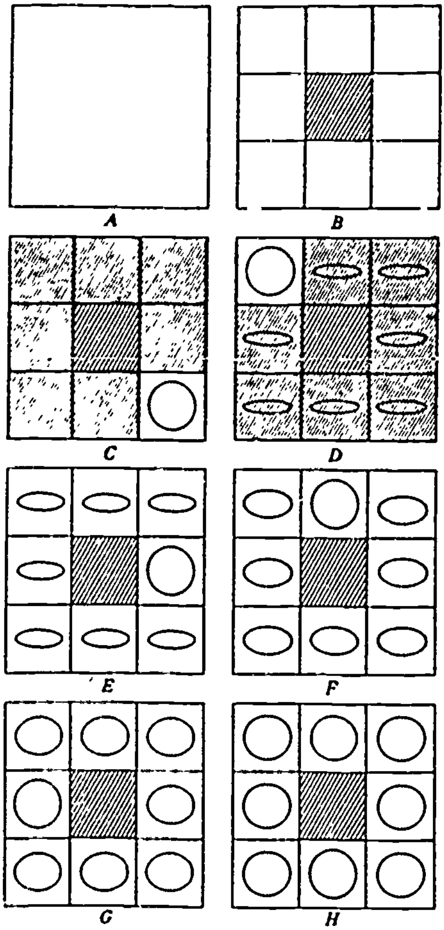


FIGURE 5. A programme designed to teach subtle form discrimination. Reinforcement was contingent on: (*A*) a response moving a large panel, (*B*) a response moving any one of nine smaller panels (with the exception of the centre panel), (*C*) a response moving only the one panel on which a circle is projected, (*D*) as before except that flat ellipses appear faintly on the other panels, (*E,F,G*) a response to the panel bearing a circle, appearing in random position among ellipses the shorter axis of which is progressively lengthening; (*H*) a response to the panel bearing a circle among ellipses closely approximating circles.

The intellectual accomplishments of this microcephalic idiot in the forty-first year of his life have exceeded all those of his first 40 years. They were possible only because he has lived a few hours of each week of that year in a well programmed environment. No very bright future beckons (he has already lived longer than most people of his kind), and it is impossible to say what he might have achieved if he had been subject to a similar programme from birth, but he has contributed to our knowledge by demonstrating the power of a method of instruction which could scarcely be tested on a less promising case. (The bright futures belong to the normal and exceptional children who will be fortunate enough to live in environments which have been designed to maximize *their* development, and of whose potential achievements we have now scarcely any conception.)

A fourth kind of programming has to do with maintaining behaviour under infrequent reinforcement. A pigeon will continue to respond even though only one response in every hundred, say, is reinforced, but it will not do so unless the contingencies have been programmed. A fresh pigeon is no more likely to peck a disk a hundred times than to pace a figure eight. The behaviour is built up by reinforcing every response, then every other response, then every fifth response, and so on, waiting at each stage until the behaviour is reasonably stable. Under careful programming pigeons have continued to respond when only every ten-thousandth response has been reinforced, and this is certainly not the limit. An observer might say that the pigeon is 'greatly interested in his work', 'industrious', 'remarkably tolerant to frustration', 'free from discouragement', 'dedicated to his task', and so on. These expressions are commonly applied to students who have had the benefit of similar programming, accidental or arranged.

The effective scheduling of reinforcement is an important element in educational design. Suppose we wish to teach a student to read 'good books' -books which, almost by definition, do not reinforce the reader sentence by sentence or even paragraph by paragraph but only when possibly hundreds of pages have prepared him for a convincing or moving dénouement. The student must be exposed to a programme of materials which build up a tendency to read in the absence of reinforcement. Such programmes are seldom constructed deliberately and seldom arise by accident, and it is therefore not surprising that few students even in good universities learn to read books of this sort and continue to do so for the rest of their lives. In their pride, schools are likely to arrange just the wrong conditions, they are likely to maintain so-called 'standards' under which books are forced upon students before they have had adequate preparation.

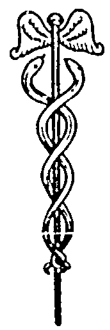
Other objectives in education need similar programming. The dedicated scientist who works for years in spite of repeated failures is often looked upon as a happy accident, but he may well be the product of a happy if accidental history of reinforcement. A programme in which exciting results were first common but became less and less frequent could generate the capacity to continue in the absence of reinforcement for long periods of time. Such programmes should arise naturally as scientists turn to more and more difficult areas. Perhaps not many effective programmes are to be expected for this reason, and they are only rarely designed by teachers of science. This may explain why there are so few dedicated scientists. Maintaining a high level of activity is one of the more important achievements of programming. Repeatedly, in its long history, education has resorted to aversive control to keep its students at work. A proper understanding of the scheduling of reinforcement may lead at long last to a better solution of this problem.

Let us look at these principles of programming at work in one or two traditional educational assignments. Instruction in handwriting will serve as one example. To say that a child is to learn 'how to write' tells us very little. The so-called signs of 'knowing how to write' provide a more useful set of behavioural specifications. The child is to form letters and words which are legible and graceful according to taste. He is to do this first in copying a model, then in writing to dictation (or self-dictation as he spells out words he would otherwise speak), and eventually in writing as a separate nonvocal form of verbal behaviour. A common method is to ask the child to copy letters or words and to approve or otherwise reinforce his approximations to good copy. More and more exact copies are demanded as the hand improves—in a crude sort of programming. The method is ineffective largely because the reinforcements are too long deferred. The parent or teacher comments upon or corrects the child's work long after it has been performed.

A possible solution is to teach the child to discriminate between good and bad form before he starts to write. Acceptable behaviour should then generate immediate, automatic self reinforcement. This is seldom done. Another possibility is to make reinforcement immediately contingent upon successful responses. One method now being tested is to treat paper chemically so that the pen the child uses writes in dark blue when a response is correct and yellow when it is incorrect. The dark blue line is made automatically reinforcing through generous commendation. Under such contingencies the proper execution of a letter can be programmed; at first the child makes a very small contribution in completing a letter, but through progressive stages he approaches the point at which he composes the letter as a whole, the chemical response of the paper differentially reinforcing good form throughout. The model to be copied is then made progressively less important by separating it in both time and space from the child's work. Eventually words are written to dictation, letter by letter, in spelling dictated words, and in describing pictures. The same kind of differential reinforcement can be used to teach good form, proper spacing, and so on. The child is eventually forming letters skilfully under continuous automatic reinforcement. The method is directed as much toward motivation as toward good form. Even quite young children remain busily at work for long periods of time without coercion or threat, showing few signs of fatigue, nervousness, or other forms of escape.

As a second example we may consider the acquisition of a simple form of verbal behaviour. A behavioural specification is here likely to be especially strongly resisted. It is much more in line with traditional educational policy to say that the student is to 'know facts, understand principles, be able to put ideas into words, express meanings, or communicate information.' In *Verbal behaviour* (Skinner 1957) I tried to show how the behaviour exhibited in such activities could be formulated without reference to ideas, meanings, or information, and many of the principles currently used in programming verbal knowledge have been drawn from that analysis. The field is too large to be adequately covered here, but two examples may suggest the direction of the approach.

What happens when a student memorizes a poem? Let us say that he begins by reading the poem from a text. His behaviour is at that time under the control of the text, and it is to be accounted for by examining the process through which he has learned to read. When he eventually speaks the poem in the absence of a text, the same form of verbal behaviour has come under the control of other stimuli. He may begin to recite when asked to do so — he is then under control of an external verbal stimulus — but, as he continues to recite, his behaviour comes under the control of stimuli he himself is generating (not necessarily in a crude word-by-word chaining of responses). In the process of 'memorizing' the poem, control passes from one kind of stimulus to another.



caduceus

FIGURE 6

A classroom demonstration of the transfer of control from text to self generated stimuli illustrates the process. A short poem is projected on a screen or written on a chalkboard. A few unnecessary letters are omitted. The class reads the poem in chorus. A second slide is then projected in which other letters are missing (or letters erased from the chalkboard). The class could not have read the poem correctly if this form had been presented first, but because of its recent history it is able to do so. (Some members undoubtedly receive help from others in the process of choral reading.) In a third setting still other letters are omitted, and after a series of five or six settings the text has completely disappeared. The class is nevertheless able to 'read' the poem. Control has passed mainly to self generated stimuli.

As another example, consider what a student learns when he consults an illustrated dictionary. After looking at a labelled picture, as in figure 6, we say that he knows something he did not know before. This is another of those vague expressions which have done so much harm to education. The 'signs or symptoms of such knowledge' are of two sorts. Shown the picture in figure 6 without the text the student can say 'caduceus' (we say that he now knows what the object pictured in the figure is called) or, shown the word *caduceus*, he can now describe or reconstruct the picture (we say that he now knows what the word *caduceus* means). But what has actually happened?

The basic process is similar to that of transferring discriminative control in the Terrace experiment. To begin with, the student can respond to the picture in various ways: he can describe it without naming it, he can find a similar picture in an array, he can draw a fair copy, and so on. He can also speak the name by reading the printed word. When he first looks at the picture and reads the word, his verbal response is primarily under the control of the text, but it must eventually be controlled by the picture. As in transferring the control exerted by red and green to vertical and horizontal lines, we can change the control efficiently by making the text gradually less important, covering part of it, removing some of the letters, or fogging it with a translucent mask. As the picture acquires control the student can speak the name with less and less help from the text. Eventually, when the picture exerts enough control, he 'knows the name of the pictured object.' The normal student can learn the name of one object so quickly that the 'vanishing' technique may not be needed, but it is a highly effective procedure in learning the names of a large number of objects. The good student learns how to make progressive reductions in the effectiveness of a text by himself. he may glance at the text out of the corner of his eye, uncover it bit by bit, and so on. In this way he improvises his own programme in making the text less and less important as the picture acquires control of the verbal response.

In teaching the student 'the meaning of the word *caduceus*' we could slowly obscure the picture, asking the student to respond to the name by completing a drawing or description or by finding a matching picture in an array. Eventually in answer to the question *What is a caduceus?* he describes the object, makes a crude sketch, or points to the picture of a caduceus. The skilful student uses techniques of this sort in studying unprogrammed material.

'Knowing what a caduceus is' or 'knowing the meaning of the word caduceus' is probably more than responding in these ways to picture or text. In other words, there are other 'signs of knowledge.' That is one reason why the concept of knowledge is so inadequate. But other relevant behaviour must be taught, if at all, in substantially the same way.

These examples do scant justice to the many hundreds of effective programmes now available or to the techniques which many of them use so effectively, but they must suffice as a basis for discussing a few general issues. An effective technology of teaching, derived not from philosophical principles but from a realistic analysis of human behaviour, has much to contribute, but as its nature has come to be clearly seen, strong opposition has arisen.

A common objection is that most of the early work responsible for the basic formulation of behaviour was done on so called lower animals. It has been argued that the procedures are therefore appropriate only to animals and that to use them in education is to treat the student like an animal. So far as I know, no one argues that because something is true of a pigeon, it is therefore true of a man. There are enormous differences in the topographies of the behaviours of man and pigeon and in the kinds of environmental events which are relevant to that behaviour— differences which, if anatomy and physiology were adequate to the task, we could probably compare with differences in the mediating substrata but the basic processes in behaviour, as in neural tissue, show helpful similarities. Relatively simple organisms have many advantages in early stages of research, but they impose no limit on that research. Complex processes are met and dealt with as the analysis proceeds. Experiments on pigeons may not throw much light on the 'nature' of man, but they are extraordinarily helpful in enabling us to analyse man's environment more effectively. What is common to pigeon and man is a world in which certain contingencies of reinforcement prevail. The schedule of reinforcement which makes a pigeon a pathological gambler is to be found at race track and roulette table, where it has a comparable effect.

Another objection is to the use of contrived contingencies of reinforcement. In daily life one does not wear glasses in order to get food or point to circles in order to receive chocolate. Such reinforcers are not naturally contingent on the behaviour and there may seem to be something synthetic, spurious, or even fraudulent about them. The attack on contrived contingencies of reinforcement may be traced to Rousseau and his amazing book, *Émile*. Rousseau wanted to avoid the punitive systems of his day. Convinced as he was that civilization corrupts, he was also afraid of all social reinforcers. His plan was to make the student dependent upon *things* rather than people. John Dewey restated the principle by emphasizing real life experiences in the schoolroom. In American education it is commonly argued that a child must be taught nothing until he can reap natural benefits from knowing it. He is not to learn to write until he can take satisfaction in writing his name in his books, or notes to his friends. Producing a purple rather than a yellow line is irrelevant to handwriting. Unfortunately, the teacher who confines himself to natural reinforcers is often ineffective, particularly because only certain subjects can be taught through their use, and he eventually falls back upon some form of punishment. But aversive control is the most shameful of irrelevancies. it is only in school that one parses a Latin sentence to avoid the cane.

The objection to contrived reinforcers arises from a misunderstanding of the nature of teaching. The teacher expedites learning by arranging special contingencies of reinforcement, which may not resemble the contingencies under which the behaviour is eventually useful. Parents teach a baby to talk by reinforcing its first efforts with approval and affection, but these are not natural consequences of speech. The baby learns to say *mama, dada, spoon*, or *cup* months before he ever calls to his father or mother or identifies them to a passing stranger or asks for a spoon or cup or reports their presence to someone who cannot see them. The contrived reinforcement shapes the topography of verbal behaviour long before that behaviour can produce its normal consequences in a verbal community. In the same way a child reinforced for the proper formation of letters by a chemical reaction is prepared to write long before the natural consequences of effective writing take over. It was necessary to use a 'spurious' reinforcer to get the boy to wear glasses, but once the behaviour had been shaped and maintained for a period of time, the natural reinforcers which follow from improved vision could take over. The real issue is whether the teacher prepares the student for the natural reinforcers which are to replace the contrived reinforcers used in teaching. The behaviour which is expedited in the teaching process would be useless if it were not to be effective in the world at large in the absence of instructional contingencies.

Another objection to effective programmed instruction is that it does not teach certain important activities. When required to learn unprogrammed material for an impending examination the student learns how to study, how to clear up puzzling matters, how to work under puzzlement, and so on. These may be as important as the subject matter itself. The same argument could have been raised with respect to a modern experimental analysis of learning when contrasted with early studies of that process. Almost all early investigators of learning constructed what we now call terminal contingencies of reinforcement to which an organism was immediately subjected. Thus, a rat was put into a maze, a cat was put into a puzzle box, and so on. The organism possessed little if any behaviour appropriate to such a 'problem', but some responses were reinforced, and over a period of time an acceptable terminal performance might be reached. The procedure was called 'trial and error.' A programme of contingencies of reinforcement would have brought the organism to the same terminal performance much more rapidly and efficiently and without trial and error, but in doing so it could have been said to deprive the organism of the opportunity to learn how to try, how to explore—indeed, how to solve problems.

The educator who assigns material to be studied for an impending test presents the student with an opportunity to learn to examine the material in a special way which facilitates recall, to work industriously at something which is not currently reinforcing, and so on. It is true that a programme designed simply to impart knowledge of a subject matter does not do any of this. It does not because it is not designed to do so. Programming undertakes to reach one goal at a time. Efficient ways of studying and thinking are separate goals. A crude parallel is offered by the current argument in favour of the cane or related aversive practices on the ground that they build character, they teach a boy to take punishment and to accept responsibility for his conduct. These are worthwhile goals, but they should not necessarily be taught at the same time as, say, Latin grammar or mathematics. Rousseau suggested a relevant form of programming through which a child could be taught to submit to aversive stimuli without alarm or panic. He pointed out that a baby dropped into a cold bath will probably be frightened and cry, but that if one begins with water at body temperature and cools it one degree per day, the baby will eventually not be disturbed by cold water. The programme must be carefully followed. (In his enthusiasm for the new science, Rousseau exclaimed 'Use a thermometer!') Similar programmes can teach a tolerance for painful stimuli, but caning a boy for idleness, forgetfulness, or bad spelling is an unlikely example. It only occasionally builds what the eighteenth century called 'bottom' as it only occasionally eliminates idleness, forgetfulness, or bad spelling.

It is important to teach careful observation, exploration, and inquiry, but they are not well taught by submitting a student to material which he must observe and explore effectively or suffer the consequences. Better methods are available. There are two ways to teach a man to look before leaping. he may be severely punished when he leaps without looking or he may be positively reinforced (possibly 'spuriously') for looking before leaping. He may learn to look in both cases, but when simply punished for leaping without looking he must discover for himself the art of careful observation, and he is not likely to profit from the experience of others. When he is reinforced for looking, a suitable programme will transmit earlier discoveries in the art of observation. (Incidentally, the audiovisual devices mentioned earlier which undertake to attract attention do not teach careful observation. On the contrary, they are much more likely to deprive the student of the opportunity to learn such skills than effective programming of subject matters.)

Learning how to study is another example. When a teacher simply tests students on assigned material, few ever learn to study well, and many never learn at all. One may read for the momentary effect and forget what one has read almost immediately, one obviously reads in a very different way for retention. As we have seen, many of the practices of the good student resemble those of the programmer. The student can in a sense programme material as he goes, rehearsing what he has learned, glancing at a text only as needed, and so on. These practices can be separately programmed as an important part of the student's education and can be much more effectively taught than by punishing the student for reading without remembering.

It would be pleasant to be able to say that punishing the student for not thinking is also not the only way to teach thinking. Some relevant behaviours have been analysed and can therefore be explicitly programmed. Algorithmic methods of problem solving are examples. Simply leading the student through a solution in the traditional way is one kind of programming. Requiring him to solve a series of problems of graded difficulty is another. More effective programmes can certainly be prepared. Unfortunately, they would only emphasize the rather mechanical nature of algorithmic problem solving. Real thinking seems to be something else. It is sometimes said to be a matter of 'heuristics.' But relevant practices can be formulated as techniques of solving the problem of solving problems. Once a heuristic device or practice is formulated and programmed, it cannot be distinguished in any important way from algorithmic problem solving. The will-of-the-wisp of creative thinking still leads us on.

Human behaviour often assumes novel forms, some of which are valuable. The teaching of truly creative behaviour is, nevertheless, a contradiction in terms. Original discovery is seldom if ever guaranteed in the classroom. In Polya's little book, *How to solve it* (Polya, 1945), a few boys in a class eventually arrive at the formula for the diagonal of a parallelopiped. It is possible that the teacher did not tell them the formula, but it is unlikely that the course they followed under his guidance resembled that of the original discoverer. Efforts to teach creativity have sacrificed the teaching of subject matter. The teacher steers a delicate course between two great fears—on the one hand that he may not teach and on the other that he may tell the student something. Until we know more about creative thinking, we may need to confine ourselves to making sure that the student is in full possession of the contributions of earlier thinkers, that he has been abundantly reinforced for careful observation and inquiry, that he has the interest and industry generated by a fortunate history of successes.

It has been said that an education is what survives when a man has forgotten all he has been taught. Certainly few students could pass their final examinations even a year or two after leaving school or the university. What has been learned of permanent value must therefore not be the facts and principles covered by examinations but certain other kinds of behaviour often ascribed to special abilities. Far from neglecting these kinds of behaviour, careful programming reveals the need to teach them as explicit educational objectives. For example, two programmes prepared with the help of the Committee on Programmed Instruction at Harvard—a programme in crystallography constructed by Bruce Chalmers and James G. Holland and a programme in neuroanatomy by Murray and Richard Sidman—both reveal the importance of special skills in three-dimensional thinking. As measured by available tests, these skills vary enormously even among scientists who presumably make special use of them. They can be taught with separate programmes or as part of crystallography or neuroanatomy when specifically recognized as relevant skills. It is possible that education will eventually concentrate on those forms of behaviour which 'survive when all one has learned has been forgotten.'

The argument that effective teaching is inimical to thinking, whether creative or not, raises a final point. We fear effective teaching, as we fear all effective means of changing human behaviour. Power not only corrupts, it frightens; and absolute power frightens absolutely. We take another—and very long—look at educational policy when we conceive of teaching which really works. It has been said that teaching machines and programmed instruction will mean regimentation (it is sometimes added that regimentation is the goal of those who propose such methods), but in principle nothing could be more regimented than education as it now stands. School and state authorities draw up syllabuses specifying what students are to learn year by year. Universities insist upon 'requirements' which are presumably to be met by all students applying for admission. Examinations are 'standard.' Certificates, diplomas, and honours testify to the completion of specified work. We do not worry about all this because we know that students never learn what they are required to learn, but some other safeguard must be found when education is effective.

It could well be that an effective technology of teaching will be unwisely used. It could destroy initiative and creativity, it could make men all alike (and not necessarily in being equally excellent), it could suppress the beneficial effect of accidents upon the development of the individual and upon the evolution of a culture. On the other hand, it could maximize the genetic endowment of each student, it could make him as skilful, competent, and informed as possible, it could build the greatest diversity of interests, it could lead him to make the greatest possible contribution to the survival and development of his culture. Which of these futures lies before us will not be determined by the mere availability of effective instruction. The use to which a technology of teaching is to be put will depend upon other matters. We cannot avoid the decisions which now face us by putting a stop to the scientific study of human behaviour or by refusing to make use of the technology which inevitably flows from such a science.

The experimental analysis of behaviour is a vigorous young science which will inevitably find practical applications. Important extensions have already been made in such fields as psychopharmacology and psychotherapy. Its bearing on economics, government, law, and even religion are beginning to attract attention.

It is thus concerned with government in the broadest possible sense. In the government of the future the techniques we associate with education are most likely to prevail. That is why it is so important that this young science has begun by taking its most effective technological step in the development of a technology of teaching.

Preparation of this lecture has been supported by Grant K6 MH-21,775-01 of the National Institute of Mental Health of the U.S. Public Health Service, and by the Human Ecology Fund.

References

Ayllon, T. & Azrin, N. H. 1965 An objective method for the measurement and reinforcement of adaptive behaviour of psychotics. *J. Exp. Anal. Beh.* (In the Press.)

Barzun, J. 1963 Review of Bruner, J. S. *Essays for the left hand*, *Science*, 25, 323.

Blough, D. 1956 Dark adaptation in the pigeon. *J. Comp. Physiol. Psychol.* 49, 425-430.

Ferster, C. B. & Skinner, B. F. 1957 *Schedules of reinforcement.* New York. Appleton-Century-Crofts.

James, W. 1889 *Talks to teachers on psychology.* New York. Henry Holt.

Lindsley, Ogden R. 1960 Characterization of the behavior of chronic psychotics as revealed by free operant conditioning methods. *Diseases of the Nervous System*, Monograph Supplement, 21, 66-78.

Polya, G. 1945 *How to solve it.* Princeton, New Jersey. Princeton University Press.

Pressey, S. J. 1926 A simple apparatus which gives tests and scores and teaches. *Sch. Soc.* 23, 373-376.

Rousseau, J. J. 1762 *Émile ou de l'éducation.* Le Haye: Néaulme.

Skinner, B. F. 1938 *The behaviour of organisms.* New York. Appleton-Century-Crofts.

Skinner, B. F. 1957 *Verbal behaviour.* New York: Appleton-Century-Crofts.

Sidman, M. 1964 Personal communication.

Terrace, H. S. 1963 Errorless transfer of a discrimination across two continua. *J. Exp. Anal. Behav.* 6, 223-232.

Wolf, M., Mees, H. & Risley, T. 1964 Application of operant conditioning procedures to the behaviour problems in the autistic child. *Behav. Res. Therapy*, 1, 305-312.

TREATMENT OF NONREADING IN A CULTURALLY DEPRIVED JUVENILE DELINQUENT: AN APPLICATION OF REINFORCEMENT PRINCIPLES

ARTHUR W. STAATS

WILLIAM H. BUTTERFIELD

A 14-year-old, Mexican-American delinquent boy, who had a long history of school failure and misbehavior and second-grade reading achievement, was given 40 hours of reading training which extended over a 4½ month period. Science Research Associates reading materials were adapted for use in conjunction with a token system of reinforcement. During the training, S's attention and participation were maintained in good strength by the reinforcers, he made many reading responses and learned and retained 430 new words, his reading achievement increased to the 4.3-grade level, he passed all his courses for the first time, and his misbehaviors in school decreased to zero.

Staats (1964c, Staats & Staats, 1963) has previously discussed behavior problems and their treatment in terms of learning principles. In doing so it

Staats' address. Department of Educational Psychology, University of Wisconsin, Madison.

[1] The present methods of reading training were formulated, and the present paper written, by the first author as part of a long-term project applying learning principles and procedures to the experimental study of language-learning and reading. The methods were applied by the second author in his position as an officer of the Maricopa County Juvenile Probation Department. The second author also collected and tabulated the data and aided in its graphic presentation. Appreciation is expressed to Chief Probation Officer John H. Walker for lending cooperation in the conduct of the study. In addition, Mary J. Butterfield made important contributions in the preparation of the reading materials used in the study, Brenda Shields typed the materials, and Janet Munir typed the present manuscript.

CHILD DEVELOPMENT, 1965, Vol. 36, pp. 925-942.

was indicated that problem behaviors can arise in part (1) because behavior that is necessary for adjustment in our society is absent from the individual's repertoire, (2) because behaviors considered undesirable by the society are present in the individual's repertoire, or (3) because the individual's motivational (reinforcement) system was inappropriate in some respect.

Although a complete account is not relevant here, several points pertinent to the above conceptions will be made in introducing the present study. The notion that many behavior problems consist of deficits in behavior is important in the study of child development. Behaviorally speaking, a child is considered to be a problem when he does not acquire behaviors as other children do. It is conceivable that a deficit in behavior could arise because the child simply cannot acquire the behavior involved, even though the conditions of learning have been entirely adequate.

It would be expected, however, that behavioral deficits would also arise in cases where the conditions of learning have been defective. Learning conditions can be defective in different ways. For example, the child may never have received training in the behavior he must later exhibit. Or the training may be poor, even though the "trainers," parents or teachers, and so on, have the best intentions.

In addition, however, a child may be exposed to learning conditions that are appropriate for most children but, due to the particular child's past history of learning, are not appropriate for him. It is especially in these cases that people are most likely to conclude erroneously that since other children learn in the same circumstances, the child's deficit must be because of some personal defect. For example, in cases where the training is long term, adequate reinforcement must be available to maintain the attentional and work behaviors necessary for learning. As Staats has indicated (1964c, Staats & Staats, 1963, Staats, Staats, Schutz, & Wolf, 1962), the reinforcers present in the traditional schoolroom are inadequate for many children. Their attentional behaviors are not maintained, and they do not learn. Thus, a deficit in an individual's behavioral repertoire may arise although he has been presented with the "same" training circumstances from which other children profit. Learning does not take place because the child's previous experience has not provided, in this example, the necessary reinforcer (motivational) system to maintain good "learning" behaviors. It would seem that in such a circumstance the assumption that the child has a personal defect would be unwarranted and ineffective.

However, after a few years of school attendance where the conditions of learning are not appropriate for the child, he will not have acquired the behavioral repertoires acquired by more fortunate members of the class—whose previous experiences have established an adequate motivational system. Then, lack of skilled behavior is likely to be treated aversively. That is, in the present case, the child with a reading deficit (or other evidence of underachievement) is likely to be gibed at and teased when he is still young and ignored, avoided, and looked down upon when he is older. Although the individuals doing this may not intend to be aversive, such actions constitute the presentation of aversive stimuli. Furthermore, this presentation of aversive stimuli by other "successful" children, and perhaps by a teacher, would be expected to result in further learning, but learning of an undesirable nature. These successful children, teachers, academic materials, and the total school situation can in this way become learned negative reinforcers, which may be translated (see Staats, 1964b) to say the child acquires negative attitudes toward school.

At this point, the child is likely to begin to "escape" the school situation in various ways (daydreaming, poor attendance, and so on) and to behave aversively in turn to the school and its inhabitants (vandalism, fighting, baiting teachers and students, and the like). Thus, a deficit in behavior, resulting from an inappropriate motivational system, can lead to the further development of inappropriate reinforcers and inappropriate behaviors.

The foregoing is by no means intended as a complete analysis of delinquency, dropouts, and the like. However, it does indicate some of the problems of learning that may occur in school. In addition, it does suggest that an analysis in terms of laboratory-established learning principles, when applied to problems such as in classroom learning of the above type, can yield new research and applied hypotheses. It was with this general strategy that the study of reading acquisition employing learning principles and reinforcement procedures were commenced (Staats, 1964a, Staats *et al.*, 1962, Staats, Finley, Minke, & Wolf, 1964a, Staats, Minke, Finley, Wolf, & Brooks, 1964b). The present study is a replication and an extension of these various findings to the development of a program for training nonreaders to read. The program, which adapts standard reading materials, is based upon the principle of the reinforcer system employed in the previous studies with the younger children, thus testing the principles of reinforcement in the context of remedial reading training, as well as the feasibility of using the type of reinforcement system with a new type of S. As such, the study has implications for the study of nonreading children of pre-adolescent, adolescent, and young adult ages. In the present case, S was also a culturally deprived delinquent child—and the study thus involves additional information and implications for the special problems associated with education in this population of children.

METHODS

Subject

The S was a 14-year-and 3-month-old boy of Mexican-American ancestry. He was the fifth child in a family of 11 children and the mother and father. The parental techniques for controlling their children's behavior consisted of physical and verbal abuse. Both parents described their own childhood conditions as primitive. The father was taken out of school after completing the fifth grade to help with his father's work. Each of S's four older brothers had been referred to the juvenile court for misbehavior. The parents appeared to be at loss as to how to provide effective control for family members.

The S had a history of various miscreant behaviors, having been referred to the juvenile department nine times for such things as running away, burglary, incorrigibility, and truancy. During the course of the Study S was again referred on a complaint (with three other boys) of malicious mischief for shooting light bulbs and windows in a school building with a BB gun. He associated with a group of boys who had been in marked difficulty with the law. The S smoked, and on occasion he drank excessively.

The study commenced when S was residing with his family. However, after the complaint on malicious mischief S was sent to a juvenile detention home During his stay there he was allowed to attend school in the daytime. The study was finally concluded when S was committed to an industrial school for juvenile-delinquent boys. This occurred because S baited the attendants at the detention home and caused disturbances which, although not serious, were very unpleasant and disruptive.

On the Wechsler Bellevue Form I, given when S was 13-10, he received Verbal and Performance IQ's of 77 and 106, respectively, for a Full Scale IQ of 90 The examiner concluded that S was probably within the normal range for this test. On the basis of this test and HTP Projective Drawings, S was characterized as having a poor attention span and poorly integrated thought processes and as lacking intellectual ambitiousness. He was also described as seeking satisfaction in fantasy and as having good conventional judgment.

The S had continually received failing grades in all subjects in school. He was described as having "been incorrigible since he came here in the second grade He has no respect for teachers, steals and lies habitually and uses extremely foul language." The S had been promoted throughout his school career simply to move him on or to "get rid of him." He was disliked by the teachers and administrators in grade school because of his troublesome behavior and was described by the principal as mentally retarded even though one of the tests taken there indicated a score within the normal range. Another test taken there gave him an IQ of 75. During the study S was attending a local high school and taking classes for low-le..l students.

Reinforcer System

In previous studies (Staats, in press, Staats et al., 1964a; 1964b), a reinforcer system was demonstrated that was capable of maintaining attention and work behaviors for long-term experimental studies. This system worked well with preschool children of ages 2 to 6 and with educable and trainable retardates of ages 8 to 11. The principle of the system was based upon token reinforcers. The tokens were presented contingent upon correct responses and could be exchanged for items the child could keep. In the previous studies toys of various values could be obtained when a sufficient number of tokens had been accrued in visible containers.

This system was adapted for use with the adolescent S of the present study. In the adaptation there were three types of token, distinguished by color. The tokens were of different value in terms of the items for which the tokens could be exchanged. A blue token was valued at $1/10$ of one cent. A white token was valued at $1/5$ of a cent. A red token was worth $1/2$ of a cent.

The child's acquisition of tokens was plotted so that visual evidence of the reinforcers was available. The tokens could be used to purchase a variety of items. These items, chosen by the subject, could range in value from pennies to whatever the subject wished to work for. Records were kept of the tokens earned by S and of the manner in which the tokens were used.

Reading Materials

The reading material used was taken from the Science Research Associates (SRA) reading-kit materials. The SRA kits consist of stories developed for and grouped into grade levels. Each story includes a series of questions which can be used to assess the reader's comprehension of the story. The reading training program was adapted from the SRA materials as follows:

Vocabulary words.—A running list was made of the new words that appeared in the series of stories. The list finally included each different word that appeared in the stories that were presented. From this list, the new vocabulary for each story was selected, and each word was typed on a separate 3 × 5 card.

Oral reading materials.—Each paragraph in the SRA stories was typed on a 5 × 8 card. Each story could thus be presented to S paragraph by paragraph.

Silent-reading and comprehensive-question materials.—Each SRA story, with its comprehensive questions, was typed on an 8½ × 13 sheet of white paper.

Procedure

Vocabulary presentation.—The procedure for each story in the series commenced with the presentation of the new words introduced in that story. The words were presented individually on the cards, and S was asked to pronounce them. A correct response to a word-stimulus card was reinforced with a midvalue token. After a correct response to a word, the card was dropped from the group of cards yet to be presented. The S was instructed to indicate words that he did not know the meaning of, and this information was provided in such cases.

When an incorrect response to a word stimulus occurred, or when S gave no response, E gave the correct response. The S then repeated the word while looking at the stimulus word. However, the word card involved was returned to the group of cards still to be presented. A card was not dropped from the group until it was read correctly without prompting. After an error on a word stimulus, only a low-value token was given on the next trial when the word was read correctly without prompting. The vocabulary presentation phase of the training was continued until each word was read correctly without prompting.

Oral reading.—Upon completion of the vocabulary materials, each paragraph was individually presented to S in the order in which the paragraph occurred in the story. When correct reading responses were made to each in the paragraph, a high-value token was given upon completion of the paragraph. When a paragraph contained errors, S was corrected, and he repeated the word correctly while looking at the word. The paragraph was put aside, and when the other paragraphs had been completed, the paragraph containing errors was again presented. The paragraph was repeated until it was done correctly in its entirety—at which time a midvalue token was presented. When all paragraphs in a story had been completed correctly, the next phase of the training was begun.

Silent reading and comprehensive questions.—Following the oral reading S was given the sheet containing the story and questions. He was instructed to read the story silently and to answer the questions beneath the story. He was also instructed that it was important to read to understand the story so that he could answer the questions.

Reinforcement was given on a variable interval schedule for attentive behavior during the silent-reading phase. That is, as long as S appropriately scanned the material he was given a low-value reinforcer an average of every 15 seconds. The exact time for reinforcement was determined by a table of random numbers varying from 1 to 30 seconds. Whenever he did anything else than peruse the material, no reinforcement was given. The next interval was then timed from the moment S returned to the silent reading, with the stipulation that no reinforcement be given sooner than 5 seconds after S returned to the reading. If the interval was less than 5 seconds, a token was not given until the next interval had also occurred. Timing was done by a continuously running stopwatch. The S was also given an extra midvalue token at the end of the silently read story on those occasions where he read without moving his lips.

Upon completion of the story, S wrote his answers to the questions typed below the story and gave his answers to E. For each correct answer, S received a high value token. For an answer with a spelling error, S was reinforced with a midvalue token when he had corrected the answer. For incorrect answers S had to reread the appropriate paragraph, correct his answer, and he then received a midvalue token.

Vocabulary review.—Some of the vocabulary words presented to S in the first phase of training were words he already could read. Many others, however, were words that the procedure was set up to teach. The oral-reading-phase performance indicated the level of S's retention of the words he had learned—and also provided further training trials on the words not already learned. A further assessment of S's retention of the words that he did not know in the vocabulary training was made after each 20 stories of the SRA materials had been read. This test of individually presented words, for each story, was started about 3 days after completion of the 20 stories and constituted fairly long-term retention.

This test was also used as a review for S, and further training on the words was given. This was first done by reinforcing S with a low-value token for every word he read correctly. However, S's attention was not well maintained by this reinforcement, and the procedure was changed to provide a midvalue token for correctly read words. When S could not read a word, or missed one, he was prompted and had to correctly repeat the name of the word while looking at the word. This word card was then put aside and presented later, at which time S was reinforced with a low-value token if he read it correctly. If not, the procedure was repeated until a correct unprompted trial occurred.

Achievement tests.—Prior to the commencement of the training, S was tested to assess his reading performance, and during the period of experimental training he was given two additional reading-achievement tests. The first one given was the Developmental Reading Test. (At this time the S's vision and hearing were also tested and found to be normal.) After 45 training sessions another reading test was given S, this time the California Reading Test, Form BB, for grades 1, 2, 3, and L-4. Twenty-five sessions later, just before the termination of the study, S was given the California Reading Test, Form BB, for grades 4, 5, and 6. The S's performance on the three reading tests constituted one of the measures of his progress. The tests were given at the Arizona State University Reading Center.

Training sessions.—The training sessions would ordinarily last for 1 hour or less, although a few sessions were as short as 30 minutes or as long as 2 hours. Not all of this time was spent in reading, however. A good deal of time was spent in arranging the materials, recording S's performance, keeping count of the reinforcers, plotting the reinforcers accrued, and so on. The time spent actually reading was tabulated. During the 4½-month experimental period, 70 training sessions were conducted, with an average of about 35 minutes spent per session for a total of 40 hours of reading training.

RESULTS AND CONCLUSIONS

During the period of training S made many reading responses. Figure 1 shows the number of single-word reading responses S made as a function

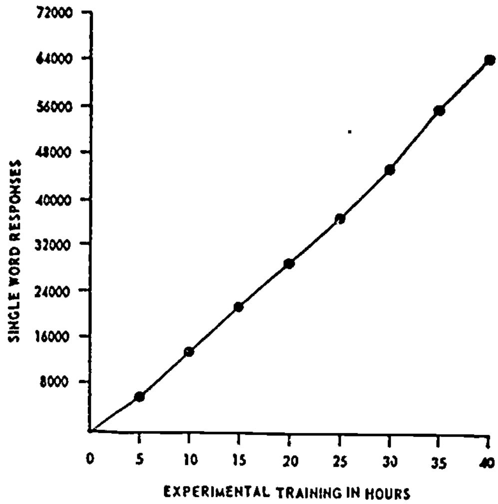


FIG 1.—Number of single-word reading responses as a function of the time in experimental reading training.

of the hours of time spent in training. An estimate of the number of single-word reading responses was obtained from tabulating each presentation of a word card, the number of words in the stories, and the reading-comprehension questions at the end of each story, as well as the words presented to S in the later single word retention test. Actually, the number of words in the stories was an estimate obtained from the mean number of words in two out of each five stories. Thus, rather than giving the true absolute number of reading responses made, the figure gives an estimate. However, the most important aspect of the figure is to indicate the rate of this single-word reading-response measure as a function of time in experimental training. As can be seen, as the training progressed S covered the reading material at a slightly more rapid rate, as is shown by the slight positive acceleration in the curve The importance of this result is to indicate that the child's behavior of attending to the task and making the appropriate reading responses did not diminish throughout the period of training. Thus, the reinforcement system employed was capable of maintaining the behavior for a long period of time. During this time the attentional and cooperative behaviors instigated resulted in many, many, learning trials—a *sine qua non* for the acquisition of achievement in any skill.

Before reading each story S was presented with individual cards for all the words included in that story which had not been presented in a previous story. When these words were presented, S would read a certain proportion correctly on first presentation, the other words being missed on the first presentation. The ones missed were considered to be new words for S, words that he had not previously learned. These words were separately tabulated. The cumulative number of these new words as a function of every 5 SRA stories read is shown by the top curve of Figure 2. (The data for the first 10 stories are not presented since they were not available for all three curves.) As this curve indicates, 761 new words were presented to S during the training.

Thus, S missed 761 words when they were first presented to him. However, he was given training trials on these words, and then he then read them again in the oral reading of the paragraph. The number of these words that he missed in this oral-reading phase is plotted in the bottom curve of Figure 2. This curve then indicates the number of errors made on the second reading test of the words that had been previously learned. Thus, only 176 words out of the 761 (about 23 per cent) were missed in the oral-reading phase—showing retention for 585 words. The results indicate that the criterion of one correct unprompted reading trial in the original vocabulary-learning phase produced considerable learning when the words were read in context.

The middle curve in Figure 2 involves a measure of long-term retention of the words that had been learned. This measure was obtained by testing S on the words, presented singly, that had been learned in the preceding 20 stories. This test was given 10 to 15 days after the training occurred. The training thus included the previous single-word presentations of the words, as well as those same words read orally and silently. In addition, however, S had also learned a considerable number of other words by the time of this test. As the middle curve shows, when tested 10–15 days later, S read 430 of the 761 words correctly, or, conversely, 331 words (about 43 per cent) were missed. Thus, the procedures produced retention when the words were later presented out of context after a considerable intervening period.

The results appearing in Figure 2 indicate that the child covered a considerable amount of reading material, that he learned to read a number of new words when presented individually or in context, and that he retained a good proportion of what he had learned. The results also indicate that the child improved during the training in his retention. That is, his rate of getting new words in the first-presentation phase continues at a high rate throughout the study. (This supports the results shown in Fig. 1 indicating that the child's behavior did not weaken during the training.) However, his "rate" of missing the new words on the second and third presentations decreased, that is, he retained more of the words he had learned. Thus, tabulation indicated that for the first 35 stories only about 33 per cent of the words learned were retained 10–15 days later, whereas S's subsequent retention increased to about 55 per cent. It should be noted that this improvement occurred even though the difficulty of the words (as shown in Fig. 2

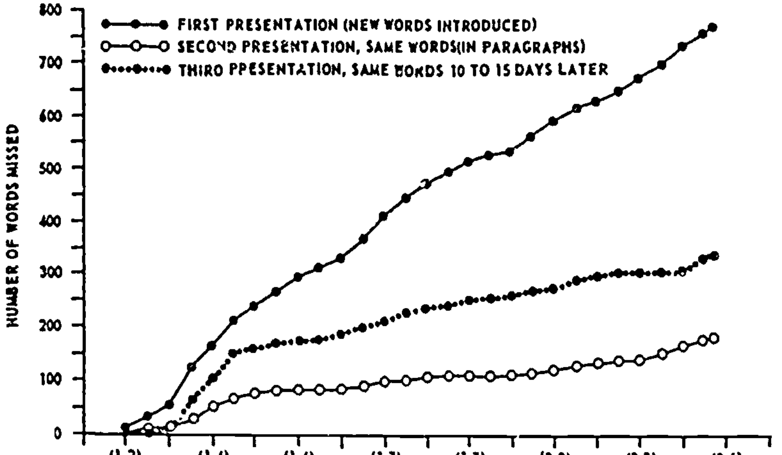


FIG. 2.—Number of words missed on first, secor ', and third presentations for the 150 SRA stories.

by the numbers in parentheses) became progressively greater during the training, moving from the 1.2 grade level of difficulty to the 2.6 grade level.

These results receive support from the data presented in Figure 3. As already indicated, on the first presentation of the vocabulary of a story, some words were missed out of the total presented—and S was then presented with training on these words. Figure 3 shows the number of the words presented and missed in ratio to the total number presented, as this ratio is related to the number and difficulty of the stories presented. A smaller ratio indicates that S missed fewer of the total vocabulary words when they were presented for the first time. As can be seen in Figure 3, as the child read more stories in his training (even though they become more difficult), he missed fewer and fewer words that were presented to him. It should be stressed that he was thus improving in the extent to which he correctly responded to new words on first presentation. This improvement appeared to be correlated with other observations that indicated that S was also beginning to learn to sound out words as a function of the training. For example, he remarked when in the judge's office that he thought a sign said "information," because he could read the "in" and the "for" and the "mation." In addition, S reported a number of times that the training was helping him in

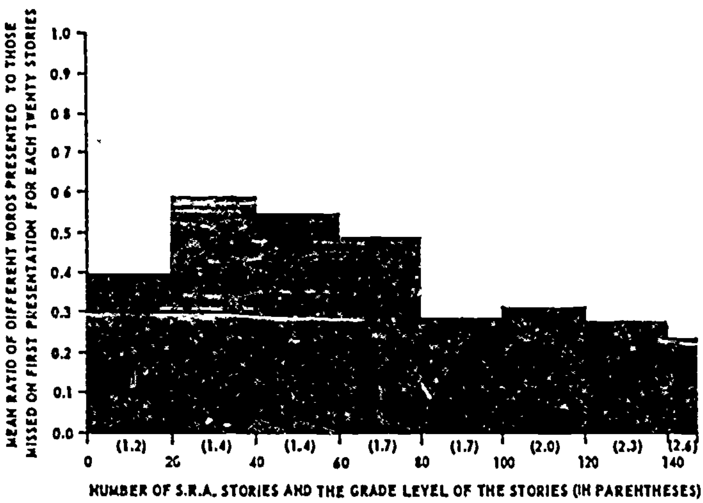


FIG. 3.—Ratio of words presented to those missed on first presentation for the 150 SRA stories.

school, that reading was getting easier for him in school, that he liked the reading training better as he went along, and so on. It would be expected (as will be supported by other data) that as the reading training improved his reading in school, the things he learned in school would also improve his performance in the reading training. It is this effect that may also be reflected in his increasing ability to read the new words presented to him.

In addition to this direct evidence of the child's progress in reading training, and the foregoing indirect evidence that the reading training was having general effects upon the child's behavior, the study was formulated to obtain other sources of information concerning the child's progress. One means of doing this was to give the child reading-achievement tests before beginning the reading training as well as during the training. The results of these tests are shown in Figure 4. The first point on the curve is a measurement obtained by use of the Developmental Reading Test giving a total score of reading achievement showing that S was performing at the grade-2 level. After 45 reading-training sessions, S's performance on the California Reading Test showed a gain to the 3.8-grade level. By the end of the training, after 25 more training sessions, S had advanced to the 4.3-grade level on the California Reading Test.

Another indication of the general effect of the reading training came from the child's performance in school, both in school achievement and deportment. The period of reading training coincided with a school term. The boy received passing grades in all subjects. A *C* in physical education, a *D* in general shop, a *D* in English, and a *D* in mathematics. It should be em-

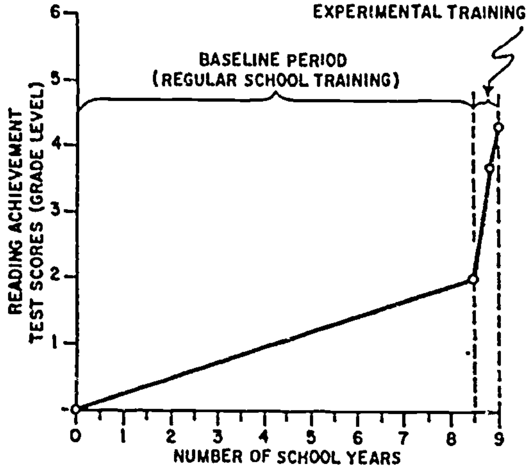


FIG. 4.—Reading-achievement test scores as a function of 8½ years of school training and 4½ months of experimental training.

phasized that these grades represent the first courses that this child had ever passed, and thus his finest academic performance.

Furthermore, S began to behave better while in school. The boy had always been a behavior problem in school, and this continued into the period during which S received reading training. As Figure 5 shows, during the first month of the training S committed 10 misbehaviors that resulted in the receipt of demerits. The behaviors were as follows. disturbance in class (2 times), disobedience in class (five times), loitering (2 times), and tardiness. In the second month he was given demerits for scuffling on the school grounds and also for creating a disturbance. In the third month he was given demerits for cutting a math class and for profantiy in class. As the figure shows, however, no misbehaviors occured in the fourth month or in the half month after this until the conclusion of the school term.

The S requested that the tokens be exchanged for items that he wanted in sessions 12, 17, 25, 31, 35, 43, 49, 55, and in the last session he was given the value of the remaining tokens in cash. Items included were a pair of "beatle" shoes, hair pomade, a phonograph record, an ice cream sundae, a ticket to a school function, money for his brother who was going to reform school, and so on. Further information regarding the reinforcement system is given in Figure 6. The vertical axis of the graph represents the ratio of the number of tokens obtained by S relative to the number of single-word reading responses which he emitted. Lesser ratios thus indicate more reading responses per reinforcer. This ratio was plotted as a function of the progress S made in the training program, as given by the number of SRA stories he had completed. As the training progressed S gradually made an increasingly

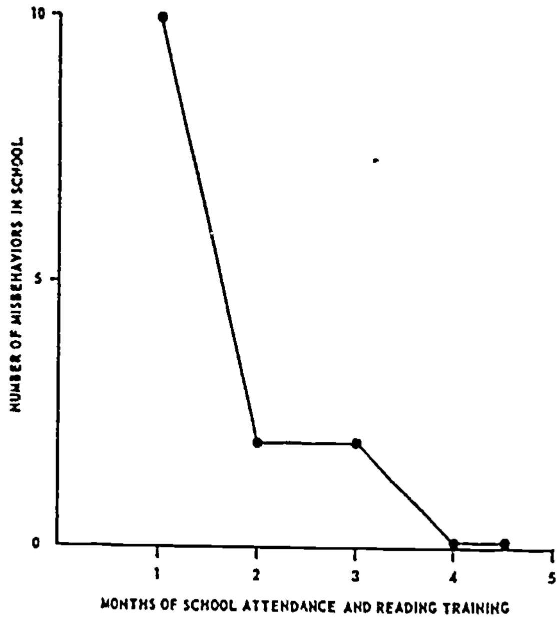


FIG. 5—Number of official misbehaviors in school as a function of time in the experimental training.

greater number of reading responses per reinforcer. This effect was not accomplished by changing the rules by which the reinforcers were administered. The effect, which was planned in the training program, resulted from the fact that the SRA stories became longer as the grade level was raised. Since, for example, paragraph reading was reinforced by the paragraph, the longer the paragraph, the greater the number of reading responses that had to be emitted before reinforcement was obtained. At the end of training, thus, S was getting about half as much reinforcement per response as at the beginning of training. It should also be indicated that the stories were more difficult as the training progressed, so the effort involved in reading was increasing—although reinforcement for the reading was decreasing.

During the 4½ months of training, which involved 40 hours of reading training and the emission of an estimated 64,307 single-word reading responses, S received $20.31.

DISCUSSION

In this section the various aspects of the reading-training procedures will first be discussed. Then the implications of the results and analysis will be outlined both for further studies of remedial reading training as well as for a learning conception of certain aspects of cultural deprivation and delinquency.

The method of reading training used in the present study was derived from previous study (Staats, 1964a, in press, Staats et al., 1962) with preschool children in which words were first presented singly, then in sentences, and finally in short stories. The present study indicated that SRA materials can be adapted for a similar type of presentation in conjunction with the type of reinforcer system previously developed (Staats et al., 1964a; 1964b). From the SRA materials it was possible to present single-word training trials and oral-reading training and to develop a silent-reading training procedure, all involving reinforcement.

When the training of reading, at least in part, is considered as operant discrimination learning, the learning task consists of having S emit the correct speech response while looking at the verbal stimulus—this process being followed by reinforcement. This basic procedure was elaborated in the present study to include two levels of reinforcement. An unprompted reading response on the first trial was reinforced more heavily than one that had been previously missed. This procedure appeared to produce learning that was retained very well when the child later read the words orally in a paragraph, with considerable retention also occurring when the child was tested on the individual words 10–15 days later.

It may seem incongruous at first to attempt to reinforce silent reading, since this behavior is not observable. However, it should be remembered that the subject actually has two types of behavior in the silent-reading act. He looks at the verbal stimuli—that is, attends—and he makes "reading" verbal responses to the verbal stimuli. While the reading responses cannot be monitored when they are covert, the attending behavior can be. Of course, there is a danger involved in einforcing the behavior of just looking at something. Perhaps the child will do nothing else. If he is heavily reinforced for sitting and looking at a page, and the actual reading responses are effortful, he may not emit the reading responses. The present procedure was set up to eliminate this possibility by using a double contingency. The child was reinforced for simple attention, but the reinforcement was low in value. The opportunity for a greater amount of reinforcement came during the answering of the questions. Thus, although simple attention was reinforced lightly, attention and reading responses were reinforced much more heavily. In this way it was possible to use reinforcement in a procedure designed to maintain reading for "understanding," in addition to simple "word-naming." (These results could be generalized to other types of learning.) Furthermore, this procedure provided an opportunity to train the subject to read silently. Although he had a tendency to make vocal or lip responses while reading, it was possible to strengthen reading without these other responses through differentially reinforcing the correct silent reading.

Thus, it may be concluded that the reading program increased the child's reading vocabulary as shown by the various measures of retention used in the study, the tests of reading achievement, as well as the child's improved school performance and his verbal description of improved attitude toward and performance in reading in school. There were also suggestions that the child was acquiring a "unit-reading repertoire," that is, the general ability to sound out words through making the correct response to single letters and syllables. Thus, for example, the child made errors on fewer and fewer of the new words presented as the training progressed, even though the words were of greater difficulty. In addition, he retained a greater proportion of the words he learned as he went on. Further research of the present type must be conducted to test the possibilities for using a more phonic system of remedial reading training with the present type of subject.

A final point should be made concerning the training procedures used in the present study. The procedures are very specific and relatively simple. Thus, it was not necessary to have a person highly trained in education to administer the training. In the present case the trainer-experimenter was a probation officer. It might also be suggested that anyone with a high-school education and the ability to read could have administered the training. This has implications for the practical application of the present methods, since one of the questions that arises in this context concerns the economy of the procedures. Although the procedures as described involved a one-trainer-to-one-student ratio, as many remedial teaching procedures do, in the present case the simplicity of the procedures suggests the possibility that savings may be effected because the trainer need not be so highly trained. Thus, the procedures could be widely applied or adapted by various professionals, for example, social workers, prison officials, remedial teachers, tutors, and so on. In an even more economical application, helpers of professionals could be used to actually administer the procedures, for example, selected delinquents (or prisoners) could administer the procedures to other delinquents. Thus, the procedures could be utilized in various situations, such as settlement houses, homes for juvenile delinquents, prison training programs, parts of adult education, and so on. All that is needed is a suitable system of reinforcers to back up the tokens. These conclusions are supported by another study by A. W. Staats now being prepared for publication in which analogous methods developed for work with preschool children were applied by a graduate student who was untrained in teaching; the results suggest possibilities for economic innovations in education generally.

In the same context, it may be worthwhile pointing out that the results indicated that the child advanced as many years in reading achievement, as measured by the tests, during the experimental training as he had in his previous school history. A comparison of the relative costs—in the present case, about 70 hours of time of a person not necessarily trained in teaching and $20.31 for the reinforcers versus 8½ years of trained teachers' time, albeit in a group situation—suggests that the procedure introduced in the present study may not be uneconomical, even without improvements in the method. And, as will be further described, the child's failure in school may in many cases be considered as a contributor to the child's delinquency—which also carries a high cost to society. The present results, in suggesting that the training procedures may also effect general improvements in behavior, including misbehaviors in school, thus have further implications concerning the economy of the procedures.

The present study, among other things, tests the feasibility of using the type of reinforcing system, previously applied successfully to younger children, to the study of learning in older children—in this case a 14-year-old juvenile delinquent. The reinforcer system worked very well with the present S, maintaining his attention and working behaviors in good strength for a long period of time. And there was every reason to expect that the study could have been continued for a much longer period, probably as long as it would have taken to train the child to read normally.

It should be noted that although the amount of reinforcement given decreases during the training, as shown in Figure 6, the reading behavior is

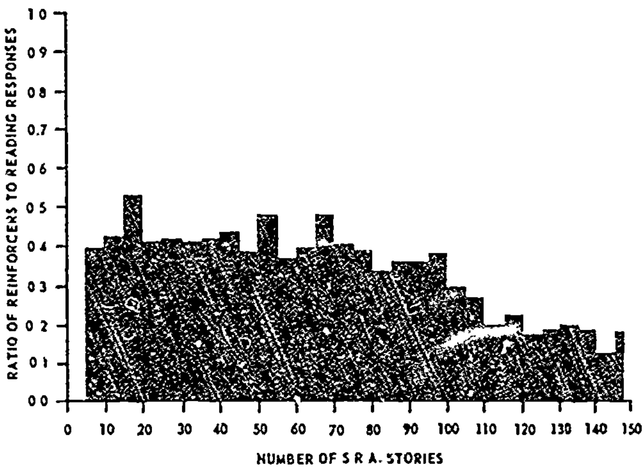


FIG. 6.—Ratio of the number of tokens received divided by the number of reading responses made as a function of the number of SRA stories read.

maintained in good strength throughout the study, as shown in Figures 1 and 2, thus, less and less reinforcement is needed to maintain the behavior even though the material increases in difficulty. As already described, this occurred because a progressively greater number of reading responses was necessary per reinforcer. This is analogous to gradually raising the ratio of responses to the reinforcers as considered in terms of ratio schedules of reinforcement. Staats has suggested that this type of gradual increase must occur to produce good work behaviors in humans (Staats & Staats, 1963).

This result in the present study is in part an answer to the question whether the use of extrinsic reinforcers in training will produce a child who is dependent upon these reinforcers. It is not possible to discuss this topic fully now. However, it may be said that the extrinsic reinforcement can be gradually decreased until, as was happening with the present child, reading becomes reinforcing itself, or other sources of reinforcement maintain the behavior.

A word should be said concerning the relevance of reinforcement variables in the treatment of nonlearning in culturally deprived children. Typically, as in the present case, such children do not, as a result of their home experiences, acquire "reinforcer systems" appropriate for maintaining learning in the traditional classroom. Rosen (1956) has shown that, in the present terminology, lower-class children do not have experiences that make school achievement and learning itself positively reinforcing. This deficit, among others that affect the reinforcer system, can be expected to lead to poor school learning and other behavioral deficits. In such cases, there are increased opportunities for other poor social attitudes and undesirable behaviors to develop, as suggested in the introduction and exemplified in the present case.

The present study suggests that these conditions can be reversed through the application of learning principles and reinforcement variables to the task of repairing the child's behavioral-achievement deficit. There were indications that this treatment resulted in improvement in the reinforcement value of (attitudes toward) school for this child and consequently in the decrease in incidence of misbehaviors in school. The results thus suggest that under appropriate conditions the deficit in behavior stemming from the child's inadequate reinforcing system may be, at least in part, repaired by a properly administered, effective reinforcement system, resulting in a decrease in undesirable behaviors.

A comment should be made about the possibility of a Hawthorne effect, that is, that the social reinforcement by the E and possible extraexperimental reinforcement contributed to the results in the present study. It would be expected that such reinforcers could contribute to the overall effect—and in the present case the expenditure for the material reinforcers was small. In general, it can be expected that individuals will vary in the extent to which social reinforcers will be effective. For example, in preschool children social reinforcement is ineffective for long-term training (Staats, 1964c, Staats et al., 1962), and the same would be expected for many individuals with behavior problems. Ordinarily, it might be expected that the weaker other sources of reinforcement are for the individual, the stronger must be the reinforcer system of the treatment procedure.

In conclusion, the present study helps support and replicate the previous findings and extends the general procedures and principles to the study of an adolescent child who is culturally deprived and is also a juvenile delinquent. The various sources of data used suggest that the present procedures and principles are applicable to this population also. Based upon these suggestions, further studies will be conducted on culturally deprived children, delinquent and nondelinquent, as well as studies of other types of non-achieving or underachieving readers.

It should also be indicated that the present study indicates the possibility for developing procedures for the objective application and test of laboratory-derived learning principles within the context of an actual problem of behavior. As previously indicated (Staats, 1964a), verification of learning principles in the context of a problem of human behavior constitutes one way to further the generality of the principles themselves. It may thus be suggested that such studies have two types of implication. they have implications for people interested in dealing with the problems of human behavior, as well as for those interested in the extension and verification of the basic science.

REFERENCES

Rosen, B. C. The achievement syndrome. A psychocultural dimension of social stratification. *Amer. sociol. Rev.*, 1956, 21, 203–211.

Staats, A. W. A case in and a strategy for the extension of learning principles to problems of human behavior. In Staats, A. W. (Ed.), *Human learning*. New York: Holt, Rinehart & Winston, 1964. (a)

Staats, A. W. Conditioned stimuli, conditioned reinforcers, and word meaning. In Staats, A. W. (Ed.), *Human learning*. New York: Holt, Rinehart & Winston, 1964. (b)

Staats, A. W. (Ed.) *Human learning*. New York. Holt, Rinehart & Winston, 1964. (c)

Staats, A. W. An integrated-functional learning approach to complex human behavior In B. Kleinmuntz (Ed.), *Problem solving: research, method and theory*. New York: Wiley, in press.

Staats, A. W., Finley, J. R., Minke, K. A., & Wolf, M. Reinforcement variables in the control of unit reading responses. *J. exp. anal. Behav.*, 1964, 7, 139–149. (a)

Staats, A. W., Minke, K. A., Finley, J. R., Wolf, M., & Brooks, L. O. A reinforcer system and experimental procedure for the laboratory study of reading acquisition. *Child Develpm.*, 1964, 35, 209–231. (b)

Staats, A. W., & Staats, C. K. *Complex human behavior*. New York. Holt, Rinehart & Winston, 1963.

Staats, A. W., Staats, C. K., Schutz, R. E., & Wolf, M. The conditioning of textual responses utilizing "extrinsic" reinforcers. *J. exp. anal. Behav.*, 1962, 5, 33–40.

PRODUCTION AND ELIMINATION OF DISRUPTIVE CLASSROOM BEHAVIOR BY SYSTEMATICALLY VARYING TEACHER'S BEHAVIOR[1]

DON R. THOMAS, WESLEY C. BECKER, AND MARIANNE ARMSTRONG

The effects of teacher behaviors on the classroom behaviors of children were investigated by systematically varying approving (praise, smiles, contacts, *etc.*) and disapproving (verbal reprimands, physical restraint, *etc.*) classes of teacher behavior. Measures were taken on both teacher and child behaviors. Each day a sample of 10 children was observed. The subject pool was a class of 28 well behaved children in a middle primary public school class. The results demonstrated that approving teacher responses served a positive reinforcing function in maintaining appropriate classroom behaviors. Disruptive behaviors increased each time approving teacher behavior was withdrawn. When the teacher's disapproving behaviors were tripled, increases appeared most markedly in the gross motor and noise-making categories of disruptive behavior. The findings emphasize again the important role of the teacher in producing, maintaining, and eliminating disruptive as well as pro social classroom behavior.

Teachers are sometimes unaware of the effects of their actions on the behavior of their students. Many teachers assume that if a child performs disruptive acts in the classroom then the child must have a problem at home, or at the very least, must not have reached a stage of sufficient maturity to function adequately in the school situation. However, an increasing body of evidence indicates that many of the behaviors which teachers find disruptive are actually within their control. A teacher can modify and control the behavior of her students by controlling her own responses.

Contingent use of social reinforcement has been shown to control such motor behaviors as walking, standing, and running (Bijou and Baer, 1963), talking and crying (Kerr, Meyerson, and Michael, 1965, Hart, Allen, Buell, Harris, and Wolf, 1964), and classroom conduct (Becker, Madsen, Arnold, and Thomas, 1967, Zimmerman and Zimmerman, 1962).

Becker *et al.* (1967) worked in public schools with teachers who had problem children in their classes. Behaviors exhibited by the students were observed and the frequency of these behaviors was estimated for each child. Each teacher was taught to use praise, smiles, *etc.* to reinforce good behavior. The rate of appropriate classroom behaviors increased in most cases as soon as teacher approval and recognition were made contingent on such behavior.

The present study evolved from prior research showing the importance of social reinforcement, and Becker's work, which suggests that specific procedures, or definable classes of teacher behaviors can be used by the teacher to increase appropriate classroom behaviors. In order to provide more convincing data on the role of different teacher behaviors, the present study was designed to produce and remove problem behavior in students by systematically varying teacher behaviors in an initially well-behaved class.

METHOD

Subjects

Students. A class of 28 elementary students at the middle primary level was selected. According to the teacher her class was "a good class, with an above-average distribution of ability and no 'bad' kids." Most of the children were from upper-middle- and middle-

[1]The authors wish to thank Urbana School District #116 and the principal of Thomas Paine School, Mr. Richard Sturgeon, for their cooperation. The observers (Loretta Nielson, Barbara Goldberg, Marilyn Goldberg, and Darlene Zientarski) deserve thanks for their conscientious work. This research was supported, in part, by National Institute of Child Health and Human Development Grant HD 00881 05 Reprints may be obtained from Wesley C Becker, Bureau of Educational Research, University of Illinois, Urbana, Illinois 61801.

JOURNAL OF APPLIED BEHAVIOR ANALYSIS, 1968, Vol. 1, pp 35 45.

income range families. Ages at the beginning of the study ranged from 6 yr, 11 months to 7 yr, 11 months, I Q range (group test) was from 99 to 134.

Teacher The teacher, age 23, obtained her student teaching experience with a class of "maladjusted" children. In addition, she had 1 yr experience with a class of "slow learners". Preliminary observations indicated that she rarely attended in an approving manner to children who behaved inappropriately, and rarely reprimanded children who were performing their assigned tasks She volunteered to participate in the study because of its potential contribution to teacher training in the future.

Observation Procedures

The basic data for the study consisted of the relative frequency of occurrence of classes of child behaviors in relation to classes of teacher behaviors utilizing rating schedules to be described One to three observers were placed in the classroom each morning from approximately 9.15 to 10.00 a.m. while the students were completing reading and reading workbook assignments. To insure obtaining a daily sample of both child and teacher behaviors during this 45 min work period, a 20 min observation time was decided on for both child and teacher observations. Thus, even if only one observer was present, the relevant information could be obtained. This time restriction limited the number of children who could be observed each day. Ten children were selected for observation each morning by drawing numbers from a hat. During Baseline$_1$ and the first No Praise condition a no replacement procedure was used so that all children had to be observed before a child's number could be drawn a second time At the start of Baseline$_2$ this restriction was removed Through the use of a numbered seating chart, the observers recorded the behaviors of selected children in the order in which they were chosen. Five extra numbers were drawn each day to provide observation targets in case one or more of the first 10 subjects drawn were not available for observation. Target children were observed for 2 min each. Each minute was divided into six 10-sec intervals Observers were trained to record classes of behavior which occurred in a given interval. Recordings were made during the first five intervals of each minute. During the sixth 10 sec interval the observers made notes, checked for synchronization, and/or prepared to switch to a new child. Thus, the daily child observation sample consisted of ten 10-sec observation intervals on each of 10 children.

Teacher behaviors were recorded on a similar schedule, the only difference being that for teacher behaviors each occurrence of a response in a specified class was recorded (frequency measure), whereas for child behaviors a given class of behavior could be rated only once in a 10-sec interval. This difference in procedure was necessitated by the greater difficulty in separating child behaviors into discrete response units. Observers used a clipboard, stopwatch, and a recording sheet which had spaces for 100 observation intervals, guides for computing reliability, and a place for comments.

Undergraduate university students were hired and trained to collect the data. Each observer memorized the definitions of classes of child and teacher behaviors. Pre-baseline training in recording of behavior was carried out in the experimental classroom to allow the children to become accustomed to the presence of the observers. The children were already well adapted to the classroom before observer training was started. Observers were instructed to avoid all interactions with the students and teacher while in the class or on the school grounds. At the scheduled time they would enter the class, walk directly to chairs provided for them, sit down, and begin the observations. A hand signal was used to insure synchronization of observation times. Initially two observers were scheduled to observe on Monday, Wednesday, and Friday, and two on Tuesday and Thursday. When a systematic difference developed between the two sets of observers, one of the Tuesday-Thursday observers was placed on a three-day-a-week schedule to tie the two sets of observations together with reliability checks Thus, on some days there were as many as three observers in the classroom. The number of observers in the classroom varied from one to three. Due to illness or the need to obtain observations in other classroom, there were times when only one observer was available. Observers were not informed of changes in experimental conditions.

Classes of Teacher Behaviors: The Independent Variable

The behaviors emitted by the teacher were defined as belonging to three general classes: (1) Disapproving Behavior, (2) Approving Behavior, and (3) Instructional Behavior. Disapproving and Approving Behaviors were rated only when they immediately followed discriminable child behaviors falling into inappropriate or appropriate classes (see below).[2] Listings were made of the teacher behaviors that could occur within each class.

The general class of Disapproving Behaviors included Physical Contact, Verbal, and Facial subclasses. The subclasses of Physical behaviors included forcibly holding a child, grabbing, hitting, spanking, shaking, slapping, or pushing a child into position. The Verbal subclass of Disapproving Behaviors included yelling, scolding, raising voice, belittling, or making fun of the child, and threats. Threats included "if then" statements of loss of privilege or punishment at some future time. For example, the teacher might say to the class, "If you don't remain quiet, you will have to stay in from recess." The Facial subclass of Disapproving Behaviors included frowning, grimacing, side-to-side head shaking, gesturing, *etc.*

The general class of Approving Behaviors also included Physical Contact, Verbal, and Facial subclasses. Approving Physical Contacts included embracing, kissing, patting, holding hand or arm of child, or holding the child in the teacher's lap. Approving Verbal comments included statements of affection, approval, or praise. Approving Facial response was rated whenever the teacher smiled, winked, or nodded at one or more of the children.

The general class of Instructional Behavior included any response from teacher to children which involved giving instructions, information, or indicating correct responses.

In addition to recording the above classes of teacher behavior, note was taken of those times when the teacher terminated social interaction by turning out lights and saying nothing, turning her back on the class and waiting for silence, or stopping talking and waiting for quiet.

As noted earlier, the observers recorded every teacher response falling in a given class. Thus, the measures of teacher behaviors are frequency counts.

Child Behaviors: The Dependent Variable

The classes of child behaviors were developed by categorization of behaviors occurring with some frequency in the repertoire of problem children (Becker *et al.*, 1967). It was assumed that certain behaviors, because of their common topography, could be grouped together. Five classes of Disruptive Behavior (Gross Motor, Noise Making, Orienting, Verbalizations, and Aggression) and one class of Appropriate Behavior (Relevant) were defined. Behaviors not specifically defined were rated in a separate category (Other Task). Disruptive Behaviors were essentially behaviors apparently incompatible with good classroom learning conditions.

Included in the category of behaviors labeled as Gross Motor activities were. getting out of seat, standing up, walking around, running, hopping, skipping, jumping, rocking chair, moving chair, sitting with chair in aisle, kneeling in chair, arm flailing, and rocking body without moving chair.

The category of Noise Making was rated with the stipulation that the observers must hear the noise as well as see the noise making action, and included tapping feet, clapping, rattling papers, tearing papers, throwing books or other objects onto desks, slamming desk top, tapping objects on desk, kicking desk or chair, and scooting desk or chair.

The Verbalization category was rated only when the observer could hear the response. Lip movements alone were not rated. Carrying on conversations with other children, calling out teacher's name to get her attention, crying, screaming, singing, whistling, laughing, and coughing were included in the category.

The Orienting class of behaviors required that the child be seated. Turning of head or head and body toward another person, showing objects to another child, and looking at another child were rated. Looking behaviors

[2] As it turned out, approval following inappropriate behavior occurred only three times and disapproval following appropriate behavior did not occur. Also, this teacher did not make non-response contingent approval or disapproval comments. Thus, we were dealing essentially with two response contingent classes of teacher behavior.

of less than 1 sec duration were not rated except for any turn of more than 90 degrees from the desk. When an Orienting response overlapped two rating intervals, and could not be rated in the first interval, because it began too late in the interval to meet the .4-sec criterion, it was rated in the second interval.

Aggression was defined as hitting, pushing, shoving, pinching, slapping, striking with objects, poking with objects, grabbing objects or work belonging to another, knocking neighbor's property off desk, destroying another's property, throwing objects. No judgments of intent were made.

Appropriate behaviors were labeled Relevant and were made more easily identifiable by restricting the observations to a period in the morning when all of the children were preparing reading and workbook assignments. Specific Relevant Behaviors were looking at the teacher when she was speaking to the entire class or to the child being observed, answering questions of the teacher, raising hand and waiting for teacher to respond, writing answers to workbook questions, looking at pages of text in which reading was assigned. It was required that the entire 10 sec interval be filled with on task behavior before the Relevant rating was made.

When a child being observed performed a response not defined by one of the categories of Disruptive Behaviors or by Relevant Behavior, a rating of Other Task was made. The Other Task rating was incompatible with Relevant, but could be recorded in the same interval as any or all of the categories of Disruptive Behavior.

When rating the children, the observers were instructed to record each class of behaviors which appeared in an interval regardless of how many other classes had already been recorded in that interval. All five categories of Disruptive Behaviors and the Other Task category were compatible with each other. Relevant Behavior was incompatible with the other categories. No category of behavior was rated more than once in an interval. If a child was conversing with his neighbor, and he made two verbal responses in one interval, this class of behaviors was recorded only once. Thus, each child behavior measure was a record of intervals in which the response occurred, rather than a count of the number of discrete responses as in the recording of teacher's behavior.

The overall level of Disruptive Behaviors was defined as the percentage of intervals in which one or more Disruptive Behaviors occurred.

Reliability

Two types of reliability were calculated. Reliability I reflects simply the degree to which two observers obtained the same score for each category of behavior during a 20-min observation period. The smaller score is divided by the larger. Reliability I most appropriately applies to the data as reported in Fig. 1, since these are averages for an observation period. Random errors tend to cancel each other out when a score is based on a series of observations and a reliability measure should reflect the gain in accuracy obtained by averaging. For training purposes, and for greater confidence in the accuracy of the observation procedure, a second type of reliability was also calculated (Reliability II). Reliability II required that the same behavior category be recorded in the same interval by each observer to define an agreement. Reliability II was calculated by dividing the number of agreements by the number of agreements plus disagreements.

During the pre-baseline observer training, reliability checks were required for every observation. Before baseline observations were started, consistent reliabilities (Type II) above 80% were required. Reliability I data based on a weighted average of the reliabilities of the child-behavior codes are reported in Fig. 1, as are the average reliabilities by conditions for teacher behaviors. Comparable Reliability II data averaged 82.6% for child behaviors and 83.2% for teacher behaviors. Reliabilities for individual categories are well represented by these averages.

Sequence of Conditions

The first phase of the study ($Baseline_1$) consisted of measuring both teacher and child behaviors. No attempt was made to manipulate teacher behavior.

The second phase (No $Approval_1$) was defined by the absence of Approval Behaviors. The teacher discontinued the use of praise statements and used only contingent Disapproving Behaviors to control the children.

These phases were then repeated (Baseline$_2$, No Approval$_2$). At the beginning of No Approval$_2$ and throughout the rest of the study, the teacher carried a small "supermarket" adding machine with her to count the frequency of Disapproval Behaviors so that she could better monitor her behavior.

The fifth phase of the study, Frequent Disapproval, involved increasing the level of Disapproving Behaviors to approximately three times that given during Baseline$_1$ while continuing to withhold Approving Behaviors.

Phase 6 returned to the lower level of Disapproval (No Approval$_3$) and Phase 7 again returned to the baseline conditions (Baseline$_3$).

The teacher was instructed to maintain experimental conditions throughout the day, not just during the observation period. During the periods when praise was withheld beginning with No Approval$_2$, checks of the daily counts of Disapproving Responses obtained by the teacher with her counter corresponded closely to those which would have been predicted by extrapolation from the observation periods.

RESULTS

The relationships of greatest interest are the effects of presence and absence of Approval Behaviors on Relevant Behaviors and the effects of levels of Disapproval Behaviors on Disruptive Behaviors. Because of a systematic rater bias which entered into the data for Other Task Behavior (discussed later), and therefore also affected Relevant Behaviors incompatible with Other Task, greater emphasis is given to the analysis of Disruptive Behaviors in presenting the results.

Average Level of Disruptive Behavior

In Baseline$_1$ Disruptive Behaviors occurred in an average of 8.7% of the intervals observed. When Approving Behaviors were discontinued (No Approval$_1$), Disruptive Behavior increased to an average of 25.5% (Fig. 1). Approving Behaviors were again provided

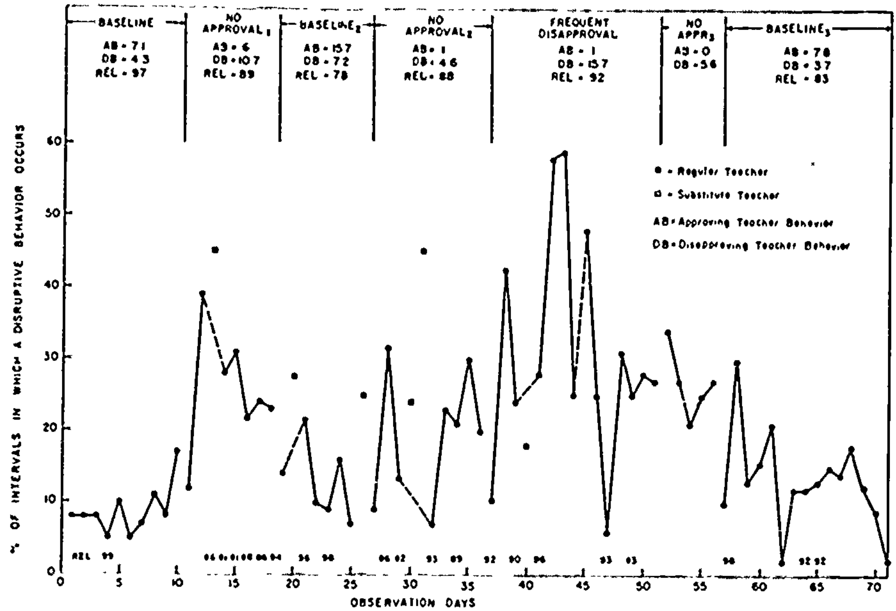


Fig. 1. Disruptive classroom behaviors as a function of nature of teacher behavior. Data points represent 2-min samples on 10 children each day. Dotted lines cross observations where the regular teacher was absent due to a recurrent illness, including a 10 day hospitalization between Days 39 and 41. The dotted line connecting days 44 and 45 represents the Easter vacation break. The data for Day 26 were taken with the teacher out of the room.

(Baseline$_2$) and Disruptive Behavior dropped to an average of 12.9%. In order to show more conclusively that the changes in Disruptive Behavior were related to the changes in teacher behavior, Approving Behaviors were again discontinued (No Approval$_2$) and the level of Disruptive Behaviors stabilized near the same level as in No Approval$_1$ condition (average 19.4%). When the Disapproving Behaviors (critical comments) were tripled (Frequent Disapproval), while Approving Behaviors were still withheld, Disruptive Behavior increased to an average of 31.2% with high points far above any observed before. The behavior stabilized, however, near the level at which the two previous No Approval phases had stabilized. When the rate of disapproval was lowered (No Approval$_3$), no great reduction in Disruptive Behavior occurred. The average level of Disruptive Behaviors over No Approval$_2$, Frequent Disapproval, and No Approval$_3$ was 25.9%. At the end of No Approval$_3$, Approval was again added to the low level of Disapproving Behaviors, and Disruptive Behavior dropped to an average of 13.2%, with the trend indicating a level far below this average.[3]

Analysis of Classes of Behavior

Discontinuation of approving behaviors. In reviewing the changes in the individual categories of behavior through the first two withdrawals of Approving Behavior, the majority of the increase in Disruptive Behaviors could be attributed to changes in Verbalization and Orienting categories (Table 1) The mean of Verbalization in No Approval$_1$ was 22 6% due to one extremely high observation on the second day of the condition, however, these behaviors stabilized between 9% and 17% (Fig. 2). Orienting showed a slight decrease across No Approval$_1$ (Fig. 2) The second time Approval was discontinued, Orienting increased across the condition while Verbalization remained relatively stable except for two high observations. Gross Motor behaviors followed the same pattern as Orienting and Verbalization through No Approval (1 and 2), increasing each time Approving Behavior was discontinued and decreasing when Approving Behaviors were present (Fig 2)

Noise Making and Aggression followed a pattern through No Approval$_1$ and $_2$ which was distinctly different from the other categories of disruptive behavior Both of these categories of behavior were already occurring at a low frequency in the Baseline condition (Table 1), but they occurred even less often when only Disapproving Behavior was given.

Increase of disapproving behaviors. In the Frequent Disapproval condition, Noise Making, Gross Motor, and Orienting all increased (Table 1). Verbalizations showed a decline

[3]A conservative statistical analysis was performed (*F* test) to compare those three conditions where approval responses were available with those two conditions where approval responses were withdraw For this test the Frequent Disapproval and No Approval$_{2,3}$ conditions were collapsed into one condition In order to insure independence of observations, the average values within each condition were used, thus providing four degrees of freedom Significant differences were found for Relevant Behavior ($p < 0.01$). Noise Making ($p < 0.05$), Gross Motor ($p < 0.025$), and for the overall level of Disruptive Behavior ($p < 0.01$)

Table 1

Average Percentages for Specific Behavior Classes for Each Experimental Phase

Behavior Classes	*Baseline$_1$*	*No Approval$_1$*	*Baseline$_2$*	*No Approval$_2$*	*Frequent Disapproval*	*No Approval$_3$*	*Baseline$_3$*
Disruptive Behaviors[a]	8.7	25.5	12.9	19.4	31.2	26.8	13.2
Gross Motor	2.7	6.7	2 0	4.8	,2.3	10.4	2.4
Noise	0.9	0.1	0.7	0 09	4.1	4.4	0.9
Verbalization	4.6	22 6	7.7	9.6	7.9	6.0	3.9
Orienting	1.4	6 5	4.1	7.1	11.5	10.2	7.6
Aggression	0.25	0.01	0.2	0.01	0.04	0 04	0.1
Other Task	7.0	10.4	5.9	10.7	5.9	4.2	1.2
Relevant	84.1	65 3	83.9	72.1	64.3	69.4	85 6

[a]The addition of percentages for the five classes of Disruptive Behaviors will usually lead to a sum higher than that reported as percentage of Disruptive Behaviors, since the latter does not reflect the occurrence of more than one subclass of Disruptive Behaviors in a given 10-sec interval.

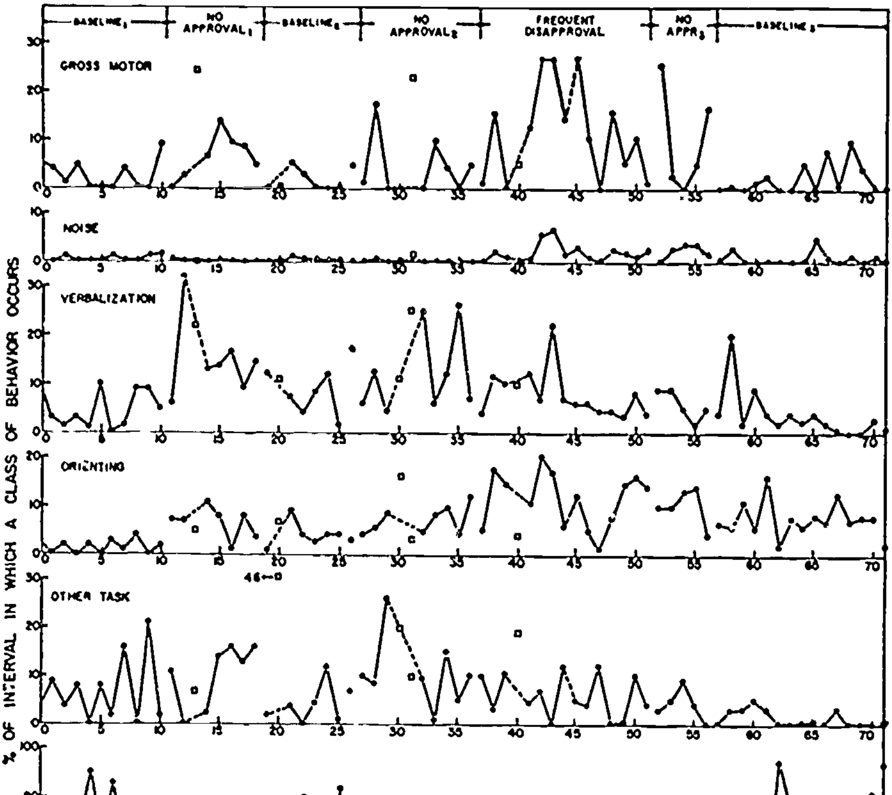


Fig 2 Analysis of specific behavior classes by condition. Data points represent 2-min samples on 10 children each day. See notes under Fig. 1.

over this condition and continued to decline through the rest of the study.

Changing from a high level of Disapproving Behaviors to a lower level did not markedly change the frequency of the various categories of Disruptive Behaviors relative to their terminal level under the Frequent Disapproval condition.

When Approving Behaviors were again used by the teacher (Baseline$_3$), the frequency of Gross Motor, Noise Making, and Orienting behaviors decreased noticeably (Fig. 2). Verbalization continued to show the steady decrease in frequency which had started in the Frequent Disapproval condition In Baseline, Aggression again occurred, but rarely. All Disruptive Behaviors except Orienting dropped to the level of the initial Baseline (or below) during the final Baseline.

Relevant behavior. Appropriate behaviors were initially high in the classroom (Fig. 2). Behaviors such as getting out of seat to move to a reading group or to check a completed workbook assignment were rated in the Gross Motor category. The requirements for such behaviors, however, remained constant through all conditions so changes in the level of Relevant Behaviors cannot be attributed to changes in classroom requirements. Relevant Behavior decreased each time Approving Behavior was discontinued and increased each time the Approval was reinstated. Relevant Behavior was at a slightly higher level during the final Baseline than during the initial Baseline.

Other task: Behavior not specifically defined. As indicated earlier, a systematic rater difference was encountered early the study in rating Other Task behaviors. In Fig. 2 this bias can be seen by contrasting data collected on Days 2, 4, 6, 8, and 10 from one set of observers with that collected on Days 1, 3, 5, 7, and 9 by another set of observers. While an attempt was made to correct this bias by interlocking reliability checks, it is apparent that the bias continued to some extent throughout the study. Since Other Task is by definition incompatible with Relevant Behavior, Relevant Behavior shows the same bias. By looking at Disruptive Behavior, defined so as to exclude Other Task behaviors, the systematic bias was largely eliminated from the data presented in Fig. 1.

Teacher Behaviors

The behavior of the teacher remained under good control throughout the study. Averages by conditions for Approving and Disapproving Behaviors are given in the upper part of Fig. 1. As the conditions were changed, little difficulty was found in withholding behaviors in the Approving category. Some difficulty was reported by the teacher in regulating the frequency of Disapproving Behaviors while withholding Approving Behaviors, but a partial solution to this problem was found. The teacher found that by carrying a small hand-counter (mentioned earlier) she could more accurately judge the frequency of her critical comments. In the Frequent Disapproval Phase there were days when the children were not emitting enough Disruptive Behaviors for critical comments to be appropriate at the programmed frequency. Rather than make inappropriate comments, the rate of Disapproving comments was adjusted to the frequency of the Disruptive Behaviors. When enough Disruptive Behaviors were available, Disapproving Behaviors were dispensed at a maximum rate of one per minute throughout the day, thus, many of the responses of the children were reprimanded very quickly.

General frequency of instructional comments did not change appreciably across conditions. However, the teacher did increase the frequency with which she would say in a neutral tone whether responses were correct or incorrect in the phases where Approval was not given.

The behaviors characterized as Terminating Social Interaction occurred only twice during the study and were, therefore, not subject to further analysis.

Substitute teachers. Observations taken on the days when a substitute teacher was in charge of the classroom appear in four conditions of the study. The frequency of Disruptive Behaviors increased in the presence of a temporary teacher as long as the regular teacher was in either Baseline or No Approving Behavior phases. When the Disapproving Behavior was being dispensed at a high rate, however, the level of Disruptive Behaviors decreased in the presence of a temporary teacher (Fig. 1).

Day 26. The data for this day were taken while the teacher was out of the room. Since the experimental conditions were not operative, this point should have been omitted altogether.

DISCUSSION

The results indicate that some aspects of the behaviors included in the category of Approving Behaviors were reinforcing for task appropriate behaviors. The frequency of Relevant Behaviors was high whenever Approving Behaviors followed Relevant child

Behavior, and decreased whenever Approving Behaviors were discontinued.

In each change of conditions that involved discontinuation of Approving Behaviors, there appeared a reliable transition effect (observation Days 11 and 27). This effect may be an example of the typical increase in rate found when a positive reinforcer is removed. In support of this explanation, the teacher reported, "When I stop praising the children, and make only negative comments, they behave very nicely for three or four hours. However, by the middle of the afternoon the whole classroom is chaotic." Since observations were taken during a study period in the morning, the periods of good behavior show up in the data each time a condition was changed. A similar low deviant behavior point occurred at the transition to Frequent Disapproving Behaviors (day 37), but it is not clearly explained. "The children seemed stunned."

Reviewing the individual classes of Disruptive Behaviors brings out certain similarities and differences among the classes. During the first alternations of Baseline with discontinuation of Approving Behaviors, Gross Motor, Orienting, and Verbalization Behaviors increased with discontinuation of Approval, while Noise Making and Aggressive Behaviors remained at their already low frequency. The increases are interpreted as suggesting that some responses in the disruptive classes may be reinforced by peer attention or other environmental circumstances when control through approving teacher responses to incompatible behaviors is withdrawn. For example, Orienting behaviors, such as looking around the room or out the window may be reinforced by seeing other children playing, by observing a custodian cleaning up the schoolyard, or by seeing any of numerous events which have no relationship to the classroom. Observational evidence for this inference was clearest in the Frequent Disapproval phase (below). It is also possible to attribute the increases in Disruptive Behaviors during No $Approval_1$ to the increase in use of Disapproval. However, the data for No $Approval_2$, where Disapproval was held to the Baseline level, would argue that the effect was primarily related to the withdrawal of approval.

Increasing Disapproving Behaviors to a high level produced four days where Disruptive Behaviors were above 40%. Several individual categories of behavior also showed marked changes. The increase in Gross Motor Behaviors was related to an increase in interactions with other students. During the Frequent Disapproval condition, two or three children would make alternate trips to check their workbooks at a table provided for that purpose. Only one child was permitted at the table at a time. During Baseline and No Approval phases, it was rare to see a child make more than one trip to the table, in the Frequent Disapproval phase, some of the children would check their papers several times. Others responded by pushing their papers off of their desks and then getting up to get them. There was a noticeable "pairing off" with two or more children exhibiting the same behaviors.

Another consequence of the Frequent Disapproval phase was a marked increase in the noise level in the room. A majority of the noises during this period were created by children scooting their desks and chairs. One observer reported, "I waited for a few minutes after the regular observation period was over and counted the noises. During one 40-sec period, I counted 17 separate chair scraping noises. They came in bursts of two or three at a time. It looked as though the kids were trying to irritate the teacher." The noises in "bursts of two or three" seemed similar to the "pairing off" of children noted with the Gross Motor behaviors, and strengthens an hypothesis that reinforcement from peers is one of the elements which accounts for the increase in Disruptive Behaviors during this time. Peer attention cannot be the only element affecting the behavior of the children, however, because the Verbalization category of behaviors showed a constant decrease throughout the Frequent Disapproval condition. The inhibition of Verbalization could be due to interfering emotional responses being elicited by the high level of critical comments by the teacher. More probable, however, is that the children simply talked more quietly to avoid being caught by the teacher. Observers' reports indicate that a substantial number of verbalizations would have been recorded during the Frequent Disapproving Behaviors condition if there had been no requirement that the responses be heard by the observers. The children could be seen to turn their heads, and lip movements could be seen frequently, but the verbalizations could not be heard.

Work by Lovaas, Freitag, Kinder, Rubenstein, Schaeffer, and Simmons (1964) suggests that for some children any adult attention may be reinforcing Some of the present findings under the Disapproval conditions could also be interpreted as indicating that teacher behavior of the Disapproving variety was positively reinforcing. The level of Disruptive Behaviors during each of the conditions when only Disapproving and Instructional attentions were available does appear to vary with the level of Disapproving Behaviors dispensed by the teacher Unfortunately, the illness-caused absences and Easter break make the results less clear than hoped. It should be apparent that the effect of Frequent Disapproval on the behavior of the children is not subject to a simple interpretation. Some criticized behaviors decreased, some increased, and several possible controlling stimuli could have been operating with contradictory effects on behavior It is obviously difficult in a field experimental study of this complexity to maintain control of all the possibly relevant variables at once.

Another limitation of the present design should be noted. Because of a shortage of observation time under the desired classroom conditions, a sample of 10 children was observed daily A procedure which included all children each day would have provided a stronger basis for analysis of effects on individuals A rough analysis of individuals with the present data confirms, however, that an average of 76% of the students made changes in the same direction as the group changes. From Baseline, to No Approval, 81% of the students showed increases in Disruptive Behavior When Approving Behavior became available, 75% of the students improved within two weeks. Discontinuing Approving Behavior a second time resulted in 78% of the students being more disruptive, while the final addition of Approving Behavior showed an increase in appropriate behavior for 71% of the children Across condition changes, 5% of the children showed no change on the average, and 19% showed change (usually minor) in an opposite direction. Procedures which permitted specifications of which children were praised or criticized for which behaviors would be needed to clarify fully individual effects. It is quite possible that the children who changed opposite to the group trend were being responded to differently. Of course, there are many ways one can speculate here. In an as yet unpublished study we have shown that praising some children but not others leads to changes in the behavior only for the children who are praised. Results of this sort emphasize the importance of looking at individual contingencies.

Brief mention should be made of the possible ethical considerations involved in producing disruptive behaviors. One needs to weigh the potential gains in knowledge against the short-term or long-term deleterious effects on the children or teacher. On the basis of prior research and the return to baseline after the first No Approving Behaviors condition, the teacher and the experimenters were confident that appropriate behaviors could be readily reinstated at any time it was felt necessary. It may also be reassuring to know that this accelerated middle primary class did achieve well academically during the year. The children completed all second and third grade work and were all performing on a fourth grade level by the end of the year.

IMPLICATIONS

This further demonstration of the importance of specific teacher behaviors in influencing classroom behavior has a double implication. First, the teacher who uses her Approving Behaviors as immediate consequences for good behavior should find that the frequency and duration of appropriate behaviors increase in her classroom (at least for most children). On the other hand, the teacher who cuddles the miscreant, tries pleasantly to get a child to stop behaving disruptively, talks with a child so that he "understands" what he was doing wrong, or who pleasantly suggests an alternative activity to a child who has been performing inappropriately, is likely to find an increase in the very behaviors she had hoped to reduce. This view of the functional importance of teacher's behavior in creating, maintaining, or reducing classroom behavior problems contrasts sharply with that generated by psychodynamic models of problem behaviors and what to do about them. Work of this sort also suggests a need to re-evaluate the popular cliche about the importance of the interaction of the "personality" of the teacher with that of the child in looking at classroom management procedures.

The suggestive evidence that peer reinforcement (among other stimuli) takes over when social reinforcement is not provided by teacher is given support by the recent work of Wahler (1967). Wahler has shown how preschool children can systematically control the behavior of their peers by differential use of social reinforcement. The more general implication for the teacher is this: unless an effort is made to support desirable classroom behaviors with appropriate consequences, the children's behavior will be controlled by others in ways likely to interfere with the teacher's objectives.

Finally, the possibility that critical comments may actually function to increase some behaviors upon which they are contingent cannot be overlooked. A recent study (Madsen, Becker, Thomas, Koser, and Plager, 1967), gives clear evidence that some forms of critical comment do function to strengthen behavior. The more often a teacher told first graders to "sit down", the more often they stood up. Only praising sitting seemed to increase sitting behavior.

REFERENCES

Becker, W. C., Madsen, C. H., Jr., Arnold, Carole R., and Thomas, D. R. The contingent use of teacher attention and praise in reducing classroom behavior problems. *Journal of Special Education*, 1967, 1, 287-307.

Bijou, S. W. and Baer, D. M. Some methodological contributions from a functional analysis of child development. In L. P. Lipsitt and C. S. Spiker (Eds.), *Advances in child development and behavior*. New York: Academic Press, 1963. Pp. 197-231.

Hart, Betty M., Allen, K. Eileen, Buell, Joan S., Harris, Florence R., and Wolf, M. M. Effects of social reinforcement on operant crying. *Journal of Experimental Child Psychology*, 1964, 1, 145-153.

Kerr, Nancy; Meyerson, L., and Michael, J. A procedure for shaping vocalizations in a mute child. In L. P. Ullman and L. Krasner (Eds.), *Case studies in behavior modification*. New York: Holt, Rinehart, & Winston, Inc., 1965. Pp. 366-370.

Lövaas, O. I., Freitag, G., Kinder, M. I., Rubenstein, D. B., Schaeffer, B., and Simmons, J. B. Experimental studies in childhood schizophrenia—Establishment of social reinforcers. Paper delivered at Western Psychological Association, Portland, April, 1964.

Madsen, C. H., Jr., Becker, W. C., Thomas, D. R., Koser, Linda, and Plager, Elaine. An analysis of the reinforcing function of "sit down" commands. In R. K. Parker (Ed.), *Readings in educational psychology*. Boston: Allyn and Bacon, 1968 (in press).

Wahler, R. G. Child-child interactions in free field settings: Some experimental analyses. *Journal of Experimental Child Psychology*, 1967, 5, 278-293.

Zimmerman, Elaine H. and Zimmerman, J. The alteration of behavior in a special classroom situation. *Journal of the Experimental Analysis of Behavior*, 1962, 5, 59-60.

Learning Theory Approaches to Classroom Management: Rationale and Intervention Techniques

HARVEY F. CLARIZIO, Ed.D.
STEPHEN L. YELON, Ph.D.

JOURNAL OF SPECIAL EDUCATION, 1967.
Vol. 3, pp 267 274

Today, more than ever before, there is acute concern about the mental health of children. Traditionally, we have modeled intervention efforts after the clinical concept of treatment. Dissatisfaction with the limitations of psychotherapeutic intervention (Levitt, 1957) together with the professional manpower shortage in the mental health field has led, however, to suggestions, e.g., Redl (1962), that we need new modes of treatment, closer to real-life situations, if we are to tackle children's problems more effectively.

When psychodynamic models were the preferred method of treatment, teachers were accorded at best a second-string status on the clinical team helping emotionally handicapped children. The increasing popularity of behavior therapy and other approaches based on learning theory now offers teachers opportunities for an integral role in the quest for better mental health for children. Indeed, it might well be the mental health specialist who will now assume the supportive role (Gallagher & Chalfant, 1966) in the "treatment" of children.

In the application of learning theory principles to the modification of deviant behavior, the emphasis is on the changing of behavior with little attention devoted to the etiology of the behavior. Why should teachers focus primarily on the behavior rather than on its causes? There are several reasons:

1. First, teachers by virtue of their orientation are not trained to probe the causes of behavior that even mental hygiene specialists often consider obscure and uncertain. Hence, is it really helpful to ask the teacher to understand the causes underlying children's disturbed behavior?

2. Teachers in any case are rarely in a position wherein they can directly manipulate the causes so as to modify their influence on the child's classroom adjustment. For example, if the problem lies in the parent-child relations or in a brain lesion, there are few if any constructive intervention techniques that the teacher can employ. Yet the child's troublesome behavior persists and must be handled as effectively as possible when it occurs in the classroom.

3. Even in such occasional cases where the causes can be identified and manipulated directly, the maladaptive behaviors may persist. Thus, despite the discovery and correction of the contributing role of poor vision and faulty child-rearing practices in a reading disability case, a pupil may continue to experience difficulty with his reading until attention is *specifically* devoted to his reading behavior and unless he can experience success in this specific area, his mental health will continue to be impaired.

4. Behaviors or symptoms or habits may in their own right be incapacitating and disturbing, and current persisting symptoms may themselves be producing emotional disturbance (Franks, 1965) above and beyond the core disturbance from which the child is suffering. And, as research indicates (White & Harris, 1961), it is difficult to disentangle educational and emotional maladjustments in the school-age child (Gallagher & Chalfant, 1966).

5. There is little substantial evidence to indicate that if the teacher assists the child in modifying his behavior or symptoms, other undesirable behaviors will inevitably take their place in the manner of symptom substitution (Grossberg, 1964).

6. Finally, and most importantly, as already implied, the teacher most commonly has no resort other than to deal with the pupil's behavior as it appears in the here and now. As Lewis (1965) attests:

> If we cannot aspire to reconstruction of personality that will have long range beneficial effects, we can modify disturbing behavior in specific ways in present social contexts. This more modest aspiration may not only be more realistic but it may be all that is required of the child-helping professions in a society that is relatively open and provides a variety of opportunity systems in which a child can reconcile his personal needs with society's expectations of him.

Having argued that the teacher should be primarily concerned with behavior *per se* rather than with its causes, let us turn to techniques emanating from learning theory which have relevance to the modification of deviant behavior in the classroom (see Glossary of Terms on the next page). Although the techniques to be presented are discussed separately for the sake of clarity, it should be recognized that more than one of them may be operating at any given time in real-life attempts to modify behavior. Moreover, common to all of these techniques is the use of "systematic environmental contingencies to alter the object's responsiveness to stimuli" (Krasner & Ullman, 1965).

THE TECHNIQUES

Extinction. There is a growing body of research demonstrating that simple withdrawal of reinforcers can reduce or eliminate such troublesome behavior as excessive talking, tantrum behavior and academic errors (Warren, 1965; Williams, 1959; Zimmerman & Zimmerman, 1962). Extinction is not always, however, the most economic and effective means of producing behavioral change (Bandura & Walters, *Glossary of Terms*

Behavior Therapy. A therapeutic process in which the primary goal is to change overt behavior rather than to restructure an individual's personality makeup. The process uses principles of learning for its methodological source.

Extinction. The decrease and eventual disappearance of a response learned under conditions of reinforcement when the reinforcement is withheld.

Reinforcement. Whatever serves to maintain the occurrence or increase the strength of a response, e.g., food, water or the avoidance of punishment.

Partial Reinforcement. A condition in which subjects receive reinforcement only at various time intervals or after a certain number of responses.

Positive Reinforcement. Much the same as reinforcement; i.e., *presenting* a pleasant stimulus when a response occurs, as opposed to negative reinforcement where an unpleasant stimulus is *removed* when a response occurs.

Modeling. A condition where the behavior to be acquired is demonstrated for the learner.

Punishment. A condition where a learner is made to feel uncomfortable by being presented an unpleasant stimulus, e.g., the infliction of pain by hitting, and/or a condition where a pleasant stimulus is withdrawn so that the learner is made to feel discomfort, e.g., having treats withdrawn.

1963). Certain cautions should be recognized:

1. Spontaneous remission—the return of undesirable behavior—may occur following the extinction trials, thus necessitating additional extinction sessions.

2. When behavior is maintained on a partial reinforcement schedule, removal of the reinforcers may actually produce an increase in the frequency and intensity of the deviant responses. Moreover, it is sometimes extremely difficult not to reinforce maladaptive behaviors in a school setting, since circumstances may be beyond the teacher's control. The aggressive youngster who kicks the teacher or a classmate cannot help but be reinforced by the look of pain on the victim's face. The needed cooperation of classmates in the application of extinction procedures may also be difficult to secure, so that by necessity the deviant behavior is established on a partial reinforcement schedule.

3. General observation suggests that certain behaviors do not diminish and disappear simply because reinforcers are withdrawn, and sometimes teachers cannot or will not wait long enough to permit the completion of the extinction process. These limitations are particularly acute in situations in which emotional contagion is a distinct possibility. Behaviors seriously injurious to the self would also seemingly not lend themselves well to this technique. In brief, this method of behavior change has proven to be of value with acting-out as well as inhibited youngsters. Yet, its limitations suggest that other methods of behavioral modification are at times more economical and effective (see also Ausubel, 1957, Bandura, 1965).

Positive Reinforcement. Operant conditioning techniques constitute one of the main tools of behavior modification. In this technique, emphasis is placed on the response made by the individual, and only minimal attention is given to the stimuli eliciting the response. Essentially, the teacher presents a reward whenever the child emits the desired response. While teachers have been cognizant of the value of positively reinforcing "good" behavior, there is ample evidence to suggest that even "good" teachers not uncommonly reinforce undesirable behavior. One of the merits of the positive reinforcement technique stems from its applicability to antisocial youngsters as well as to withdrawn children (Bandura & Walters, 1963).

There has been a dearth of psychotherapeutic approaches designed for the conduct problem child, despite such pupils typically being the most disruptive of classroom procedures. The application of positive reinforcement principles to seriously aggressive children involves the manipulation of three variables, the schedules of reinforcement, the interval factor and the type of reward. With respect to the concept of reinforcement schedules, a distinction must be enforced between the acquisition and the maintenance of behavior. For the former, continuous or full-schedule reinforcement or reward after each appearance of the desired behavior is most effective, whereas for the latter, partial or intermittent reinforcement is most economical and effective. The interval variable merely refers to the passage of time between the production of a response and the presentation of the reward or reinforcer. The delay factor should usually be quite short initially, because acting-out children typically have difficulty in postponing gratification. Step by step, the interval can be lengthened as the child acquires more adequate behavioral controls.

The rewards for such pupils, at the start, may have to be tangible or physical in nature but should always be paired with verbal social reinforcers, e.g., "You handled yourself well in that situation today" (Quay, 1963). Gradually, the reinforcers can be shifted away from the concrete into language and other symbolic forms of reward until the child can respond satisfactorily to them. In deciding upon the most suitable reinforcers, consideration should be given to such factors as the child's developmental level and socio-cultural background.

The main unresolved question with the technique of positive reinforcement centers around the question of how to make the child initiate the response in the first place so that he can be rewarded (Franks, 1965). The technique of social modeling may well provide at least a partial answer to this problem (Baer, 1963, Ferster, 1961, Hewett, 1964, Staats, 1960, Wolf, Risley & Mees, 1964).

Modeling. Modeling is based on the premise that a child will imitate the behavior of others. Modeling is important in that children commonly acquire social skills through imitation of and identification with examples of socially approved behaviors presented by suitable models. School teachers thus have a unique opportunity to influence the behavior of entire groups of children. However, this technique has been typically overlooked in the management or modification of deviant behavior in schools. Modeling procedures may represent a more effective means than positive reinforcement of establishing new response patterns in children (Bandura, 1965). Moreover, a behavior pattern, once acquired through imitation, is often maintained without deliberate external reinforcement, because human beings learn to reinforce themselves for behaving in certain ways. Teacher training institutions have long recognized the importance of modeling procedures in the training of future teachers and, accordingly, attempt to provide adequate models in the form of critic teachers. However, attention should now be devoted to the teacher's use of modeling procedures in influencing the behavior of the pupils.

There are three effects of exposure to models: the *modeling effect*, the *inhibitory* or *disinhibitory effect*, and the *eliciting effect* (Bandura, 1965). Through the *modeling effect* children come to acquire responses that were not previously a part of their behavior. As noted earlier, modeling procedures may be considerably more economical in establishing new responses than the method of operant conditioning based on positive reinforcement, especially when a combination of verbalizing models and demonstration procedures are used. The strengthening or weakening of inhibitory responses already existing in the observer (the *inhibitory* or *disinhibitory effect*) can also be accomplished through modeling procedures. Children, for example, who see a model punished or rewarded for aggressive behavior tend to decrease or increase their aggressive behavior accordingly. The *eliciting* or *response facilitation effect* refers to the teacher's eliciting responses that precisely or approximately match those exhibited by the model. Thus, observation of the teacher's response provides discriminative clues that trigger similar responses already in the pupil's behavior repertoire. This eliciting effect is distinguished from the modeling and the disinhibiting effects in that the imitated behavior is neither new nor previously punished.

The probability that a child will imitate a model is a function of several variables. Modeling is partly dependent upon the reinforcing consequences of the model's behavior. Thus, if a model is rewarded for his socially approved behavior, the likelihood that the observer will behave in a socially approved manner is increased. Other factors include the process of attending to the model's behavior, e.g., previous training in observation, and various environmental stimuli, e.g., the complexity of the stimuli (Baldwin, 1967, Bandura, 1962b, Bandura & Hutson, 1961, Bandura & Kupers, 1964, Bandura, Ross & Ross, 1963).

Punishment. Aversive conditioning or punishment is an intervention technique which has been used primarily to discourage undesirable behavior. This technique consists in the presentation of either physically or psychologically painful stimuli or the withdrawal of pleasant stimuli when undesirable behavior occurs. The use of punishment as a technique for behavioral modification has been contraindicated for the following reasons:

1. Punishment does not eliminate the response, it merely slows down the rate at which the troublesome behaviors are emitted.

2. This technique serves notice to stop certain negative behaviors, it does not indicate what behaviors are appropriate in the situation.

3. Aggressive behaviors on the teacher's part may provide an undesirable model for the pupil.

4. The emotional side effects of punishment, such as fear, tenseness and withdrawal are maladaptive.

5. Punishment serves as a source of frustration which is apt to elicit additional maladaptive behaviors.

Some psychologists, who are currently reconsidering the concept of punishment, contend that it can have a beneficial effect if applied to specific responses rather than to general behavior (Marshall, 1965).

Teachers, whatever their motivations, use verbal reprimands and other forms of correction in their approach to classroom management, and the judicious use of punishment as an intervention technique is most likely necessary in that it is impossible to guide behavior effectively with positive reinforcement and extinction alone. As Ausubel (1957) asserts, "It is impossible for children to learn what is *not* approved and tolerated simply by generalizing in reverse from the approval they receive for the behavior that *is* acceptable." Thus, punishment of specific responses can have an informative and beneficial effect. A particular positive value that may accrue from the use of punishment is that undesirable behaviors are held in abeyance, thus permitting the teaching of desirable modes of behavior through such intervention techniques as social imitation or positive reinforcement. Although punishment techniques have been used primarily with acting-out pupils, they have also been found to be of value in certain cases of withdrawn behavior (Bandura, 1962a, Church, 1963, Lovaas, 1965, Meyer & Offenbach, 1962, Redl, 1965, Sears, Maccoby & Levin, 1957, Solomon, 1964).

Discrimination Learning. Children sometimes engage in maladaptive behavior because they have transferred behaviors acceptable in one setting to a second setting where these behaviors are considered inappropriate and maladaptive. Thus, for example, the child who is overly dependent upon his mother may behave in a very dependent way toward his teacher. Such cases of inappropriate generalization can sometimes be remediated through the use of discrimination learning. Essentially this process consists of labeling given behaviors as appropriate within a specific environmental context. The teacher in the above case, for example, may inform the child in a nonpunitive way that she is not his mother but his teacher and that as such she will require him to become more self-reliant. This labeling by the teacher makes the child more aware of both inappropriate and appropirate behaviors. Interestingly, children do not always have to be able to express such discriminations verbally in order to achieve "insight" into their behavior. It is rather required, to insure effective results, that appropriate responses be rewarded and undesirable responses discouraged. Discrimination learning thus may be of service in conjunction with most other techniques in managing conduct and personality problems in the classroom (Ayllon & Michael, 1959, Barrett & Lindsley, 1962, Brackbill & O Hara, 1958, Penny & Lupton, 1961, Stevenson, Weir & Zigler, 1959).

Desensitization. Densensitization as an intervention technique has been used principally with the fearful and phobic child. The basic objective is to have the child achieve a relaxed response in the presence of what were previously anxiety-producing stimuli. To accomplish this relaxed response, the subject is encouraged to perform approximations of previously punished acts within non-punishing or actually rewarding situations. Or through gradual exposure to the feared object or situation, a subject may become able to perform a formerly feared act or approach the feared object in a relaxed manner (Bentler, 1962, Garvey & Hegrenes, 1966, Jersild & Holmes, 1935, Lazarus, 1960; Wolpe, 1958).

CONCLUDING REMARKS

As evidenced by our discussion of the limitations of each technique, we do not envision management techniques emanating from learning theory as a panacea, but these intervention techniques do have certain potential advantages:

1. The fruitfulness of these techniques in modifying human behavior has been demonstrated in laboratory settings as well as in natural settings,

2. They are consistent with the teacher's role whereby she must reflect cultural expectations and set standards for her pupils' academic and social behavior.

3. Behavioral approaches offer specific and practical techniques for use in day-to-day classroom problems. While teachers already use some or all of these techniques, they frequently do so intuitively or inconsistently thereby reducing their efficacy.

4 These techniques enable the teacher to strive toward more realistic and obtainable goals relative to their pupils' mental health.

5. One of the most important attributes of these techniques is the fact that they can be taught to teachers. While there are few if any teacher training institutions currently offering didactic and practice training in such techniques, one can envision the time when teachers will acquire such skills through laboratory courses taken in conjunction with their formal course work or through in-service meetings and workshops.

References

Ausubel D *Theory and problems of child development* New York Grune & Stratton 1957

Ayllon, T & Michael, J The psychiatric nurse as a behavioral engineer *Journal of Experimental Analysis of Behavior*, 1959 2, 323-334

Baer D Effect of withdrawal of positive reinforcement on an extinguishing response in young children *Child Development*, 1961, 32, 67-74

Baer D Social reinforcement and behavior change *American Journal of Orthopsychiatry*, 1963 591-633

Baldwin A Theories of child development *Critique of social learning theory* Chapter 16 New York Wiley 1967

Bandura A Punishment revisited *Journal of Consulting Psychology*, 1962 26, 298-301(a)

Bandura, A Social learning through imitation In M Jones (Ed), *Nebraska Symposium on Motivation* Lincoln Nebraska University of Nebraska Press, 1962 Pp 211-269(b)

Bandura A Behavioral modification through modeling procedures In L Krasner & L Ullman (Eds.), *Research in behavior modification* New York Holt, Rinehart & Winston, 1965

Bandura A & Huston A Identification as a process of incidental learning *Journal of Abnormal and Social Psychology*, 1961, 63, 311-318

Bandura A & Kupers, C The transmission of patterns of self-reinforcement through modeling. *Journal of Abnormal and Social Psychology*, 1964 69, 1-19

Bandura, A, Ross D & Ross S Imitation of film mediated aggressive models *Journal of Abnormal and Social Psychology*, 1963, 66 3-11

Bandura, A, & Walters R *Social learning and personality development* New York Holt, Rinehart & Winston, 1963

Barrett B & Lindsley O Deficits in acquisition of operant discrimination and differentiation shown by institutionalized retarded children *American Journal of Mental Deficiency*, 1962, 67, 424-436

Bentler, P An infant s phobia treated with reciprocal inhibition therapy *Journal of Child Psychology and Psychiatry*, 1962, 3, 185-189

Brackbill, Y & O'Hara, J The relative effectiveness of reward and punishment for discrimination learning in children *Journal of Comparative and Physiological Psychology*, 1958 51 747-751.

Church, R The varied effects of punishment on behavior *Psychological Review* 1963 70, 369-402

Ferster C Positive reinforcement and behavioral deficits of autistic children *Child Development*, 1961, 32, 437-456

Franks, C Behavior therapy, psychology and the psychiatrist contribution evaluation and overview *American Journal of Orthopsychiatry*, 1965, 35, 145-151

Gallagher, J & Chalfant J The training of educational specialists for emotionally disturbed and socially maladjusted children In *N S S E yearbook 1966 social deviancy among youth* Chicago University of Chicago Press, 1966 pp 398-423

Garvey, W & Hegrenes, J Desensitization techniques in the treatment of school phobia, *American Journal of Orthopsychiatry*, 1966, 36, 147-152

Grossberg, J Behavior therapy a review *Psychological Bulletin*, 1964 62, 73-88

Hewett, F Teaching reading to an autistic boy through operant conditioning *The Reading Teacher*, 1964, 17 613 618.

Jersild A & Holmes, F Methods of overcoming children's fears *Journal of Psychology*, 1935, 1, 75-104.

Krasner, L & Ullman, L *Case studies in behavior modification* New York Holt Rinehart & Winston, 1965

Lazarus, A The elimination of children's phobias by deconditioning In H Eysenck (Ed) *Behavior therapy and the neuroses* New York Pergamon Press 1960 Pp 114-122

Levitt, E E Results of psychotherapy with children an evaluation *Journal of Counseling Psychology*, 1957 21, 189-196

Lewis, W Continuity and intervention in emotional disturbance a review *Exceptional Child*, 1965, 31, 465-475

Lovaas, I Building social behavior in autistic children by use of electroshock *Journal of Experimental Research in Personality*, 1965, 1, 99-109.

Marshall, H The effect of punishment on children, a review of the literature and a suggested hypothesis *Journal of Genetic Psychology*, 1965, 106, 108-133.

Meyer, W & Offenbach, S Effectiveness of reward and punishment as a function of task complexity *Journal of Comparative and Physiological Psychology*, 1962, 55, 532 534.

Penny, R O & Lupton, A Children's discrimination learning as a function of reward and punishment *Journal of Comparative and Physiological Psychology*, 1961, 54, 449-456

Quay, H Some basic considerations in the education of emotionally disturbed children. *Exceptional Child*, 1963, 30, 27-31.

Redl, F Crisis in the children's field *American Journal of Orthopsychiatry*, 1962, 32, 759-780.

BEST COPY AVAILABLE

Redl F The concept of punishment In N. Long, W Morse & R Newan (Eds). *Conflict in the classroom* Belmont, Calif Wadsworth, 1965.

Sears R, Maccoby, E & Levin, H *Patterns of child rearing* Evanston, Ill Row Peterson, 1957.

Slack, C Experimenter-subject psychotherapy: a new method for introducing intensive office treatment for unreachable cases. *Mental Hygiene*, 1960, 44, 238-256.

Solomon, R Punishment *American Psychologist*. 1964, 19, 239-253.

Stevenson, H, Weir, M, & Zigler, E. Discrimination learning in children as a function of motive-incentive conditions *Psychological Report*, 1959, 5, 95-98

Warren, A All's quiet in the backroom Paper read at the Council for Exceptional Children, Wichita, Kans, Oct, 1965

White, M & Harris, M *The school psychologist*. New York Harper, 1961

Williams, C D The elimination of tantrum behavior by extinction procedures *Journal of Abnormal and Social Psychology*, 1959, 59, 269

Wolf, M, Risley, T & Mees, H. Application of operant conditioning procedures to behavior problems of an autistic child. *Behavior Research and Therapy*, 1964, 1, 305-312

Wolpe, J *Psychotherapy by reciprocal inhibition*. Stanford, Calif Stanford University, 1958.

Zimmerman, E & Zimmerman, J The alternation of behavior in a special classroom situation. *Journal of Experimental Analysis of Behavior*, 1962 5, 59-60.

A TOKEN REINFORCEMENT PROGRAM IN A PUBLIC SCHOOL: A REPLICATION AND SYSTEMATIC ANALYSIS[1]

K. D. O'LEARY, W. C. BECKER, M. B. EVANS, AND R. A. SAUDARGAS

A base rate of disruptive behavior was obtained for seven children in a second-grade class of 21 children. Rules, Educational Structure, and Praising Appropriate Behavior while Ignoring Disruptive Behavior were introduced successively, none of these procedures consistently reduced disruptive behavior. However, a combination of Rules, Educational Structure, and Praise and Ignoring nearly eliminated disruptive behavior of one child. When the Token Reinforcement Program was introduced, the frequency of disruptive behavior declined in five of the six remaining children. Withdrawal of the Token Reinforcement Program increased disruptive behavior in these five children, and reinstatement of the Token Reinforcement Program reduced disruptive behavior in four of these five. Follow-up data indicated that the teacher was able to transfer control from the token and back-up reinforcers to the reinforcers existing within the educational setting, such as stars and occasional pieces of candy. Improvements in academic achievement during the year may have been related to the Token Program, and attendance records appeared to be enhanced during the Token phases. The Token Program was utilized only in the afternoon, and the data did not indicate any generalization of appropriate behavior from the afternoon to the morning.

Praise and other social stimuli connected with the teacher's behavior have been established as effective controllers of children's behavior (Allen, Hart, Buell, Harris, and Wolf, 1964, Becker, Madsen, Arnold, and Thomas, 1967, Brown and Elliot, 1965, Hall, Lund, and Jackson, 1968, Harris, Johnston, Kelley, and Wolf, 1964, Harris, Wolf, and Baer, 1964, Scott, Burton, and Yarrow, 1967, Zimmerman and Zimmerman, 1962). When the teacher's use of praise and social censure is not effective, token reinforcement programs are often successful in controlling children (Birnbrauer, Wolf, Kidder, and Tague, 1965; Kuypers, Becker, and O'Leary, 1968, O'Leary and Becker, 1967, Quay, Werry, McQueen, and Sprague, 1966, Wolf, Giles, and Hall, 1968).

The token reinforcement program utilized by O'Leary and Becker (1967) in a third-grade adjustment class dramatically reduced disruptive behavior. In order to maximize the possibility of reducing the disruptive behavior of the children, O'Leary and Becker used several major variables simultaneously. The first objective of the present study was to analyze the separate effects of some of the variables utilized in the former study. More specifically, the aim was to examine the separate effects of Classroom Rules, Educational Structure, Teacher Praise, and a Token Reinforcement Program on children's disruptive behavior. Rules consisted of a list of appropriate behaviors that were reviewed daily. Educational Structure was the organization of an academic program into specified 30-min lessons such as spelling and arithmetic. The second objective was to assess whether a Token Reinforcement Program used only in the afternoon had any

[1]Portions of this paper were presented to the American Psychological Association, September, 1968, San Francisco, California. This research was supported primarily by Research Grant HD 00881 05 to Wesley C. Becker from the National Institutes of Health and secondarily by a Biomedical Science Grant 31-8200 to K. Daniel O'Leary from the State University of New York at Stony Brook. The authors are grateful to Nancy Brown, Connie Dockterman, Pearl Dorfmann, Jeanne Kappauf, Margery Lewy, Stanley Madsen, and Darlene Zientarski who were the major observers in this study. Appreciation for support of this study is expressed to Dr. Lowell Johnson, Director of Instruction, Urbana Public Schools, and to Mr. Richard Sturgeon, elementary school principal. The greatest thanks goes to Mrs. Linda Alsberg, the teacher who executed the Token Reinforcement Program and tolerated the presence of observers both morning and afternoon for eight months. Her patience and self control during the Praise and Withdrawal Phases of the program were especially appreciated. Reprints may be obtained from K. Daniel O'Leary, Dept. of Psychology, State University of New York at Stony Brook, Stony Brook, N.Y. 11790.

JOURNAL OF APPLIED BEHAVIOR ANALYSIS, 1969, Vol. 2, pp. 3-13

effect on the children's behavior in the morning Third, the present study sought to examine the extent to which the effects of the Token Reinforcement Program persisted when the Token Program was discontinued.

METHOD

Subjects

Seven members of a second-grade class of 21 children from lower middle class homes served At the beginning of the school year, the class had a mean age of 7 yr. 5 months, a mean IQ score of 95 (range 80 to 115) on the California Test of Mental Maturity, and a mean grade level of 1 5 on the California Achievement Test. The class was very heterogeneous with regard to social behaviors. According to the teacher, three of the children were quite well behaved but at least eight exhibited a great deal of undesirable behavior. The teacher, Mrs. A., had a master's degree in counseling but had only student teaching experience She was invited to participate in a research project involving her class and received four graduate credits for participating in the project.

Observation

Children. Mrs. A. selected seven children for observation. All seven children were observed in the afternoon and four of the seven (S1, S2, S4, and S6) were also observed in the morning. Morning observations were made by a regular observer and a reliability checker from 9 30 to 11.30 every Monday, Wednesday, and Friday. Afternoon observations were made by two regular observers and a reliability checker from 12.30 to 2.30 every Monday, Wednesday, and Friday. Observations were made by undergraduate students who were instructed never to talk to the children or to make any differential responses to them in order to minimize the effect of the observers on the children's behavior. Before Base Period data were collected, the undergraduates were trained to observe the children over a three-week period in the classroom, and attention-seeking behaviors of the children directed at the observers were effectively eliminated before the Base Period.

Each child was observed for 20 min each day. The observers watched the children in a random order. Observations were made on a 20-sec observe, 10-sec record basis, i.e., the observer would watch the child for 20 sec and then take 10 sec to record the disruptive behaviors which had occurred during that 20-sec period. The categories of behavior selected for observation were identical to those used by O'Leary and Becker (1967). Briefly, the seven general categories of disruptive behavior were as follows: (1) *motor behaviors:* wandering around the room, (2) *aggressive behaviors.* hitting, kicking, striking another child with an object, (3) *disturbing another's property.* grabbing another's book, tearing up another's paper, (4) *disruptive noise.* clapping, stamping feet, (5) *turning around.* turning to the person behind or looking to the rear of the room when Mrs. A. was in the front of the class, (6) *verbalization.* talking to others when not permitted by teacher, blurting out answers, name-calling, and (7) *inappropriate tasks.* doing arithmetic during the spelling lesson.

The present study was a systematic replication of O'Leary and Becker (1967). To facilitate comparison of the two studies, the dependent measure reported is the percentage of intervals in which one or more disruptive behaviors was recorded. Percentages rather than frequencies were used because the length of the observations varied due to unavoidable circumstances such as assemblies and snow storms. Nonetheless, most observations lasted the full 20 min, and no observation lasting less than 15 min was included.

Teacher. In order to estimate the degree to which the teacher followed the experimental instructions, Mrs. A. was observed by two undergraduates for 90 min on Tuesday and Thursday afternoons. Teacher behavior was not observed on Monday, Wednesday, and Friday when the children were observed because Mrs. A. understandably did not wish to have as many as five observers in the room at one time. Furthermore, because Mrs. A. was somewhat reluctant to have three regular observers and one or two graduate students in the room at most times, she was informed of the need for this observational intrusion and the mechanics thereof. This explanation made it impossible to assess the teacher's behavior without her knowledge, but it was felt that deception about teacher observation could have been harmful both to this project and future projects in the school. Nonetheless, frequent teacher observations by two graduate students who were often in the room the entire week ensured some uniformity of her behavior throughout the week. The graduate students frequently met with Mrs. A. to alert her to any deviations from the experimental instructions, and equally important, to reinforce her "appropriate" behavior. Observations of the teacher's behavior were made on a 20 sec observe, 10 sec record basis. The categories of teacher behavior selected for observation were as follows:

I. Comments *preceding* responses.

A. *Academic instruction* "Now we will do arithmetic", "Put everything in your desk", "Sound out the words."

B. *Social instruction* "I'd like you to say 'please' and 'thank you'", "Let me see a quiet hand", "Let's sit up."

II. Comments *following* responses.

A. *Praise* "Good", "Fine", "You're right", I like the way I have your attention."

B. *Criticism.* "Don't do that", "Be quiet", "Sit in your seat!"

C. *Threats* "If you're not quiet by the time I count three . . . . ", "If you don't get to work you will stay after school", "Do you want to stay in this group?"

The teacher's praise, criticism, and threats to individual children were differentiated from praise, criticism, and threats to the class as a whole. For example, "Johnny, be quiet!" was differentiated from "Class, be quiet!". Thus, eight different classes of teacher behavior were recorded: two classes of comments preceding responses and six classes following responses.

Procedure

The eight phases of the study were as follows: (1) Base Period, (2) Classroom Rules, (3) Educational Structure, (4) Praising Appropriate Behavior and Ignoring Disruptive Behavior, (5) Tokens and Back-up Reinforcement, (6) Praising Appropriate Behavior and Ignoring Disruptive Behavior (Withdrawal), (7) Tokens and Back-up Reinforcement, and (8) Follow up. Three procedures, Educational Structure and both of the Token Reinforcement Phases, were instituted for a 2-hr period during the afternoon. The remainder of the procedures were in effect for the entire day. The eight procedures were in effect for all 21 children. The first four conditions were instituted in the order of hypothesized increasing effectiveness. For example, it was thought that Rules would have less effect on the children's behavior than the use of Praise. In addition, it was thought that the combination of Rules and Praise would have less effect than the Tokens and Back-up Reinforcers.

Base Period. After the initial three-week observer training period, the children were observed on eight days over a six-week Base Period to estimate the frequency of disruptive pupil behavior under usual classroom conditions.[4] The teacher was asked to handle the children in whatever way she felt appropriate. During the Base Period, Mrs. A. instructed all the children in subjects like science and arithmetic or took several students to small reading groups in the back of the room while the rest of the class engaged in independent work at their seats. Neither the particular type of activity nor the duration was the same each day. Stars and various forms of peer pressure were sporadically used as classroom control techniques, but they usually had little effect and were discontinued until experimentally reintroduced during the Follow-up Phase.

Classroom Rules. There were seven observations over a three-week period during the second phase of the study. The following rules or instructions were placed on the blackboard by the teacher: "We sit in our seats, we raise our hands to talk, we do not talk out of turn; we keep our desks clear, we face the front of the room, we will work very hard, we do not talk in the hall, we do not run, and, we do not disturb reading groups." Mrs. A. was asked to review the rules at least once every morning and afternoon, and frequent observations and discussions with Mrs. A. guaranteed that this was done on most occasions. The classroom activities again consisted of reading groups and independent seat work.

---

[4] Ten of the 18 observations during the Base Period were eliminated because movies were shown on those days, and disruptive behavior on those days was significantly less than on days when movies were not shown. Although movies were seldom used after Base Period, the seven subsequent observations when movies occurred were eliminated.

Educational Structure. It has been stated that a great deal of the success in token reinforcement programs may be a function of the highly structured regimen of the program and not a function of reinforcement contingencies. Since the Token Phase of the program was designed to be used during structured activities that the teacher directed, Mrs. A. was asked to reorganize her program into four 30 min sessions in the afternoon in which the whole class participated, *e.g.*, spelling, reading, arithmetic, and science. Thus, the purpose of the Educational Structure Phase was to assess the importance of structure *per se*. Mrs. A. continued to review the rules twice a day during this phase and all succeeding phases. During this phase there were five observations over a two-week period.

Praise and Ignore. In addition to Rules and Educational Structure, Mrs. A. was asked to praise appropriate behavior and to ignore disruptive behavior as much as possible. For example, she was asked to ignore children who did not raise their hands before answering questions and to praise children who raised their hands before speaking. In addition, she was asked to discontinue her use of threats. During this phase there were five observations over a two-week period.

Token I. Classroom Rules, Educational Structure, and Praise and Ignoring remained in effect. The experimenter told the children that they would receive points or ratings four times each afternoon. The points which the children received on these four occasions ranged from 1 to 10, and the children were told that the points would reflect the extent to which they followed the rules placed on the blackboard by Mrs. A. Where possible, these points also reflected the quality of the children's participation in class discussion and the accuracy of their arithmetic or spelling. The children's behavior in the morning did not influence their ratings in the afternoon. If a child was absent, he received no points. The points or tokens were placed in small booklets on each child's desk. The points were exchangeable for back-up reinforcers such as candy, pennants, dolls, comics, barrettes, and toy trucks, ranging in value from 2 to 30 cents. The variety of prizes made it likely that at least one of the items would be a reinforcer for each child. The prizes were on display every afternoon, and the teacher asked each child to select the prize he wished to earn before the rating period started.

During the initial four days, the children were eligible for prizes just after their fourth rating at approximately 2:30. Thereafter, all prizes were distributed at the end of the day. For the first 10 school days the children could receive prizes each day. There were always two levels of prizes. During the first 10 days, a child had to receive at least 25 points to receive a 2 to 5¢ prize (level one prize) or 35 points to receive a 10¢ prize (level two prize). For the next six days, points were accumulated for two days and exchanged at the end of the second day. When children saved their points for two days, a child had to receive 55 points to receive a 10¢ prize or 70 points to receive a 20¢ prize. Then, a six-day period occurred in which points were accumulated for three days and exchanged at the end of the third day. During this period, a child had to receive 85 points to receive a 20¢ prize or 105 points to receive a 30¢ prize. Whenever the prizes were distributed, the children relinquished all their points. During Token I, there were 13 observations over a five week period.

For the first week, the experimenter repeated the instructions to the class at the beginning of each afternoon session. Both the experimenter and Mrs. A. rated the children each day for the first week in order to teach Mrs. A. how to rate the children. The experimenter sat in the back of the room and handed his ratings to Mrs. A. in a surreptitious manner after each rating period. Mrs. A. utilized both ratings in arriving at a final rating which she put in the children's booklets at the end of each lesson period. The method of arriving at a number or rating to be placed in the child's booklet was to be based on the child's improvement in behavior. That is, if a child showed any daily improvement he could receive a rating of approximately 5 to 7 so that he could usually earn at least a small prize. Marked improvement in behavior or repeated displays of relatively good behavior usually warranted ratings from 8 to 10. Ratings from 1 to 5 were given when a child was disruptive and did not evidence any daily improvement. Although such a rating system involves much subjective judgment on the part of the teacher, it is relatively easy to implement, and a subsidiary aim of the study was to assess whether a token system could be implemented by one teacher in a class of average size. After the first week, the teacher administered the Token Program herself, and the experimenter was never present when the children were being observed. If the experimenter had been present during the Token Phases but not during Withdrawal, any effects of the Token Program would have been confounded by the experimenter's presence.

Withdrawal. To demonstrate that the token and back-up reinforcers and not other factors, such as the changes that ordinarily occur during the school year, accounted for the observed reduction in disruptive behavior, the token and back-up reinforcers were withdrawn during this phase. There were seven observations over a five week period. When the prizes and the booklets were removed from the room, Mrs. A. told the children that she still hoped that they would behave as well as they had during the Token Period and emphasized how happy she was with their recent improvement. Rules, Educational Structure, and Praise and Ignoring remained in effect.

Token II. When the tokens and back-up reinforcers were reinstated, the children obtained a prize on the first day if they received 25 to 35 points. For the next four days there was a one-day delay between token and back-up reinforcement, the remainder of the Token Reinstatement Period involved a two-day delay of reinforcement. The prize and point system was identical to that during Token I. During this phase, there were five observations over a two-week period.

Follow-up. The token and back-up reinforcers were again withdrawn in order to see if the appropriate behavior could be maintained under more normal classroom conditions. In addition to the continued use of Praise, Rules, and Educational Structure, it was suggested that Mrs. A initiate the use of a systematic star system. Children could receive from one to three stars for good behavior twice during the morning and once during the afternoon. In addition, the children received extra stars for better behavior during the morning restroom break and for displaying appropriate behavior upon entering the room at 9:15 and 12:30. At times, extra stars were given to the best behaved row of children. The children counted their stars at the end of the day, if they had 10 or more stars, they received a gold star that was placed on a permanent wall chart. If a child received 7 to 9 stars, he received a green star that was placed on the chart. The boys gold stars and the girls gold stars were counted each day, and each member of the group with the greater number of gold stars at the end of the week received a piece of candy. In addition, any child who received an entire week of gold stars received a piece of candy. All children began the day without stars so that, with the exception of the stars placed on the wall chart, everyone entered the program at the same level.

Such a procedure was a form of a token reinforcement program, but there were important procedural differences between the experimental phases designated Token and Follow-up. The back-up reinforcers used during the Token Phases were more expensive than the two pieces of candy a child could earn each week during the Follow-up Phase. In addition, four daily ratings occurred at half-hour intervals in the afternoons during the Token Phases but not during Follow-up. On the other hand, stars, peer pressure, and a very small amount of candy were used in the Follow-up Phase. As mentioned previously, both stars and peer pressure had been used sporadically in the Base Period with little effect. Most importantly, it was felt that the procedures used in the Follow-up Phase could be implemented by any teacher. During this phase there were six observations over a four-week period.

Reliability of Observations

The reliabilities of child observations were calculated according to the following procedure: an agreement was scored if both observers recorded one or more disruptive behaviors within the same 20-sec interval; a disagreement was scored if one observer recorded a disruptive behavior and the other observer recorded none. The reliability of the measure of disruptive behavior was calculated for each child each day by dividing the number of intervals in which there was agreement that one or more disruptive behaviors occurred by the total number of agreements plus disagreements. An agreement was scored if both observers recorded the same behavior within the same 20-sec interval. A disagreement was scored if one observer recorded the behavior and the other did not. The reliability of a particular class of teacher behavior on any one day was calculated by dividing the total number of agreements for that class of behaviors by the total number of agreements plus disagreements for that class of behaviors. Reliabilities were calculated differently for child behaviors and teacher behaviors because different types of dependent measures were utilized for children and the teacher, and it was felt that reliability measures should be reported for the specific dependent measures used.

At least one reliability check was made during the afternoon on every child during the Base Period, and one child had three.[3] The average reliability of the measure of disruptive behavior during the afternoons of the Base Period for each of the seven children ranged from 88 to 100%. The following figures represent the number of reliability checks and the average of those reliability checks after the Base Period through the first Token Period for each child. S1, 6, 86%, S2, 7, 91%, S3, 6, 91%, S4, 6, 93%, S5, 6, 87%, S6, 6, 84%, S7, 6, 97%. Because of the repeated high reliabilities, reliability checks were discontinued when the token and back-up reinforcers were reinstated, i.e., no reliability checks were made during or after the Withdrawal Phase.

Adequate morning reliabilities were not obtained until the Rules Phase of the study. The following figures represent the number of reliability checks and the average of those reliability checks during the Rules Phase. S1, 3, 93%, S2, 4, 68%, S4, 3, 91%, S6, 3, 88%. Morning reliability checks after the Rules Phase were made approximately every three observations (approximately seven occasions) through the first Token Period. Average reliabilities of the four children during the Rules, Educational Structure, Praise and Ignore, and Token 1 Phases ranged from 92 to 99%.

Eleven reliability checks for the various classes of teacher behavior before the Praise and Ignore Phase was introduced yielded average reliabilities as follows: academic instruction, 75%; social instruction, 77%; praise to individuals, 77%; praise to the class, 91%; criticism to individuals, 73%; criticism to the class, 72%; threats to individuals, 83%; and threats to the class, 83%.

RESULTS

Child Behavior

Figures 1 and 2 present morning and afternoon data; some of the variability within conditions can be seen. Figure 3 presents data of individual children as well as an average of seven children across afternoon conditions. An analysis of variance was performed on the percentages of combined disruptive behavior, averaged within the eight afternoon experimental conditions, for the seven subjects (See Fig. 3). The analysis of variance for repeated measures (Winer, 1962, p. 111) indicated differences among the eight experimental conditions ($F = 7.3$, $df = 7, 42$, $p < 0.001$). On the other hand, the percentages of combined disruptive behavior of the four children observed in the morning, averaged within conditions, did not change during Rules, Educational Structure, Praise and Ignore, or Token 1 ($F = 1.0$, $df = 4, 12$). Differences among afternoon conditions were assessed by t-tests. Significant and nonsignificant differences are grouped individually in Table 1.[4]

It should be emphasized that comparisons between Follow-up and Praise and Ignore are more meaningful than comparisons between Follow-up and Base, Rules, or Educational Structure. Praise and Follow-up were similar procedures, both included Rules, Educational Structure, and Praise and Ignore. The Base Period did not include any of these. Furthermore, after Rules and Educational Structure were initiated, Mrs. A. stated that she required more academic work from the children than during Base Period. A statistical analysis of the group data suggests that a token reinforcement program can reduce disruptive behavior and that a token reinforcement program can be replaced with a variant of a token program without an increase in disruptive behavior. However, a more detailed analysis of the data for individual children indicated that the Token Reinforcement Program was more effective for some children than others.

---

[3]Before 10 of the 18 observation days during the Base Period were eliminated because movies were shown on those days, at least three reliability checks had been made during the afternoon on each child.

[4]Two tailed tests.

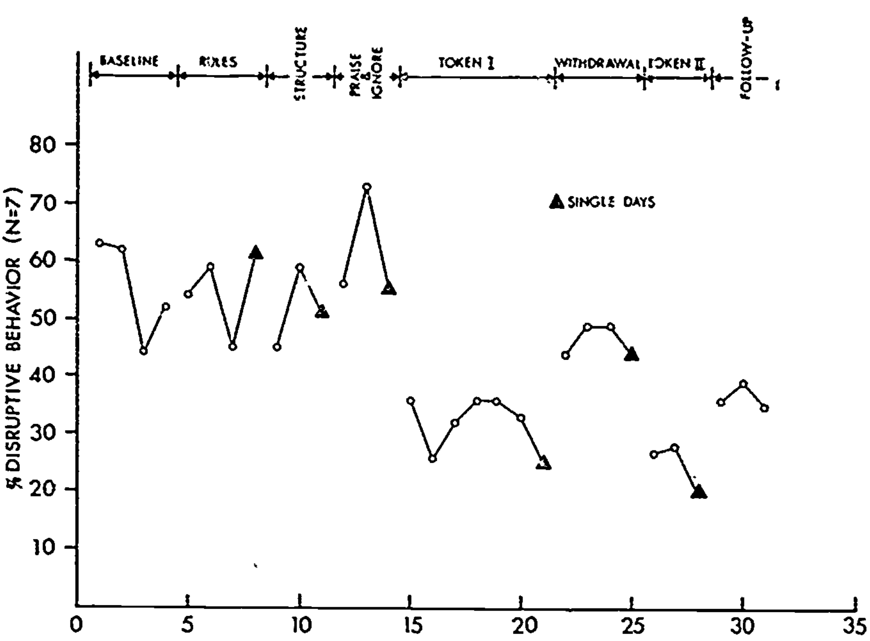


Fig. 1 Average percentage of combined disruptive behavior of seven children during the afternoon over the eight conditions Base, Rules, Educational Structure, Praise and Ignore, Token I, Withdrawal, Token II, Follow-up.

The introduction of Rules, Educational Structure, and Praise and Ignore did not have any consistent effects on behavior (see Fig 3). Praising Appropriate Behavior and Ignoring Disruptive Behavior deserve special mention. Although Mrs. A. used criticism occasionally during the Praise and Ignore Phase, she generally ignored disruptive behavior and used praise frequently. Initially, a number of children responded well to Mrs. A.'s praise, but two boys (S2 and S4) who had been disruptive all year became progressively more unruly during the Praise and Ignore Phase. Other children appeared to observe these boys being disruptive, with little or no aversive consequences, and soon became disruptive themselves. Relay races and hiding under a table contributed to the pandemonium. Several children were so disruptive that the academic pursuits of the rest of the class became impossible. The situation became intolerable, and the Praise and Ignore Phase had to be discontinued much earlier than had been planned.

The disruptive behavior of S7 was reduced to a very low level of 15% by a combination of Rules, Educational Structure, and Praise and Ignore. In the previous token program (O'Leary and Becker, 1967), in which a number of variables including rules, praise, educational structure, and a token program were simultaneously introduced, disruptive behavior during the token period was reduced to a level of 10%. Thus, the present Token Reinforcement Program probably would not be expected further to reduce disruptive behavior in this child.

During Token I, there was a marked reduction (= 18%) in the disruptive behavior of five children (S1, S2, S3, S4, and S6) and a reduction of 3% in S5. Withdrawal of the Token Program increased disruptive behavior from [illegible]% to 45% in these six children. Reinstatement of the Token Program led to a decrease in five of these six children (S1, S2, S3, S4, S5). The disruptive behavior of five children (S1, S2, S4, S5, and S6) ranged from 8% to 39% lower during the Follow-up than dur-

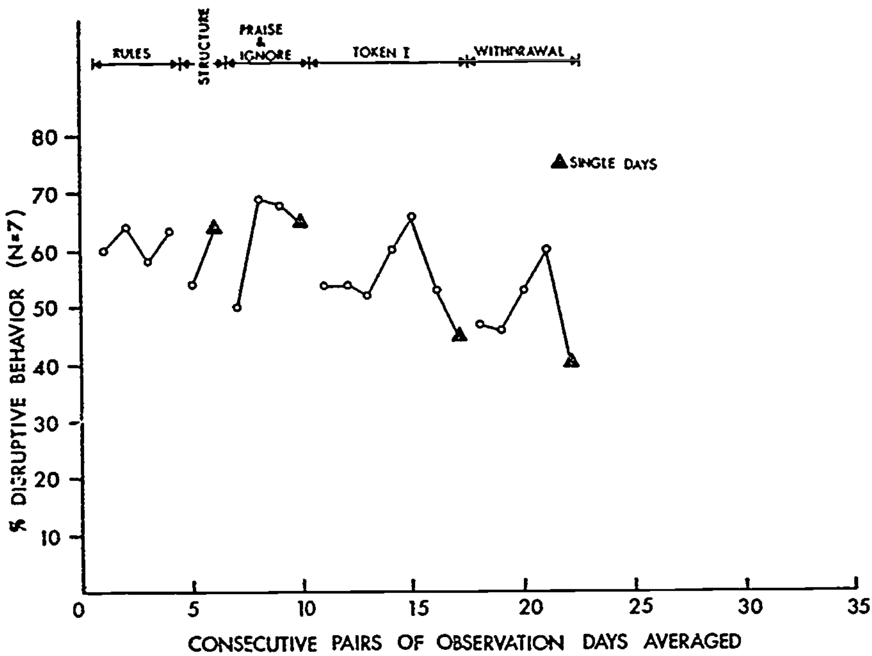


Fig. 2. Average percentage of combined disruptive behavior of four children during the morning over five conditions. Base, Rules, Educational Structure, Praise and Ignore, Token I, Withdrawal, Token II, Follow-up.

ing the Praise and Ignore Phase of the study. Since on no occasion did the Follow-up procedures precede Token I and/or Token II, this study did not demonstrate that Token I and/or Token II were necessary conditions for the success of the Follow-up procedures.

In summary, Token I and Token II were definitely associated with a reduction of disruptive behavior, *and* the Follow-up procedure was effective with three of the six children (S1, S2, and S4) who had more than 15% disruptive behavior during the Praise and Ignore Phase (S7 had 15% disruptive behavior during the Praise and Ignore Phase). Token I and Token II were associated with marked reductions of disruptive behavior of S3, but the frequency of disruptive behavior during the Follow-up was not substantially lower than during the Praise and Ignore Phase. Definitive conclusions concerning the effects of the Token Program cannot be drawn for S5 and S6, although some reduction of disruptive behavior was associated with either Token I and Token II for both of these children. In

Table I

Significant		*Non-Significant*	
Token I *vs.* Withdrawal	$t = 3.3^{**}$	Rules *vs.* Educational Structure	$t = 0.8$
Token II *vs.* Withdrawal	$t = 2.9^{*}$	Educational Structure *vs.* Praise	$t = 1.0$
Token I *vs.* Praise	$t = 3.4^{**}$	Base *vs.* Withdrawal	$t = 1.2$
Token II *vs.* Praise	$t = 3.0^{*}$	Token I *vs.* Follow-up	$t = 1.1$
Base *vs.* Follow-up	$t = 3.2^{**}$	Token II *vs.* Follow-up	$t = 1.5$
Praise *vs.* Follow-up	$t = 3.3^{**}$		
Withdrawal *vs.* Follow-up	$t = 3.2^{**}$		

$^{**}p < 0.02,\ df = 6$
$^{*}p < 0.05,\ df = 6$

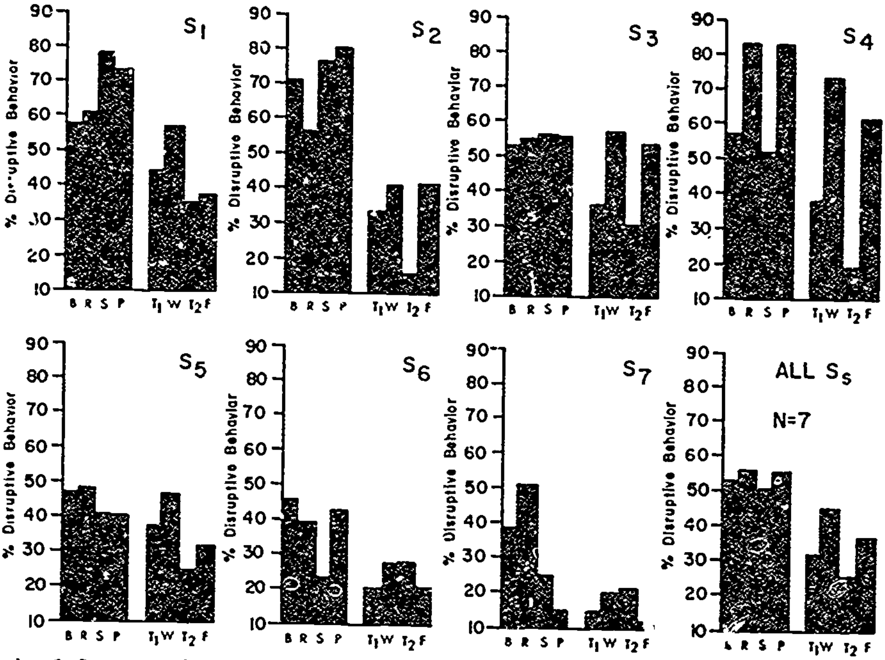


Fig 3 Percentage of combined disruptive behavior for each of seven children during the eight conditions: Base Rules Educational Structure, Praise and Ignore, Token I, Withdrawal, Token II, Follow-up.

addition, the disruptive behavior of S5 and S6 was 8% and 20% less respectively during Follow-up than during the Praise and Ignore Phase.

Teacher Behavior

On any one day, the percentage of each of the eight classes of teacher behavior was calculated by dividing the number of intervals in which a particular class of behavior occurred by the total number of intervals observed on that day. Percentages rather than frequencies were used because of slight variations from the usual 90-min time base.

The percentages of different classes of teacher behavior were averaged within two major conditions: (1) data before Praise and Ignore Phase, and (2) data in the Praise and Ignore and succeeding Phases. The data in Fig 4 show that in the Praise and Ignore Phase, Mrs A increased use of praise to individual children from 12% to 31% and decreased use of criticism to individuals from 22% to 10% Mrs A also increased use of praise to the class from 1% to 7% and decreased criticism directed to the class from 11% to 3%. Because the frequency of threats was quite low, threats to individuals and threats to the class were combined in one measure Using this combined measure, Mrs. A.'s use of threats decreased from 5% to 1%. There were no differences in Mrs. A.'s use of academic or social instruction. Consequently, the changes in the children's disruptive behavior can probably be attributed to contingencies and not to Mrs. A.'s use of cues concerning the desired behaviors.

DISCUSSION

Although a Token Reinforcement Program was a significant variable in reducing disruptive behavior in the present study, the results are less dramatic than those obtained by O'Leary and Becker, (1967). A number of factors probably contributed to the difference in effectiveness of the programs. The average of disruptive behavior during the Base Period in the 1967 study was 76%; in the present study it was 53%. The gradual in-

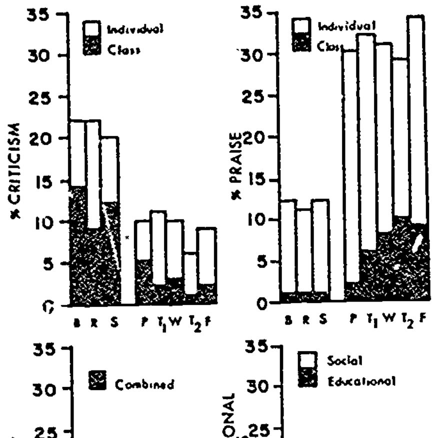


Fig. 4. Percentage of various teacher behaviors to individuals and to the class during the eight conditions: Base, Rules, Educational Structure, Praise and Ignore, Token I, Withdrawal, Token II, Follow-up.

troduction of the various phases of the program was probably less effective than a simultaneous introduction of all the procedures, as in the previous study. In the earlier study, the children received more frequent ratings. Five ratings were made each day at the introduction of the 1.5-hr token program, and they were gradually reduced to three ratings per day. In the present study, the children received four ratings per day during a 2-hr period. In the 1967 study, the class could earn points for popsicles by being quiet while the teacher placed ratings in the children's booklets; in the present study, group points were not incorporated into the general reinforcement program. In the 1967 study, the teacher attended a weekly psychology seminar where teachers discussed various applications of learning principles to classroom management. An *esprit de corps* was generated from that seminar that probably increased the teacher's commitment to change the children's behavior. Although Mrs. A. received graduate credits for her extensive participation in the project, she did not attend a seminar in classroom management. A number of children in the present study had an abundance of toys at home and it was difficult to obtain inexpensive prizes which would serve as reinforcers; in the earlier study, selection of reinforcers was not a difficult problem, since the children were from disadvantaged homes.

Related Gains

Academic. The 14 children for whom there were both pre- and post-measures on the California Achievement Test (including S1, S4, S5, S6, and S7) gained an average of 1.5 yr from October to June. The mean CAT score in October was 1.5 while the mean score in June was 3.0. Although there was no matched control group, such gains are greater than those usually obtained (Tiegs and Clark, 1963). While such gains are promising, conclusions about the effects of a token system on academic performance must await a more systematic analysis.

Attendance. Comparisons of the attendance records of the seven children during the observational days of the token and non-token phases yielded the following results: the average attendance percentage during the 45 observation days of Base, Rules, Educational Structure, Praise and Ignore, and Withdrawal was 86%. The average attendance percentage during the 20 observation days of Token I and Token II was 98%, the average attendance percentage during the 26 observation days of Token I, Token II, and Follow-up (a variant of a token program) was 99%. These attendance records are very encouraging, but because of the usual seasonal variations in attendance and the small sample of children, more definitive evidence is needed before conclusions about the effects of a token program on attendance can be made.

Cost of Program

The cost of the reinforcers in the present study was approximately $125.00. It is estimated that 3 hr of consulting time per week would be essential to operate a token reinforcement program effectively for one class in a public school. The cost of such a program and the amount of consulting time seem relatively small when compared to the hours psychologists spend in therapy with children, often without producing significant behavioral changes (Levitt, 1963). Furthermore, as evidenced in the present study, control of behavior may be shifted from reinforcers, such as toys, to reinforcers existing within the natural educational setting, such as stars and peer prestige.

Generalization

During the morning, the majority of the children were engaged in independent seat work, while four or five children were in a reading group with the teacher in the back of the room. Although there were rules and frequent instructions during the morning, there was little reinforcement for appropriate behavior, since Mrs. A. felt that it would be disruptive to the rest of the class to interrupt reading groups to praise children who were doing independent work at their seats. Ayllon and Azrin (1964) found that instructions without reinforcement had little effect on the behavior of mental patients. Similarly, Rules (instructions) without reinforcement did not influence the behavior of the children in this study.

Mrs. A. was instructed to praise appropriate behavior and ignore disruptive behavior in the morning as well as the afternoon. However, Mrs. A.'s criteria of appropriate behavior in the morning differed from her criteria in the afternoon. For example, in the morning she often answered questions when a child failed to raise his hand before speaking. In the afternoon, on the other hand, she generally ignored a child unless he raised his hand. In order to achieve "generalization" of appropriate behavior in a Token Program such as this one, the teacher's response to disruptive behavior must remain constant throughout the day. The percentage of disruptive behavior was reduced during the morning of the first few days of Token I, but the children presumably learned to discriminate that their appropriate behavior was reinforced only in the afternoon. The differences in the children's behavior between the morning and the afternoon help to stress the point that "generalization" is no magical process, but rather a behavioral change which must be engineered like any other change.

REFERENCES

Allen, K. Eileen, Hart, Betty M., Buell, Joan S., Harris, Florence R., and Wolf, M. M. Effects of social reinforcement on isolate behavior of a nursery school child. *Child Development*, 1964, 35, 511-518.

Ayllon, T., and Azrin, N. H. Reinforcement and instructions with mental patients. *Journal of the Experimental Analysis of Behavior*, 1964, 7, 327-331.

Becker, W. C., Madsen, C. H., Arnold, Carole R., and Thomas, D. R. The contingent use of teacher attention and praise in reducing classroom behavior problems. *Journal of Special Education*, 1967, 1 (3), 287-307.

Birnbrauer, J. S., Wolf, M. M., Kidder, J. D., and Tague, Celia. Classroom behavior of retarded pupils with token reinforcement. *Journal of Experimental Child Psychology*, 1965, 2, 219-235.

Brown, P. and Elliot, R. Control of aggression in a nursery school class. *Journal of Experimental Child Psychology*, 1965, 2, 103-107.

Hall, R. V., Lund, Diane, and Jackson, Deloris. Effects of teacher attention on study behavior. *Journal of Applied Behavior Analysis*, 1968, 1, 1-12.

Harris, Florence R., Johnston, Margaret K., Kelley, C. Susan, and Wolf, M. M. Effects of positive social reinforcement on regressed crawling of a nursery school child. *Journal of Educational Psychology*, 1964, 55, 35-41.

Harris, Florence R., Wolf, M. M., and Baer, D. M. Effects of social reinforcement on child behavior. *Young Children*, 1964, 20, 8-17.

Kuypers, D. S., Becker, W. C., and O'Leary, K. D. How to make a token system fail. *Exceptional Children*, 1968, 35, 101-109.

Levitt, E. E. Psychotherapy with children. A further evaluation. *Behaviour Research and Therapy*, 1963, 1, 45-51.

O'Leary, K. D. and Becker, W. C. Behavior modification of an adjustment class. A token reinforcement program. *Exceptional Children*, 1967, 33, 637-642.

Quay, H. C., Werry, J. S., McQueen, Marjorie, and Sprague, R. L. Remediation of the conduct problem child in a special class setting. *Exceptional Children*, 1966, 32, 509-515.

Scott, Phyllis M., Burton, R. V., and Yarrow, Marian R. Social reinforcement under natural conditions. *Child Development*, 1967, 38, 53-63.

Tiegs, E. W. and Clark, W. W. Manual, California Achievement Tests. Complete Battery. 1963 Norms. California Test Bureau, Monterey, California.

Winer, B. J. *Statistical principles in experimental design*. New York. McGraw-Hill, 1962.

Wolf, M. M., Giles, D. K., and Hall, R. V. Experiments with token reinforcement in a remedial classroom. *Behaviour Research and Therapy*, 1968, 6, 51-64.

Zimmerman, Elaine H. and Zimmerman, J. The alteration of behavior in a special classroom situation. *Journal of the Experimental Analysis of Behavior*, 1962, 5, 59-60.

*Educational Technology: New Myths and Old Realities**

ANTHONY OETTINGER
SEMA MARKS

Many claims have been made about the potential value of modern technological equipment and techniques for the improvement of education. In this article the authors evaluate some of these claims in depth. First, they examine the assertion that technology will promote "individualization of instruction," contrasting broad claims with a narrower reality. They then outline some sources of resistance that will make it difficult to introduce new educational technology into the schools. The authors illustrate their case by referring to two examples of educational technology, the Watertown (Mass.) Language Laboratory and the Stanford-Brentwood C.A.I. Laboratory. They reach the conclusion that the short-range claims made for educational technology are unfounded.

Many people believe that technology will revolutionize education. A Republic Steel advertisement tells us:

> Someday a single computer will give individual instruction to scores of students—in a dozen subjects at the same time. . . . The computer will very probably revolutionize teaching—and learning—within a decade. It is already happening in its early stages.
>
> Computerized instruction can practically (and pleasurably) allow each student to learn

* Adapted by the Editors from the forthcoming book *Run, Computer, Run* by Oettinger and Marks. Both the study for this book and the responses by Suppes and Becker were completed under the auspices of the Harvard Program on Technology and Society.

HARVARD EDUCATIONAL REVIEW, 1968, Vol. 38, pp. 697 717.

more, faster, but always at his own pace. Individualized instruction, the ultimate dream of effective education, is well within the range of possibility.[1]

We are also told that through its capacity to individualize instruction, the computer will reduce discipline problems, eliminate the need for compensatory education, solve the dropout problem, and free the teacher from administrative tasks. And all this is around the corner. Or is it? R. Louis Bright, associate commissioner for the U. S. Office of Education's Bureau of Research, has been quoted as saying that "computers will be ready for massive use in the classrooms of American elementary and secondary schools 'in three to four years.'"[2]

Our conclusions, however, after a study of the impact of technology on education under the Harvard/IBM Program on Technology and Society, are somewhat less enthusiastic.

Consider, for example, the claim cited above, that technology will usher in "individualized instruction, the ultimate dream of effective education." The sustained differences between persistent glorification of individuality as a goal of education and as a desired quality of pedagogical technique, and the equally persistent autocracy and regimentation of the classroom deserves examination. It turns out, as one might have expected, that the meaning of "individualized instruction" is in fact exceedingly fuzzy and of little value as anything but a flag. This, however, is not apparent in much of the literature on the upcoming "technological revolution" where the term is used with great abandon and little definition. Moreover, even if we were to assume a limited notion of individualized instruction, there would still remain very serious problems of implementation.

A full description of these problems would need to touch upon every aspect of the school and its environment. Short of that, however, one may appreciate some of the difficulties of "individualization" by looking at the more general problem of technological innovation in the schools. The reasons for the discrepancy between ultimate promise and immediate possibility become apparent through a comparative examination of the properties of a system receptive to technological innovation and the properties of school systems.

"Individualized Instruction"

Many psychologists now officially agree that there are individual differences in learning capacities. There obviously is also increasing consciousness that con-

[1] *Scientific American*, CCXVI (September, 1967), 13.
[2] "Enormous Role Seen for Computer," *Washington Monitor*, November 27, 1967, p. 78.

temporary education does not serve equally the needs and interests of all groups in our society. Current educational talk is thus all for individualized instruction. But what does "individualization" mean?

A case can be made for defining it as something like personalizing or customizing, namely taking a mass-produced object and stamping it with gold initials or heaping chrome on fins to give the illusion of individual tailoring. This is the sense in which current experimental computer programs greet you with "Good morning, Johnny" by filling in the blank in "Good morning, ———" with the name you had to give to identify yourself to the machine in the first place. This is more genteel than "Do not fold, spindle, or mutilate!," "Hey you!" or "Good to see you, 367 A 15096." It is, however, just as superficial, even when randomly selected variations heighten the effect of spontaneity.

A loftier interpretation postulates that individualizing means giving full scope to idiosyncrasy, to the freedom to pursue whatever subject suits one's fancy in a manner entirely of one's own choosing.

To current practitioners, individualization means much less than pure idiosyncrasy but usually more than golden initials. Precisely what it means is rather in doubt:

> During the past decade, the term "individualizing instruction" has become a watchword with educational reformers. Two recent yearbooks of educational organizations have had this term as title . . Oddly, both volumes were written as though everyone knows what individualization means since neither of them offers a working definition of the term. In point of fact, there is great confusion.[3]

In spite or because of its obscure meaning, individualized instruction is held up as a panacea for the ills of education. Speaking on the IPI (Individually Prescribed Instruction) Project at the Oakleaf School, Bright explains why federal officials are so enthusiastic about it:

> ——-Youngsters of all ability levels would learn more. And they would enjoy school far more, thus reducing discipline problems.
>
> ——- There would be no need for compensatory education for deprived children, on which the Federal government now is spending $1 billion of its $4 billion annual education budget.

[3] Glen Heathers, "Individualized Instruction," in U.S. Congress, Senate, Committee on Labor and Public Welfare, Subcommittee on Education, *Notes and Working Papers Concerning the Administration of Programs Authorized Under Title III of Public Law 89-10, The Elementary and Secondary Education Act of 1965 as Amended by Public Law 89-750* (Washington: U.S. Government Printing Office, 1967). Chapter 7. Section B, p. 178. This report is hereafter cited as *Notes and Working Papers.*

—The dropout problem would largely be licked.

- Teachers would cease being mere dispensers of information and would be free to tutor students individually and encourage youngsters to think and to express themselves.

Parents could take children out of school for vacations any time during the year without disrupting their learning process.[4]

The IPI people themselves describe their work in a much more modest and scholarly fashion:

The project's concern for the individualization of rates of progression should not be taken as a judgment that this represents an attack on the most important aspect of individual differences. It represents a decision to make a rather intensive study of a school program which concentrates on this one aspect. Other aspects such as differences in interests and in other personal qualities may be equally important or even more important, but this project, at least for the present, will concern itself largely with the differentiation of rates.[5]

Lindvall and his co-workers thus clearly recognize that progression at one's own rate is only one facet of any reasonable concept of individualized instruction. In this case, the misrepresentations arise when others equate the part and the whole for public exposure and consumption.

The extent to which a student may deviate from a given course, in either content or style is also limited. The IPI program specifies a sequence of behaviorally defined objectives for each subarea in the subjects involved.

Each objective should tell exactly what a pupil should be able to do to exhibit his mastery of the given content and skill . . .

Objectives should be grouped in meaningful streams of content . . .

Within each stream or area the objectives should, to the extent possible, be sequenced in such an order that each one will build on those that precede it and, in turn, be a prerequisite to those that follow. The goal here is to let the objectives constitute a "scale" of abilities . . .

Individually Prescribed Instruction lesson materials must be geared exactly to the instructional objectives.[6]

[4] "Individual Teaching Plan Excites Experts," *Boston Globe*, October 5, 1967, p. 42.

[5] C. M. Lindvall (ed.), *Defining Educational Objectives*, report of the Regional Commission on Educational Coordination and the Learning Research and Development Center (Pittsburgh, Pa.: University of Pittsburgh Press, 1964), p. 4.

[6] Research for Better Schools, *Individually Prescribed Instruction* (Undated manuscript, Philadelphia, Pa.), pp. 3-4.

If, as is true of all computerized systems of "individualized" instruction now visible as prototypes as well as of many others based on explicit definitions of "behavioral objectives" (BO's), the intent is to instruct students in such a manner that all will achieve a final level of competency which meets (or surpasses) the same set of minimally acceptable performance criteria, with variation only in style, speed, or level of achievement, the objective cannot be the cultivation of idiosyncrasy. It is, rather, what an industrial engineer might call mass production to narrow specifications with rigid quality control. Each pupil is free to go more or less rapidly exactly where he is told to go.

Semantic perversion therefore tends to mask the fact that the techniques now being developed may have great value in training to very narrow and specific "behavioral objectives" but do not address themselves to the many broader but just as basic problems of education. Training to minimal competence in well-defined skills is very important in a variety of military, industrial, and school settings. It is not, however, the whole of what the educational process should be.

Let us assume for the moment that an effective curriculum has been designed which does in fact provide for the types and ranges of learner variability encompassed by the school. What about implementing it?

Allowing students to go at their own rates creates certain problems even when all students are directed toward the same prescribed narrow goals. In a continuous progress, nongraded environment, the children move onward according to their readiness to proceed. Each child spends a different amount of time with each of different instructional resources. The task of matching time, instructors, facilities, and students already requires much effort and expense in present educational institutions. It is still more difficult when scheduling is on a "continuous progress" basis.

Patrick Suppes of the Brentwood School has found that when students are given the opportunity to progress at will, "the rate at which the brightest children advance may be five to ten times faster than that of the slowest children."[7] Although he began with a group of students "very homogeneous in initial measures of ability" (IQ range from 122 to 167, with a mean of 137.5), after a year and a half the spread was "almost two years."[8]

[7] Patrick Suppes, "The Uses of Computers in Education," *Scientific American*, CCXV (September, 1966), 218.

[8] Patrick Suppes, *Accelerated Program on Elementary School Mathematics—The Second Year*, Technical Report no. 86 (Stanford, Calif., Institute for Mathematical Studies in the Social Sciences, 1965), p. 11.

A computer simulation done by the System Development Corporation points out that two schools in different parts of the country which had independently experimented with a continuous progress plan for four years "independently decided to place their slowest students back in a lockstep group plan."[9]

One reason why the lockstep system has persisted for so long may be that it minimizes the teacher's information processing problems. Students progress in unison; they are given identical assignments and they take tests simultaneously. In the continuous progress plan school, on the other hand, at any one time each student may be working at a point in the curriculum different from every other.

Thus, each student's status must be tracked individually. To complicate the matter still further, the records must be kept in "real-time," i.e., *at any time*, not just weekly or by marking periods, the student's status as a learner must be available to the decision maker, be it computer or human guidance counselor. An effective monitoring device and evaluation scheme must be built in, because otherwise it would be impossible to know when to schedule a student for, say, extra help. Extrapolation from the SDC simulation suggests that "with a population of 900 students, there would be from 30 or 40 changes between courses and about 300 mastery tests daily."[10]

Although classroom scheduling by computer is advertised as a fait accompli, this is true only in the rather restricted sense of assigning students to conventional classroom groups and insuring that the number of groups matches the number of available teachers, and that these groups and teachers fit into available classrooms. The whole operation typically takes place once a term. This, however, is a far cry from keeping track of individual students week by week, day by day, hour by hour, or minute by minute, and matching them in turn with resources themselves parceled out in smaller packages than teachers per semester or rooms per semester. Keeping an accurate real-time list of teachers, rooms, and media is a major inventory control problem, of a type at the edge of the state of the computer art.

Thus, we remain at quite some distance from implementing truly individualized instruction in the classroom and begin to become aware of the difficulties of integrating technology into the schools on a large scale.

[9] John F. Cogswell, *Analysis of Instructional Systems* (Santa Monica, Calif., System Development Corp., TM-1493/201/00, 1966), p. 183.
[10] *Ibid.*, p. 42.

Systems and Schools

Computers are not now the great individualizers of instruction and may never even be useful in the schools unless these schools are drastically reorganized and appropriately funded to absorb reform. The discrepancy between ultimate promise and immediate possibility becomes apparent in a comparative examination of the properties of school systems and the properties of other social systems that have proved to be receptive to innovation.

It is difficult to comprehend the boundaries of the so called "school system." One of the most striking features of the educational network is the complexity of multifarious linkages between various elements of society and the school system.

A change in the school hours affects not only pupils and school personnel but every child's mother. You introduce the "new math" and shake up every parent in town. Ability grouping invites federal court decisions prohibiting it. If part of the high school burns down, it may be cheaper for local taxpayers to build a new one because the state contributes toward costs of new construction but not of renovation. An experiment with new curricula raises the specter of low performance on college boards. And, most obviously, the people who make up every other institution from the family to the Presidency are products of the school.

Granted the complexity of the system, it becomes obvious that any change in the schools which alters or even threatens to alter established linkages between the school and any other segment of society will meet at best with the delays inherent in explaining any change to those affected by it and at worst with stony resistance. Whenever one of the segments of society with which schools interact sees it in its interest to press for change in the schools, then the schools, if anxious to accept the change, must in turn still make their peace with other linked segments of society, if the change seems undesirable to the schools themselves but the pressures are strong, the schools are likely to adopt evasive tactics which suggest the form of change without commitment to its substance.

There is, in fact, widespread critical opinion that this indeed is what happens. We read, for instance, that:

Boston—and other cities—like to talk innovation. Innovation has become fashionable and profitable. The federal government will pay for almost anything billed as new or experimental. In the past two years more than two billion dollars have gone to programs associated with education for "disadvantaged youth." Around the urban schools are magnificent necklaces of special programs, head starts, pilot schools, enrichment classes, but the body of education and the results produced remain almost unchanged. In Boston, which has enough trial programs and experiments to fill a book, the life of the average child in the average classroom is virtually unaffected. The teachers, the curriculum, the school committee are the same. The books are the same. The attitudes are the same.[11]

A visit to a place we shall call Small City provided an example of this phenomenon. Overhead projectors were introduced into a brand new high school to meet the emergency created because blackboards had not been delivered before the opening of school. School officials provided enough overhead projectors to have approximately one for every two classrooms. These remained even after the blackboards came, but older teachers who had complained all along that they needed a pilot's license to use these gadgets, promptly abandoned their use. Although we were told that the younger teachers had been enthusiastic about the new devices, we saw no evidence of anyone using them in any classroom.

Not all failures to integrate new hardware into the educational system result from the intransigence of teachers, as the following observation suggests:

The shop seemed well equipped with the standard wood-working hand tools and also with an impressive array of power tools, including three lathes, a planer, a joiner, a sander, and a drill press. Had we been walking through on a more superficial tour, we would have remained deeply impressed, but when the shop teacher caught our glance toward the equipment, he quickly volunteered the information that the tools had been standing there for a year and a half and were unconnected to any power source. He also pointed to a stack of electrical cabling conduits on top of one of the cabinets, which he said represented some $800 worth of electrical equipment necessary to connect his power tools, but had lain in its resting place for an equally impressive number of months. With some bitterness, he attributed this sad state of affairs to the fact that the entire school system has only four electricians who, at Christmas time, damn well had to go and get the mayor's Christmas tree lit up. He also told us that the work benches, which were new, had been on order for ten years. The machine tools, he said, had been ordered a long time ago and appeared at a time when he had given up hope, but this manna was unaccompanied by the juice necessary to wash it down.[12]

Unfortunately there is a rather large step to take between the act of recognizing that because the schools are tied to many apron strings they find it hard to change and the acquisition of sufficient knowledge about the strings, what

[11] Peter Schrag, *Village School Downtown, Politics and Education—A Boston Report* (Boston: Beacon Press, 1967), p. 117.

[12] Anthony G. Oettinger, "Visit to a Small City" (Unpublished memorandum, 1967), pp. 20-21.

they are made of, how they are interconnected, who pulls them, how much they can stretch, etc., to provide first, understanding, and second, rational control and direction toward agreed upon and well defined goals. This particular aspect of the school system does not augur well for the introduction of technology. If we compare the organizational structure of the schools with the structure of organizations which have successfully integrated advanced technology into their functioning, it becomes obvious that the educational system is far more complex than anything we have hitherto tackled – air defense systems, moon shots, and air, rail, or communication networks notwithstanding—and that, in addition, we have far less knowledge of its component parts.

The system which is amenable to the introduction of technology is characterized by (1) enough independence from the other systems related to it for interactions with these systems to be satisfactorily accounted for or ignored, (2) well-developed and reliable research and design tools, and (3) most important, goals which are both well known and specifiable.

School systems, on the other hand, are closely tied to society. Research findings on significant educational issues are fragmented and contradictory, and we have already noted the difficulty of defining as obvious a goal as individualization.

It sometimes appears that existing American political and social systems have built-in defenses against being treated as systems, for purpose of analysis or problem solving. Socially and politically, we have substantial decentralization and localization in our decision making structure (much different than the hierarchical structure in the defense establishment).[13]

There is nothing in education remotely approximating the Secretary of Defense or the Joint Chiefs of Staff. Rather, there are 27,000 local autonomous school districts and corresponding school boards.

The schools are over centralized in the sense that universalistic practices, the standardized curriculum, and the conduct of programs of school community relationships from the central office and board result in obvious difficulties in adapting to the different problems of varying neighborhoods ... In Chicago, the local area superintendent is to "run his own district," but control over budget, staffing, and so on is still left downtown. The schools are also under centralized in the sense that it is difficult for decisions made at the top of the organization to become operationalized lower in the organization.[14]

[13] John S. Gilmore, et al., *Defense Systems Resources in the Civil Sector*, prepared for the U.S. Arms Control and Disarmament Agency (Washington. U.S. Government Printing Office, 1967), p. 65.
[14] David Street, "Public Education and Social Welfare in the Metropolis," Working Paper no. 69 (Chicago. Center for Social Organizational Studies, 1966), p. 22.

It is difficult, for example, to find an appropriate audience in the schools much less to find a source of funds and still less a boss to satisfy:

This problem of identifying and determining how most effectively to address each audience is much more severe in the civil sector than it is in the military environment. The hierarchical and authoritarian structure, as well as the often classified nature of military systems contracts, precludes much public debate and minimizes the number and variety of audiences for the report. Yet, almost all civil systems work must be responsive to many sets of requirements and many audiences—a problem which constitutes much of the so-called communications gap between defense systems people and civil servants.[15]

In contrast to the military environment, where decisions often depend on obscure knowledge and security regulations further narrow the scope of debate on a given proposal, schools are a familiar experience for everyone and consequently nearly everyone has an opinion about them. Clearly, at least the school board must be convinced of the value of going ahead with an innovation; in Boston, for instance, one would have to deal with Mrs. Hicks. Appearing at a meeting in Dorchester "she declared that there will be no redistricting, that the state suggestions were 'made by a computer which didn't take into consideration the emotions of the citizens.' "[16] This kind of barrier is impervious alike to the crew-cut ex-merchant of death and the bearded academic zealot.

Even if the decision to innovate were handed down from the school board, the task of integrating and adopting the new technologies would be left in the hands of the teachers.

If there is one thing the teacher, particularly the female teacher, is not, it is an engineer. Indeed, it is difficult to think of two world views further apart than those symbolized by the Golden Rule on the one hand and the slide rule on the other. The one calls to mind adjectives such as romantic, warm, tender minded, naive, the other calls to mind adjectives such as realistic, cold, tough-minded, efficient. One is essentially feminine, the other masculine. These two lists of adjectives undoubtedly exaggerate the real differences to be found between these two groups, but they do give us pause when we consider the likelihood of increasing the dialogue between the tender minded teachers and the tough-minded technicians. To say that they do not speak the same language is a gross understatement.[17]

[15] Gilmore, *et al., Defense Systems*, pp. 53-54.
[16] Schrag, *Village School Downtown*, p. 126.
[17] Philip W. Jackson, "The Teacher and the Machine: Observations on the Impact of Educational Technology," prepared for the Committee for Economic Development, September, 1966, p. 7.

In addition, "if classroom teachers are only mildly interested in particular techniques it is unlikely that those techniques will enjoy rapid and widespread use."[18] These observations are consistent with our own concerning resistance to the use of overhead projectors cited earlier.

If, indeed, it makes sense to introduce advanced technology into the classroom—and we will assume this purely for the sake of the present argument—then massive education and re-education, by in-service training or during sabbaticals, will be necessary to produce a breed of teachers who can feel at home with technical devices. Yet no viable mechanisms have yet been developed for the retraining of teachers. As Davies says succinctly, "in-service teacher training is the slum of American education."[19] But even if there were people in the schools who could deal with the new technology, it is the children who matter in the last analysis.

Equipment to date does not meet the demands of the school. Although the manufacturers would like to convey the impression that all one need do to install a computer system is "plug it in," implementing the "new technology" is a nontrivial task even in firmly established computer installation. A recent editorial in *Datamation* colorfully describes the results of listening "to the siren song of the salesman":

Those moans, groans, sobs and whimpers you hear from the big corner office down the hall from the beautiful, clean, glass-enclosed shrine they call the machine room are coming not from children but grown men, grizzled veterans of the edp wars, pioneering heroes who are trying to make their large scale third-generation systems work.

The problems and the agonies have to be lived with to be believed . . .

Remember, this is 1967. Third generationsville. The people we're talking about have been through the conversion snarl before. They are experienced and knowledgeable. What in the hell is going on here, anyway?[20]

Reliability is a problem with even the most modest of devices, as anyone will testify who has had a piece of chalk break in his hand or who has cursed a skipping ball point pen. The schools are not accustomed to dealing with anything less reliable or flexible than the blackboard.

The blackboard is literally at the teacher's fingertips. He can write on it, draw on it, immediately erase what he has written, or preserve it for days. He can scrawl key words

[18] *Ibid.*, p. 8.

[19] Don Davies, "Teacher Education," in *Notes and Working Papers*, Chapter 15, Section B, pp. 295-304.

[20] "EDP's Wailing Wall," *Datamation*, XIII (July, 1967), 21.

on it, produce a detailed diagram, or write out a series of essay questions. He can use the board himself, or ask his students to use it. He can place material on it in advance, or use it to capture the fleeting and ephemeral thoughts emerging from a discussion. Given this flexibility, it is no wonder that the chalk smudged sleeve has become the trademark of the teacher.[21]

Perhaps the best example of equipment in common use that must meet the same standards of reliability and cheapness as educational equipment is the telephone. This instrument is so widespread that it must lend itself to operation by practically anyone, under the most varied environmental conditions, and it is frequently subject to vandalism. Just how major are even the most innocent looking engineering improvements necessary to achieve this degree of reliability may be illustrated by the example of the telephone dial, which has been in continuous development for well over twenty-five years.

A new dial has been designed with slightly less noisy gears than the dial now in existence. It is to be somewhat more reliable and longer lived than its predecessor and somewhat less disturbing as a contributor to room noise. This "minor" change is the product of a year and a half spent by three design engineers and an expenditure of about two million dollars for retooling the manufacturing plant! The difference between a laboratory prototype and a reliable instrument amenable to wide distribution under severe conditions of use must therefore be measured in large multiples of such human and dollar costs.

Current experience with teletype terminals for computers suggests that tomorrow's logs may well include entries about chewing gum in the type mechanism. Those with graphic terminals are learning that a deposit of ear wax interferes with the operation of light pens. Since the grass always looks greener on the other side of the fence, we may look with confidence toward the early development by biomolecular genetic engineers of earless and therefore waxless pupil systems.

Our general conclusion is that the observed combination of institutional rigidity with infant technology will preclude really significant progress in the next decade if significant progress is interpreted, in accord with contemporary literature, as widespread and meaningful adoption, integration, and use of technological devices within the schools. In addition, this discussion should suggest the enormous difficulties that will have to be overcome if educational technology is to be introduced in any decade in the twentieth century.

[21] Jackson, "The Teacher and the Machine," pp. 3-4.

Two Examples

To illustrate concretely some of the general arguments made up to this point, we would like to examine two specific instances of the use of technology in education: a language laboratory currently operating in the Watertown, Massachusetts, Public Schools, and the Stanford Brentwood Computer Assisted Instruction Project.

The Language Laboratory

The choice of the language laboratory as an example deserves some explanation. It is a good example of a branch of educational technology whose adoption is sufficiently recent for the sound of drumbeats and promises still to reverberate, while, at the same time, enough language laboratories are now in enough schools for a realistic assessment to be possible.

Since it, too, has been touted as leading to individualization, we should be aware of the possibilities for perversion. The Raytheon brochure advertising their "Random Access Teaching Equipment" states:

> RATE systems are tailored to the individual student's progress, as each position permits the instructor to gauge the progress of all students on an individual basis. Therefore, the entire class is not limited to the learning capacity of the slowest members, thus permitting fast progress through any given area of instruction when possible.

A trip out to Watertown High School, where the system is operative, produced the following set of Language Laboratory Procedures:

Watertown High School Language Laboratory Procedures

1. The equipment in the laboratory is not like ordinary tape recorders. The principles involved are quite different. *Please do not ask unnecessary questions about its operation.**

2. . . . . .

3. No books, pencils, pens, papers, pocketbooks are to be taken to the booths. People assigned to the first row will leave their belongings on the bookcase at the left front of the room under the windows. Those assigned to rows B through E will leave their materials in the bookcase under the blackboard at the opposite side of the room. Take only yourself to the booth.

4. You will stand quietly behind the chair at your booth until the teacher asks you to sit. Then sit in as close to the desk as possible.

5. *No one is an individual in the laboratory.** Do nothing and touch nothing until instructions are given by the teacher. Then listen carefully and follow directions exactly.

* Italics added.

6. If you find anything out of order in your booth, report it to your teacher immediately. This is *very expensive* equipment. You must take pride in it by giving it the best of care.

Good luck to you!

Perhaps the statement that "no one is an individual in the laboratory" is an overstatement. We learned, for example, that the laboratory contains student booths, each equipped with a tape holder. The system provides the option to submit up to five independent inputs to the students. We were told, however, that this capacity is rarely used. There is difficulty, noted by the students as well as by the teachers, in picking materials that are relevant to what goes on in the classroom. The problem of multimedia integration, reflected in such time-worn complaints as that the reading in the textbook has nothing to do with the lectures, will not be automatically ameliorated by the introduction of more elaborate technology. The point is simple. if it is hard to pick one set of tapes, it is still harder to pick five and to monitor the progress of students proceeding on five different tracks.

We are told by Robert Locke, senior vice president of McGraw-Hill that the "language labs have taken over the drill and left the teacher to do the things she's trained to do ... The children, of course, have gained enormously because now they can be exposed to the language as it ought to be spoken, and they can practice their own speech without the embarrassment of doing it in front of the whole class. It's like singing in the shower."[22]

However, the fact remains that one teacher can give each of thirty students at most one and a third to two minutes of attention in a forty to sixty minute period. Consequently, private schools tend to design their language laboratories for fifteen rather than thirty students. The common notion that the introduction of technology necessarily changes education from a labor intensive to a capital intensive activity is therefore not borne out in this case.

Neither has the laboratory eliminated the problem of student motivation. One language coordinator confided to me. "The language laboratory is a magnificant environment for daydreaming. Kids who are used to having blaring transistor radios around them every waking moment have trained themselves to ignore anything coming into their ears, and therefore hear very little of what comes out of the earphones they wear in the language lab."

[22] Robert W. Locke, "Educational Technology," talk prepared for delivery at a Convention of the Chicago Educational Publishers Association, November 16, 1966, p. 8.

The difficulties lying in the path of realizing individualization are not grandiose, but rather are the accumulation of a myriad of minor frustrations and inadequacies which, in the aggregate, create a problem whose complexity lies far beyond that of any system design problem with which modern technology has successfully grappled to date.

Hard design lessons must be learned to produce equipment that will stand up under the kind of use that appears to be normal in schools. The phenomenon is not restricted necessarily to secondary schools or to schools which are in an area where the auditorium must be padlocked to maintain security. One problem, for example, is that students will pull wires out of head sets! While it is standard on many appliances designed for home, industrial, or military use to supply control knobs that are held on a shaft by a spring and therefore pull off with a reasonable amount of force, this a clearcut invitation—if not necessarily to vandalism—then at least to absent-minded, but nonetheless destructive fiddling.

The operating problems are brought home by a casual look at the teacher's log. It shows the following sprinkling of entries:

Not working at all, not audio active, tape cartridge doesn't play, no hook for headset; button missing, could not hear tape, blue ink marks on tape drive, scratches on tape drive; some spots of missing paint on desk; blue ink mark on desk.

Thus the likelihood that Watertown will be affected substantially by such new technology within the next decade is exceedingly small, for the problems of engineering the transition from laboratory to mass production have only barely begun to be faced, the economics of scale necessary to bring costs within reason can be realized only through massive production and standardization, teaching materials to be used with such systems exist only in bits and pieces, and the people capable of forwarding the state of this art are woefully few. Moreover, no one has yet faced the scheduling problems entailed by even this limited amount of individualization. As the evidence of the Systems Development Corporation studies and the Brentwood experiment points out, the rate of spread of even an initially uniform group of students is enormous. As a consequence, just as Watertown has been unable to take advantage even of the restricted amount of flexibility provided by its language laboratory equipment, so it is highly probable that Watertown or any other school in the next decade will be unable to take advantage even of the pacing flexibility of computers without the kind of major administrative revolution that seems conclusively blocked by other factors in so short a time span.

Stanford-Brentwood

And yet we are told on a visit to the Stanford-Brentwood CAI Project in Palo Alto, California that "The Stanford Brentwood Computer Aided Instruction Laboratory is the first to be an integral part of a public school", however, we then learn that the computerized classroom at Brentwood is in a *separate windowless building* on the grounds of the Brentwood School. This building was erected with federal funds and is dedicated entirely to the purposes of the experiment. It includes a set of offices, a conventional machine room where the principal components of the computer are housed, and a classroom with carrels holding the individual computer teaching terminals. There is also, immediately adjacent to the computerized classroom, a conventional classroom with low tables and chairs and a blackboard.

The public explanation of the project is given in the form of a series of questions and answers, such as the following:

Q: How does the laboratory fit in the organization of the school?

A. The school sends children to the laboratory at regular periods each day. The children stay for a half hour to study either reading or mathematics and at the close of the period return to their own classrooms. For the first year, there will be eight half hour periods per day.

Q. Will the classroom teachers go with their children to the CAI laboratory?

A. The laboratory can handle only 16 children at one time, so the classroom teacher ordinarily stays in her room, to work with the children who come to the laboratory during another period. The laboratory is always open to the teachers, and they frequently come to observe.

Being "an integral part of a public school" therefore means having a carefully insulated facility on the same grounds as the school. The laboratory essentially guarantees that the children who are delivered to it will be kept there for a full half hour, come what may. Thus, the remainder of the school is carefully insulated from possible schedule disruption effects. The insulation is provided, among other things, by the conventional classroom adjacent to the computerized classroom. when the machine is down, the children are taken into that room and taught or amused in conventional ways, until the guaranteed time period is up. It is therefore risky to generalize from this experimental set up to arbitrary school environments.

That the situation is artificial in other ways is revealed by the following questions and answers:

Q: Who operates the laboratory?

A: The laboratory is operated by the Institute. There are specially trained teachers and computer technicians on duty at all times during school hours.

Q: Who supervises the children while they are in the laboratory?

A: The Institute's staff includes experienced, certified elementary school teachers whose main job is to supervise the children and help them use the system.

Q: Do the teachers who will be in the laboratory need special training?

A: The laboratory teachers must know the curriculum material thoroughly and must also be trained to operate the CAI system.

Q: How long have you been developing curriculum materials for this project?

A: The Brentwood CAI Project began in June, 1964. However, personnel of the Institute have been involved in developing learning materials for younger children for over 10 years.

Suppes, who surely is aware of the problems raised in these questions and answers, nevertheless states that "the truly revolutionary function of computers in education ... lies in the novel area of computer-assisted instruction." He admits, however, that "This role of the computer is scarcely implemented as yet." I agree with both statements, although with some reservation about the first. He further predicts that "assuming the continuation of the present pace of technological development, it cannot fail to have profound effects in the near future."[23] Without more detailed specification of the nature of these effects, such a statement is indisputable Whether the effects will be significant to education is very much open to question.

The Brentwood experiment itself is wisely limited to what Suppes calls "drill and practice." It cannot be over emphasized that major problems of integration and training arise even with this relatively simple and elementary kind of computer use.

The Brentwood system, as it currently stands, can store only a limited segment of the mathematics and reading program at one time. Hence a new pupil cannot enter a group which has been working on the reading program for a year and begin from the beginning that part of the course would have been removed to provide room for later lessons. Of course, a past lesson can be reintroduced for this student if the operator has been notified in advance. Given the current conditions of computer operating systems, such jugglings are not necessarily trivial. In the future, scaling up to Suppes' "millions of school children" will lead to familiar difficulties.

[23] Suppes, "The Uses of Computers," p. 207.

Similarly, if the achievement spread of students using the reading program at any one time were too great, it would very likely be impossible for anyone to use the arithmetic sequence. Once again individualization now means *at most* proceeding at one's carefully synchronized individual pace through either the currently available reading materials or the currently available arithmetic materials.

Let us now postulate that, in order to achieve economies in the immediate future, a large central installation is provided to serve many schools. If the central equipment operates in the same mode as the Brentwood experiment, then every child in every school can have access to a particular instructional sequence at those, and only those, times when the sequence is scheduled for presentation. This, of course, is precisely where we left the Watertown language laboratory.

It may be desired that a child who has missed a session be able to return to continue where he left off whenever the particular subject matter is next available on the computer. The scheduling problems are obvious, and the whole situation is reminiscent not only of the language laboratory and its embellishments, but of the effort to provide classroom instruction via television from a central studio or a central source of video tapes.

In all fairness, the computer system does have one important advantage over central audio or television systems. With the latter, the progression of a given program is inexorable, by and large, and the student who has missed a piece of a French lesson is out of step with his class for the rest of the session. To the extent that the individual pacing and branching features of the drill and practice systems are successfully realized, this particular limitation need not apply within a single session with a program designed for as homogeneous a group of students as computer memory limitations dictate.

Nevertheless, we must examine the consequences of providing drill and practice in a variety of subjects simultaneously. After all, we have heard earlier a confident prediction that "someday a single computer will give individual instruction to scores of students - in a dozen subjects at the same time." The alternative of trying to present a dozen subjects in sequence, one at a time, leaving no opportunity for repeats, implies precisely the present mode of lockstep scheduling.

Even to provide a computer system with these limited capabilities to a large number of school systems, however, seems entirely unrealistic in the light of the estimated costs involved. Costs estimates for student use of a computer terminal range from $1.40 to $7.00 per hour. The lower estimate seems to be based on a number of unrealistic premises and the latter seems closer to the truth. But even

$1.40 *per student hour* looms large when one realizes that a typical school system like Watertown spends only $1.00 *per year* on books for each student. The bulk of its school budget (slightly over 80 per cent) goes for salaries. Of course, it is the claim of some that technology can replace the human teacher, thereby cutting this expense and perhaps freeing money to pay the computer bill. Even if the technologist convinced the teachers' union that this was a wise course, it is unlikely that he could make do with fewer teachers. Neither the Watertown language laboratory nor the Brentwood Project has resulted in any manpower reduction.

The Longer Range

It is curious that, in an era where the great and undeniable power of science and technology in certain realms is seen by some with ecstasy and by others with horror, we should be stuck through plain, old-fashioned *ignorance*. What holds back progress in education and in other types of social problems is not that the scientific method has failed us, but rather that it has so far revealed very little of systematic value in attaining the goals of education even if one could assume some measure of agreement on their definition.

We are dealing with problems of an order of complexity for which available mathematical and analytical tools leave us quite unprepared. Worse yet, the elementary building blocks arrayed in such monumental complexity are themselves mysteries. Contemporary psychology can tell us essentially nothing about the details of individual learning processes. Contemporary social and political science can tell us essentially nothing about the dynamic processes that come to play in the transition from one form of social organization to another. Static, descriptive accounts abound, but the dynamics of social science is still *in utero*.

Attempting to design an educational or any other social system with our contemporary scientific apparatus is rather like giving Lucretius an imperial grant to make the bomb starting with his atomic theory. Whether we are two thousand years or thirty years ahead of ourselves makes little difference, and we cannot predict with any degree of reliability how much time it will take for physical, biological, and social science to progress far enough.

We should not, therefore, persist in a naive illusion that science has answers which education can exploit if only it will organize itself properly and do the right incantations. Having owned up to this, our choice of actions is relatively simple, at least in outline.

We may, on the one hand, choose to continue playing the game with the trappings but without the substance of science. Consciously taken, this cynical approach has, as indicated throughout the essay, considerable short-term political value. Its longer range implications, however, are admirably spelled out by Brewster Denny:

Perhaps the greatest danger which science poses for democracy as far as the processes of government are concerned is scientism—a cult of faith in the mechanistic superiority of the scientific method, in the trappings of science, in the mystery and the mystique of science. This is profane perversion of science—cheap, anti-intellectual, and denies the very basic creative faith on which science is founded. Perversion it may be, but man's pathetic desire to put decision and responsibility in a mechanism, or a process, or a machine may at last have found in the period of big government and big science methodologies and technologies of decision making which will take man off the hook for good.[24]

If, on the other hand, we are prepared to admit that we are playing in ignorance, this realization need hold no special terror. This, after all, has been the condition of mankind for millenia. However, this admission urges strongly against any form of strong coordinative planning and prescription on *substantive* matters. Based on deep ignorance, such planning has a much higher probability of being disastrous than of leading to correct or useful solutions.

However wasteful in appearance, it seems best to encourage as much diversity as possible, as many different paths, as many different outlooks, as many different experiments, as many different initiatives as we can afford once the demands of education have been balanced against those of other needs of our society. We should, in short, plan for the encouragement of diversity, at least in technique.

It seems vital to encourage greater freedom of choice in a situation which, however diverse in appearance because of the 27,000 school districts, has in fact a dreary monotony. Vesting all educational authority in the federal government makes no sense whatever under the preceding arguments, but letting our schools continue as local monopolies perpetuates on the local level a crime we would not tolerate and do not tolerate nationally.

There may be other alternatives for providing a kind of large-scale evolutionary effect with enough units at stake to create a fair probability that lots of different paths will be taken, and that illuminating controversy will rage. In

[24] Brewster C. Denney, "Science and Democracy. Politics as Usual in the Nuclear Age?" talk prepared for delivery at the inaugural meeting of the Sigma Xi Club, Richland, Wash., October 12, 1967, pp. 14-15.

other words we are groping for a definition of a reasonable setting for educational experimentation, but we think it vital to shy away from prescriptions of either goal or technique.

We should have a chance to fiddle around with both goals and techniques, and we should give people enough freedom and options so they need not feel like unwilling victims of any particular experiment. We should pursue independence, diversity, professional competence, and integrity, we should encourage following through on longitudinal studies and critical comparisons, upping R & D on higher education from present a pitiful 0.2 per cent, and so on.

Critical comments on all the foregoing are cordially invited. As you can see we are quite unable to make any wise, ringing, and optimistic proposal for the future.

COMPUTERIZED INSTRUCTION AND THE LEARNING PROCESS[1]

RICHARD C. ATKINSON

IN recent years there has been a tremendous number of articles and news releases dealing with computer assisted instruction, or as it has been abbreviated, CAI. One might conjecture that this proliferation is an indicant of rapid progress in the field. Unfortunately, I doubt that it is. A few of the reports about CAI are based on substantial experience and research, but the majority are vague speculations and conjectures with little if any data or real experience to back them up. I do not want to denigrate the role of speculation and conjecture in a newly developing area like CAI. However, of late it seems to have produced little more than a repetition of ideas that were exciting in the 1950s but, in the absence of new research, are simply well worn cliches in the late 1960s.

These remarks should not be misinterpreted. Important and significant research on CAI is being carried on in many laboratories around the country, but certainly not as much as one is led to believe by the attendant publicity. The problem for someone trying to evaluate developments in the field is to distinguish between those reports that are based on fact and those that are disguised forms of science fiction. In my paper, I shall try to stay very close to data and actual experience My claims will be less grand than many that have been made for CAI, but they will be based on a substantial research effort.

In 1964 Patrick Suppes and I initiated a project under a grant from the Office of Education to develop and implement a CAI program in initial reading and mathematics. Because of our particular research interests, Suppes has taken responsibility for the mathematics curriculum and I have been responsible for the initial reading program. At the beginning of the project, two major hurdles had to be overcome. There was no lesson material in either mathematics or reading suitable for CAI, and an integrated CAI system had not yet been designed and produced by a single manufacturer. The development of the curricula and the development of the system have been carried out as a parallel effort over the last 3 years with each having a decided influence on the other.

Today I would like to report on the progress of the reading program with particular reference to the past school year when for the first time a sizable group of students received a major portion of their daily reading instruction under computer control. The first year's operation must be considered essentially as an extended debugging of both the computer system and the curriculum materials. Nevertheless, some interesting comments can be made on the basis of this experience regarding both the feasibility of CAI and the impact of such instruction on the overall learning process.

Before describing the Stanford Project, a few general remarks may help place it in perspective. Three levels of CAI can be defined. Discrimination between levels is based not on hardware considerations, but principally on the complexity and sophistication of the student-system interaction. An advanced student-system interaction may be achieved with a simple teletype terminal, and the most primitive interaction may require some highly sophisticated computer programming and elaborate student terminal devices.

At the simplest interactional level are those systems that present a fixed, linear sequence of problems. Student errors may be corrected in a variety of ways, but no real-time decisions are made for modifying the flow of instructional material as a function of the student's response history. Such systems have been termed "drill-and-practice" systems and at Stanford University are exemplified by a series of fourth-, fifth-, and sixth-grade programs in arithmetic and language arts that are designed to supplement classroom instruction. These particular programs are being used in several different areas of California and also in Kentucky and Mississippi, all under control of one central

[1] Invited address presented at the meeting of the Division of Educational Psychology, American Psychological Association, Washington, D C., September 1967.

AMERICAN PSYCHOLOGIST, 1968, Vol 23, pp. 225 239

computer located at Stanford University. Currently as many as 2,000 students are being run per day; it requires little imagination to see how such a system could be extended to cover the entire country. Unfortunately, I do not have time to discuss these drill-and-practice programs in this paper, but there are several recent reports describing the research (Fishman, Keller, & Atkinson, 1968; Suppes, 1966; Suppes, Jerman, & Green, 1966).

At the other extreme of our scale characterizing student-system interactions are "dialogue" programs. Such programs are under investigation at several universities and industrial concerns, but to date progress has been extremely limited. The goal of the dialogue approach is to provide the richest possible student-system interaction where the student is free to construct natural-language responses, ask questions in an unrestricted mode, and in general exercise almost complete control over the sequence of learning events.

"Tutorial" programs lie between the above extremes of student-system interaction. Tutorial programs have the capability for real-time decision making and instructional branching contingent on a single response or on some subset of the student's response history. Such programs allow students to follow separate and diverse paths through the curriculum based on their particular performance records. The probability is high in a tutorial program that no two students will encounter exactly the same sequence of lesson materials. However, student responses are greatly restricted since they must be chosen from a prescribed set of responses, or constructed in such a manner that a relatively simple text analysis will be sufficient for their evaluation. The CAI Reading Program is tutorial in nature, and it is this level of student-interaction that will be discussed today.

THE STANFORD CAI SYSTEM

The Stanford Tutorial System was developed under a contract between the University and the IBM Corporation. Subsequent developments by IBM of the basic system have led to what has been designated the IBM 1500 Instructional System which should soon be commercially available. The basic system consists of a central process computer with accompanying disc storage units, proctor stations, and an interphase to 16 student terminals. The central process computer acts as an intermediary between each student and his particular course material which is stored in one of the disc storage units. A student terminal consists of a picture projector, a cathode ray tube (CRT), a light pen, a modified typewriter keyboard, and an audio system which can play prerecorded messages (see Figure 1).

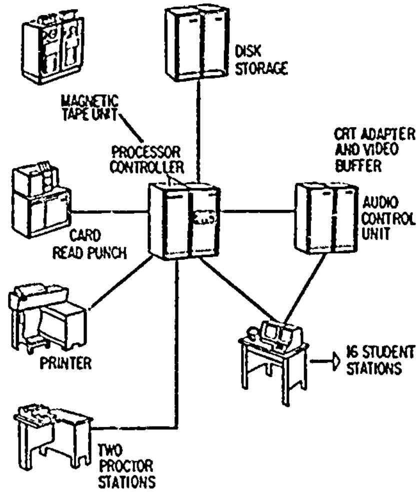


FIG. 1. System configuration for Stanford CAI System.

The CRT is essentially a television screen on which alpha numeric characters and a limited set of graphics (i.e., simple line drawings) can be generated under computer control. The film projector is a rear view projection device which permits us to display still pictures in black and white or color. Each film strip is stored in a self-threading cartridge and contains over 1,000 images which may be accessed very quickly under computer control. The student receives audio messages via a high-speed device capable of selecting any number of messages varying in length from a few seconds to over 15 minutes. The audio messages are stored in tape cartridges which contain approximately 2 hours of messages and, like the film cartridge, may be changed very quickly. To gain the student's attention, an arrow can be placed at any point on the CRT and moved in synchronization with an audio message to emphasize given words or phrases, much like the "bouncing ball" in a singing cartoon.

The major response device used in the reading program is the light pen, which is simply a light-sensitive probe. When the light pen is placed on the CRT, coordinates of the position touched are sensed as a response and recorded by the computer. Responses may also be entered into the system through the typewriter keyboard. However, only limited use has been made of this response mode in the reading program. This is not to minimize the value of keyboard responses, but rather to admit that we have not as yet addressed ourselves to the problem of teaching first grade children to handle a typewriter keyboard.

The CAI System controls the flow of information and the input of student responses according to the instructional logic built into the curriculum materials. The sequence of events is roughly as follows. The computer assembles the necessary commands for a given instructional sequence from a disc storage unit. The commands involve directions to the terminal device to display a given sequence of symbols on the CRT, to present a particular image on the film projector, and to play a specific audio message. After the appropriate visual and auditory materials have been presented, a "ready" signal indicates to the student that a response is expected. Once a response has been entered, it is evaluated and, on the basis of this evaluation and the student's past history, the computer makes a decision as to what materials will subsequently be presented. The time-sharing nature of the system allows us to handle 16 students simultaneously and to cycle through these evaluative steps so rapidly that from a student's viewpoint it appears that he is getting immediate attention from the computer whenever he inputs a response.

The CAI Reading Curriculum

The flexibility offered by this computer system is of value only if the curriculum materials make sense both in terms of the logical organization of the subject matter and the psychology of the learning processes involved. Time does not permit a detailed discussion of the rationale behind the curriculum that we have developed. Let me simply say that our approach to initial reading can be characterized as applied psycholinguistics. Hypotheses about the reading process and the nature of learning to read have been formulated on the basis of linguistic information, observations of language use, and an analysis of the function of the written code. These hypotheses have been tested in a series of pilot studies structured to simulate actual teaching situations. On the basis of these experimental findings, the hypotheses have been modified, retested, and ultimately incorporated into the curriculum as principles dictating the format and flow of the instructional sequence. Of course, this statement is somewhat of an idealization, since very little curriculum material can be said to have been the perfect end product of rigorous empirical evaluation. We would claim, however, that the fundamental tenets of the Stanford reading program have been formulated and modified on the basis of considerable empirical evidence. It seems probable that these will be further modified as more data accumulate.

The introduction of new words from one level of the curriculum to the next is dictated by a number of principles (Rodgers, 1967). These principles are specified in terms of a basic unit that we have called the vocalic center group (VCG). The VCG in English is defined as a vowel nucleus with zero to three preceding and zero to four following consonants. The sequencing of new vocabulary is determined by the length of the VCG units, and the regularity of the orthographic and phonological correspondences. Typical of the principles are the following:

1. VCG sets containing single consonant elements are introduced before those containing consonant clusters (*tap* and *rap* before *trap*).
2. VCG sets containing initial consonant clusters are introduced before those containing final consonant clusters (*stop* before *post*).
3. VCG sets containing check (short) vowels are introduced before those containing letter name (long) vowels (*met* and *mat* before *meat* or mate).
4. Single VCG sequences are introduced before multiple VCG sequences (*mat* before *matter*, *stut* before *stutter*).

More detailed rules are required to determine the order for introducing specific vowels and consonants within a VCG pattern, and for introducing specific VCG patterns in polysyllabic words. These rules frequently represented a compromise between linguistic factors, pattern productivity, item frequency, and textual "usefulness," in that order of significance.

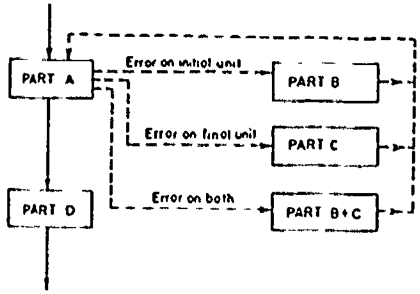


FIG. 2 Flow chart for the construction of a cell in the matrix construction task

The instructional materials are divided into eight levels each composed of about 32 lessons.[2] The lessons are designed so that the average student will complete one in approximately 30 minutes, but this can vary greatly with the fast student finishing much sooner and the slow student sometimes taking 2 hours or more if he hits most of the remedial material. Within a lesson, the various instructional tasks can be divided into three broad areas: (*a*) decoding skills, (*b*) comprehension skills, (*c*) games and other motivational devices. Decoding skills involve such tasks as letter and letter-string identification, word list learning, phonic drills, and related types of activities. Comprehension involves such tasks as having the computer read to the child or having the child himself read sentences, paragraphs, or complete stories about which he is then asked a series of questions. The questions deal with the direct recall of facts, generalizations about main ideas in the story, and inferential questions which require the child to relate information presented in the story to his own experience. Finally, many different types of games are sequenced into the lessons primarily to encourage continued attention to the materials. The games are similar to those played in the classroom and are structured to evaluate the developing reading skills of the child.

Matrix construction. To illustrate the instructional materials focusing on decoding skills let me describe a task that we have called matrix "construction." This task provides practice in learning to associate orthographically similar sequences with appropriate rhyme and alliteration patterns. Rhyming patterns are presented in the columns of the matrix, and alliteration patterns are presented in the rows of the matrix as indicated in Figure 4.

The matrix is constructed one cell at a time. The initial consonant of a CVC word is termed the initial unit, and the vowel and the final consonant are termed the final unit. The intersection of an initial unit row and a final unit column determines the entry in any cell.

The problem format for the construction of each cell is divided into four parts: Parts A and D are standard instructional sections and Parts B and C are remedial sections. The flow diagram in Figure 2 indicates that remedial Parts B and C are branches from Part A and may be presented independently or in combination.

To see how this goes, let us consider the example illustrated in Figure 3. The student first sees on the CRT the empty cell with its associated initial and final units and an array of response choices. He hears the audio message indicated by response request 1 (RR 1) in Part A of Figure 3. If the student makes the correct response (CA) (i.e., touches *ran* with his light pen), he proceeds to Part D where he sees the word written in the cell and receives one additional practice trial.

In the initial presentation in Part A, the array of multiple-choice responses is designed to identify three possible types of errors:

1 The initial unit is correct, but the final unit is not.

2 The final unit is correct, but the initial unit is not.

3 Neither the initial unit nor the final unit is correctly identified.

If, in Part A, the student responds with *fan* he is branched to remedial Part B where attention is focused on the initial unit of the cell. If a correct response is made in Part B, the student is returned to Part A for a second attempt. If an incorrect response (WA) is made in Part B, an arrow is displayed on the CRT to indicate the correct response, which the student is then asked to touch.

If, in Part A, the student responds with *rat*, he is branched to remedial Part C where additional instruction is given on the final unit of the cell.

[2] For a detailed account of the curriculum materials see Wilson and Atkinson (1967) and Rodgers (1967). See also Atkinson and Hansen (1966) and Hansen and Rodgers (1965).

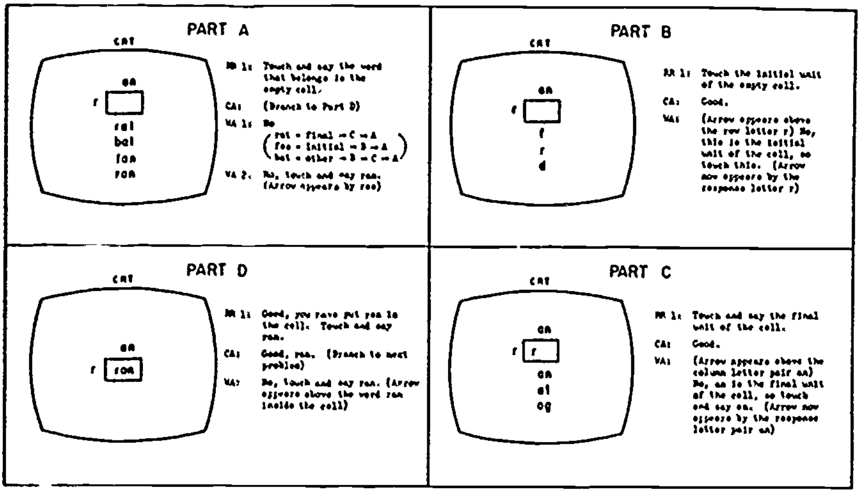


FIG 3 First cell of the matrix construction task.

The procedure in Part C is similar to Part B. However, it should be noted that in the remedial instruction the initial letter is never pronounced (Part B), whereas the final unit is always pronounced (Part C). If, in Part A, the student responds with *bat*, then he has made an error on both the initial and final unit and is branched through both Part B and Part C.

When the student returns to Part A after completing a remedial section, a correct response will advance him to Part D as indicated. If a wrong answer response is made on the second pass, an arrow is placed beside the correct response area and held there until a correct response is made. If the next response is still an error, a message is sent to the proctor and the sequence is repeated from the beginning.

When a student has made a correct response on Parts A and D, he is advanced to the next word cell of the matrix which has a problem format and sequence identical to that just described. The individual cell building is continued block by block until the matrix is complete. The upper left-hand panel of Figure 4 indicates the CRT display for adding the next cell in our example. The order in which row and column cells are added is essentially random.

When the matrix is complete, the entries are reordered and a criterion test is given over all cell entries. The test involves displaying the full matrix with complete cell entries as indicated in the lower left hand panel of Figure 4. Randomized requests are made to the student to identify cell entries. Since the first pass through the full matrix is viewed as a criterion test, no reinforcement is given. Errors are categorized as initial, final, and other, if the percentage of total errors on the criterion test exceeds a predetermined value, then remedial exercises are provided of the type shown in the two right hand panels of Figure 4. If all the errors are recorded in one category (initial or final), only the remedial material appropriate to that category is presented. If the errors are distributed over both categories, then both types of remedial material are presented. After working through one or both of the remedial sections, the student is branched back for a second pass through the criterion matrix. The second pass is a teaching trial as opposed to the initial test cycle, the student proceeds with the standard correction and optimization routines.

An analysis of performance on the matrix task is still incomplete, but some preliminary results are available. On the initial pass (Part A) our students were correct about 45% of the time; however, when an error did occur, 21% of the time it in-

BEST COPY AVAILABLE

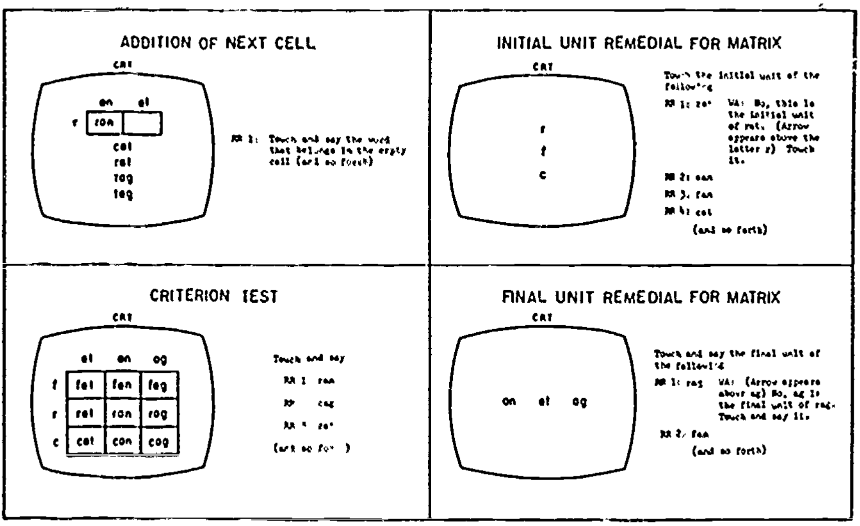


FIG. 4. Continuation of matrix construction task.

volved only the final unit, 53% of the time only the initial unit, and 26% of the time both initial and final units The pattern of performances changed markedly on the first pass through the criterion test. Here the subject was correct about 65% of the time; when an error occurred, 32% of the time it involved only the final unit, 33% of the time only the initial unit, and 35% of the time both units. Thus performance showed a significant improvement from Part A to the criterion test, equally important, initial errors were more than twice as frequent as final errors in Part A, but were virtually equal on the criterion test.

The matrix exercise is a good example of the material used in the curriculum to teaching decoding skills. We now consider two examples ("form class" and "inquiries") of tasks that are designed to teach comprehension skills.

Form class. Comprehension of a sentence involves an understanding of English syntax. One behavioral manifestation of a child's syntactic sophistication is his ability to group words into appropriate form classes. This task provides lesson materials that teach the form class characteristics of the words just presented in the matrix section of a lesson. The following type of problem is presented to the student (the material in the box is displayed on the CRT and below are audio messages, the child answers by appropriately placing his light pen on the CRT):

Dan saw the	tan fat man run	hat.

Only one of the words in the column will make sense in the sentence Touch and say the word that belongs in the sentence.

CA. Yes, Dan saw the tan hat. Do the next one.

WA. No, tan is the word that makes sense. Dan saw the tan hat Touch and say tan. (An arrow then appears above tan.)

The sentence is composed of words that are in the reading vocabulary of the student (i.e., they have been presented in previous or current lessons). The response set includes a word which is of the correct form class but is semantically inappropriate, two words that are of the wrong form class, and the correct word. A controlled variety of sentence types is employed, and the answer sets are distributed over all syntactic slots within each sentence type. Responses are categorized in rather broad terms as *nouns*, *verbs*, *modifiers*, and *other*.

The response data can be examined for systematic errors over a large number of items. Examples of the kinds of questions that can be asked are. (*a*) Are errors for various form classes in various sentence positions similarly distributed? (*b*) How are response latencies affected by the syntactic and serial position of the response set within the sentence? Answers to these and other questions should provide information that will permit more systematic study of the relationship of sentence structure to reading instruction.

Inquiries. Individual words in sentences may constitute unique and conversationally correct answers to questions. These questions take the interrogative Who ? What ? How ?" etc. The ability to select the word in a sentence that uniquely answers one of these questions demonstrates one form of reading comprehension. The inquiry exercises constitute an assessment of this reading comprehension ability. In the following example, the sentence "John hit the ball" is displayed on the CRT accompanied by these audio messages:

Touch and say the word that answers the question.

RR 1 Who hit the ball?
CA Yes, the word "John" tells us who hit the ball
WA: No, John tells us who hit the ball Touch and say John (An arrow then appears on the CRT above John.)
RR 2 What did John hit?
CA. Yes, the word "ball" tells us what John hit.
WA No, ball tells us what John hit Touch and say ball. (An arrow then appears above ball)

As in the form-class section, each sentence is composed of words from the student's reading vocabulary. A wide variety of sentence structures is utilized, beginning with simple subject-verb-object sentences and progressing to structures of increasing complexity Data from this task bear on several hypotheses about comprehension. If comprehension is equated with a correct response to an inquiry question, then the following statements are verified by our data. (*a*) Items for which the correct answer is in the medial position of the sentence are more difficult to comprehend than items in the initial or final positions; final position items are easier to comprehend than items in the initial position. (*b*) Items for which the correct answer is an adjective are more difficult to comprehend than items in which the correct answer is a noun or verb; similarly nouns are more difficult than verbs. (*c*) Longer sentences, measured by word length, are more difficult to comprehend than shorter sentences.

These are only a few examples of the types of tasks used in the reading curriculum, but they indicate the nature of the student-system interaction. What is not illustrated by these examples is the potential for long-term optimization policies based on an extended response history from the subject. We shall return to this topic later.

Problems in Implementing the Curriculum

Before turning to the data from last year's run, let me consider briefly the problem of translating the curriculum materials into a language that can be understood by the computer. The particular computer language we use is called Coursewriter II, a language which was developed by IBM in close collaboration with Stanford. A coded lesson is a series of Coursewriter II commands which causes the computer to display and manipulate text on the CRT, position and display film in the projector, position and play audio messages, accept and evaluate keyboard and lightpen responses, update the performance record of each student, and implement the branching logic of the lesson flow by means of manipulating and referencing a set of switches and counters. A typical lesson in the reading program, which takes the average student

TABLE 1

Audio Script and Film Chips with Hypothetical Addresses

Address	Message
	Audio information
A01	Touch and say the word that goes with the picture.
A02	Good. Bag Do the next one.
A03	No.
A04	The word that goes with the picture is bag Touch and say bag.
A05	Good. Card. Do the next one.
A06	No.
A07	The word that goes with the picture is card. Touch and say card.
	Film strip
I01	Picture of a bag.
I02	Picture of a card.

TABLE 2

COMPUTER COMMANDS REQUIRED TO PRESENT TWO EXAMPLES OF THE PROBLEM DESCRIBED IN THE TEXT

Commands	Explanation
PR	Problem: Prepares machine for beginning of new problem.
LD 0/S1	Load: Loads zero into the error switch (S1). The role of switches and counters will be explained later.
FP F01	Film Position: Displays frame F01 (picture of a bag).
DT 5,18/bat/	Display Text: Displays "bat" on line 5 starting in column 18 on the CRT.
DT 7,18/bag/	Displays "bag" on line 7 starting in column 18 on the CRT.
DT 9,18/rat/	Displays "rat" on line 9 starting in column 18 on the CRT
AUP A01	Audio Play: Plays audio message A01. "Touch and say the word that goes with the picture."
L1 EP 30/ABCD1	Enter and Process: Activates the light-pen; specifies the time limit (30 sec.) and the problem identifier (ABCD1) that will be placed in the data record along with all responses to this problem. If a response is made within the time limit the computer skips from this command down to the CA (correct answer comparison) command. If no response is made within the time limit, the commands immediately following the EP command are executed.
AD 1/C4	Add: Adds one to the overtime counter (C4).
LD 1/S1	Loads one into the error switch (S1).
AUP A04	Plays message A04. "The word that goes with the picture is bag. Touch and say bag."
DT 7,16/→	Displays arrow on line 7, column 16 (arrow pointing at "bag").
BR L1	Branch: Branches to command labeled L1. The computer will now do that command and continue from that point.
CA 1,7,3,18/C1	Correct Answer. Compares student's response with an area one line high starting on line 7 and three columns wide starting in column 18 of the CRT. If his response falls within this area, it will be recorded in the data with the answer identifier C1. When a correct answer has been made, the commands from here down to WA (wrong answer comparison) are executed. Then the program jumps ahead to the next PR. If the response does not fall in the correct area, the machine skips from this command down to the WA command.
BR L2/S1/1	Branches to command labeled L2 if the error switch (S1) is equal to one.
AD 1/C1	Adds one to the initial correct answer counter (C1).
L2 AUP A02	Plays audio message A02. "Good. Bag. Do the next one."
WA 1,5,3,18/W1 WA 1,9,3,18/W2	Wrong Answer. These two commands compare the student response with the areas of the two wrong answers, that is, the area one line high starting on line 5 and three columns wide starting in column 18, and the area one line high starting on line 9 and three columns wide starting in column 18. If the response falls within one of these two areas, it will be recorded with the appropriate identifier (W1 or W2). When a defined wrong answer has been made, the commands from here down to UN (undefined answer) are executed. Then the computer goes back to the EP for this problem. If the response does not fall in one of the defined wrong answer areas, the machine skips from this command down to the UN command.

Table 2—Continued

Commands	Explanation
AD 1/C2	Adds one to the defined wrong answer counter (C2).
L3 LD 1/S1	Loads one into the error switch (S1).
AUP A03	Plays message A03. "No."
AUP A04	Plays message A04. "The word that goes with the picture is bag. Touch and say bag."
DT 7,16/-√	Displays arrow on line 7, column 16.
UN	Undefined Wrong Answer: If machine reaches this point in the program, the student has made neither a correct nor a defined wrong answer.
AD 1/C3	Adds one to the undefined answer counter (C3).
BR L3	Branches to command labeled L3. (The same thing should be done for both UN and WA answers. This branch saves repeating the commands from L3 down to UN.)
PR	Prepares the machine for next problem.
LD 0/S1 FP F02 DT 5,18/card/ DT 7,18/cart/ DT 9,18/hard/	These commands prepare the display for the 2nd problem. Notice the new film position and new words displayed. The student was told to "do the next one" when he finished the last problem so he needs no audio message to begin this.
L4 EP 30/ABCD2	Light-pen is activated.
AD 1/C4 LD 1/S1 AUP A07 DT 5,16/-√ BR L4	These commands are done only if no response is made in the time limit of 30 seconds. Otherwise the machine skips to the CA command.
CA 1,5,4,18/C2	Compares response with correct answer area.
BR L5/S1/1 AD 1/C1 L5 AUP A05	Adds one to the initial correct answer counter unless the error switch (S1) shows that an error has been made for this problem. The student is told he is correct and goes on to the next problem. These commands are executed only if a correct answer has been made.
WA 1,7,4,18/W3 WA 1,9,4,18/W4	Compare response with defined wrong answer.
AD 1/C2 L6 LD 1/S1 AUP A06 AUP A07 DT 5,16/-√	Adds one to the defined wrong answer area and the error switch (S1) is loaded with one to show that an error has been made on this problem. The student is told he is wrong and shown the correct answer and asked to touch it. These commands are executed only if a defined wrong answer has been made.
UN	An undefined response has been made if the machine reaches this command.
AD 1/C3 BR L6	Adds one to the undefined answer counter and we branch up to give the same audio, etc. as is given for the defined wrong answer.

about 30 minutes to complete, requires in excess of 9,000 coursewriter commands for its execution.

A simple example will give you some feeling for the coding problem. The example is from a task designed to teach both letter discrimination and the meaning of words. A picture illustrating the word being taught is presented on the projector screen. Three words, including the word illustrated, are presented on the CRT. A message is played on the audio asking the child to touch the word on the CRT that matches the picture on the film projector. The student can then make his response using the light pen. If he makes no response within the specified time limit of 30 seconds, he is told the correct answer, an arrow points to it, and he is asked to touch it. If he makes a response within the time limit, the point that he touches is compared by the computer with the correct answer area. If he places the light pen within the correct area, he is told that he was correct and goes on to the next problem. If the response was not in the correct area, it is compared with the area defined as a wrong answer. If his response is within this area, he is told that it is wrong, given the correct answer, and asked to touch it. If his initial response was neither in the anticipated wrong-answer area nor in the correct-answer area, then the student has made an undefined answer. He is given the same message that he would have heard had he touched a defined wrong answer, however, the response is recorded on the data record as undefined. The student tries again until he makes the correct response, he then goes on to the next problem.

To prepare an instructional sequence of this sort, the programmer must write a detailed list of commands for the computer. He must also record on an audio tape all the messages the student might hear during the lesson in approximately the order in which they will occur. Each audio message has an address on the tape and will be called for and played when appropriate. Similarly a film strip is prepared with one frame for each picture required in the lesson. Each frame has an address and can be called for in any order.

Table 1 shows the audio messages and film pictures required for two sample problems along with the hypothetical addresses on the audio tape and film strip. Listed in Table 2 are the computer commands required to present two examples of the problems described above, analyze the student's responses, and record his data record. The left column in the table lists the actual computer commands, and the right column provides an explanation of each command.

While a student is on the system, he may complete as many as 5 to 10 problems of this type per minute. Obviously, if all of the instructional material has to be coded in this detail the task would be virtually impossible. Fortunately, there are ways of simplifying coding procedure if parts of the instructional materials are alike in format and differ only in certain specified ways. For example, the two problems presented in Table 2 differ only in (*a*) the film display, (*b*) the words on the CRT, (*c*) the problem identifier, (*d*) the three audio addresses, (*e*) the row display of the arrow, (*f*) the correct answer area, and (*g*) the correct answer identifier. This string of code can be defined once, given a two letter name, and used later by giving a one line macro command.

The use of macros cuts down greatly the effort required to present many different but basically similar problems. For example, the two problems presented in Table 2 can be rewritten in macro format using only two lines of code. Problem 1. CM PW]F01]bat]bag]rat]A01]ABCD1]A04]A02]A03]7]1,7,3,18]C1], Problem 2. CM PW]F02]card]cart]hard]]ABCD2]A07]A05]A06]5]1,5,',18]C2]. The command to call a macro is CM, and PW is an arbitrary two character code for the macro involving a picture-to-word match. Notice that in Problem 2 there is no introductory audio message, the "]]" indicates that this parameter is not to be filled in.

The macro capability of the source language has two distinct advantages over code written command by command. The first is ease and speed of coding. The call of one macro is obviously easier than writing the comparable string of code. The second advantage is increase in accuracy. Not only are coding errors drastically curtailed, but if the macro is defective or needs to be changed, every occurrence of it in the lesson coding can be corrected by modifying the original macro, in general, the code can stay as it is. The more standard the various problem formats, the more valuable the macro capability becomes. Apart from a few nonstandard introductory audio messages and display items, approximately 95% of the reading curriculum has been programmed using about 110 basic macros.

The macro command feature of the language has significant implications for psychological research. By simply changing a few commands in a particular macro, one can alter the flow of the teaching sequence whenever that macro is called in the program. Thus, the logic of an instructional sequence that occurs thousands of times in the reading curriculum can be redesigned by adding or modifying a few lines of code in a given macro. If, for example, we wanted to change the timing relations, the type of feedback, or characteristics of the CRT display in the task described above, it would require only a few lines of code in the PW macro and would not necessitate making changes at every point in the curriculum where the picture-to-word exercise occurred. Thus, a range of experimental manipulations can be carried out using the same basic program and display materials, and requiring changes only in the command structure of the macros.

As indicated in Table 2, a bank of switches and counters is defined in the computer and can be used to keep a running record on each student. There is a sufficient number of these registers so that quite sophisticated schemes of optimization and accompanying branching are possible. Thus, one is in a position to present a series of words and to optimize the number of correct responses to some stipulated criteria, for example, five consecutive correct responses for each of the words. Or one can select from an array of phrases choosing those phrases for presentation that have the greatest number of previous errors. As a consequence of these decisions, each student pursues a fundamentally different path through the reading materials.

Some Results from the First Year of Operation

The Stanford CAI Project is being conducted at the Brentwood School in the Ravenswood School District (East Palo Alto, California). There were several reasons for selecting this school. It had sufficient population to provide a sample of well over 100 first-grade students. The students were primarily from "culturally disadvantaged" homes. And the past performance of the school's principal and faculty had demonstrated a willingness to undertake educational innovations.

Computerized instruction began in November of 1966 with half of the first-grade students taking reading via CAI and the other half, which functioned as a control group, being taught reading by a teacher in the classroom. The children in the control group were not left out of the project, for they took mathematics from the CAI system instead. The full analysis of the student data is a tremendous task which is still underway. However, a few general results have already been tabulated that provide some measure of the program's success.

Within the lesson material there is a central core of problems which we have termed main line problems. These are problems over which each student must exhibit mastery in one form or another. Main-line problems may be branched around by successfully passing certain screening tests, or they may be met and successfully solved; they may be met with incorrect responses, in which case the student is branched to remedial material. The first year of the project ended with a difference between the fastest and slowest student of over 4,000 main line problems completed. The cumulative response curves for the fastest, median, and slowest students are given in Figure 5. Also of interest is the rate of progress during the course of the year. Figure 6 presents the cumulative number of problems completed per hour on a month-by-month basis again for the fastest, median, and slowest student. It is interesting to note that the rate measure was essentially constant over time for increase for the fast student.

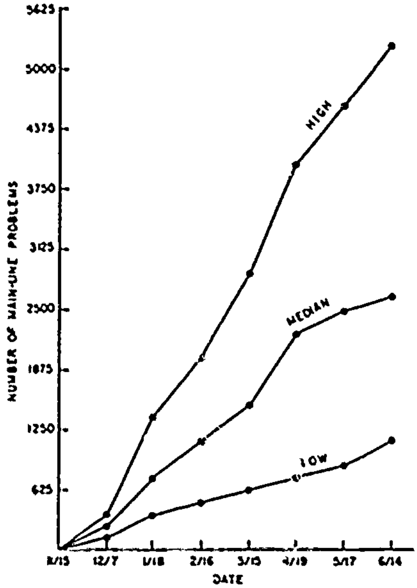


Fig. 5. Cumulative number of main-line problems for fastest, median, and slowest student.

From the standpoint of both the total number of problems completed during the year and rate of progress, it appears that the CAI curriculum is responsive to individual differences. The differences noted above must not be confused with a variation in rate of response. The difference in response rate among students was very small. The average response rate was approximately four per

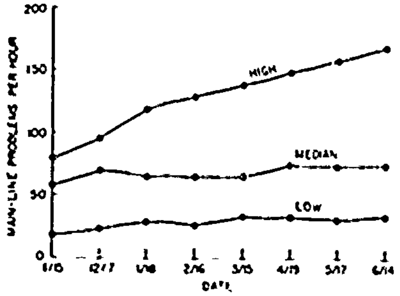


FIG. 6 Cumulative rate of progress for fastest, median, and slowest student

minute and was not correlated with a student's rate of progress through the curriculum. The differences in total number of main-line problems completed can be accounted for by the amount of remedial material, the optimization routines, and the number of accelerations for the different students.

It has been a common finding that girls generally acquire reading skills more rapidly than boys. The sex differences in reading performance have been attributed, at least in part, to the social organization of the classroom and to the value and reward structures of the predominantly female primary grade teachers. It has also been argued on developmental grounds that first-grade girls are more facile in visual memorization than boys of the same age, and that this facility aids the girls in the sight-word method of vocabulary acquisition commonly used in basal readers. If these two arguments are correct, then one would expect that placing students in a CAI environment and using a curriculum which emphasizes analytic skills, as opposed to rote memorization, would minimize sex differences in reading. In order to test this hypothesis, the rate of progress scores were statistically evaluated for sex effects. The result, which was rather surprising, is that there was no difference between male and female students in rate of progress through the CAI curriculum.

Sex differences however might be a factor in accuracy of performance. To test this notion the final accuracy scores on four standard problem types were examined. The four problem types, which are representative of the entire curriculum, were Letter Identification, Word List Learning, Matrix Construction, and Sentence Comprehension. On these four tasks, the only difference between boys and girls that was statistically significant at the .05 level was for word-list learning. These results, while by no means definitive, do lend support to the notion that when students are removed from the normal classroom environment and placed on a CAI program, boys perform as well as girls in overall rate of progress. The results also suggest that in a CAI environment the sex difference is minimized in proportion to the emphasis on analysis rather than rote memorization in the learning task. The one problem type where the girls achieved significantly higher scores than the boys, word-list learning, is essentially a paired-associate learning task.

As noted earlier, the first graders in our school were divided into two groups. Half of them received reading instruction from the CAI system; the other half did not (they received mathematics instruction instead). Both groups were tested extensively using conventional instruments before the project began and again near the end of the school year. The two groups were not significantly different at the start of the year. Table 3 presents the results for some of the tests that were administered at the end of the year. As inspection of the table will show, the group that received reading instruction via CAI performed significantly better on all of the posttests except for the comprehension subtest of the California Achievement Test. These results are most encouraging. Further, it should

TABLE 3

POSTTEST RESULTS FOR EXPERIMENTAL AND CONTROL GROUPS

Test	Experimental	Control	p value
California Achievement Test			
Vocabulary	45.91	38.10	<.01
Comprehension	41.45	40.62	—
Total	45.63	39.61	<.01
Hartley Reading Test			
Form class	11.22	9.00	<.05
Vocabulary	19.38	17.05	<.01
Phonetic discrimination	30.88	25.15	<.01
Pronunciation			
Nonsense word	6.03	2.30	<.01
Word	9.95	5.95	<.01
Recognition			
Nonsense word	18.43	15.25	<.01
Word	19.61	16.60	<.01

be noted that at least some of the factors that might result in a "Hawthorne phenomenon" are not present here, the "control" group was exposed to CAI experience in their mathematics instruction. While that may leave room for some effects in their reading, it does remove the chief objection, since these students also had reason to feel that special attention was being given to them. It is of interest to note that the average Stanford Binet IQ score for these students (both experimental and control) is 89.[3]

Owing to systems and hardware difficulties, our program was not in full operation until late in November of 1966. Initially, students were given a relatively brief period of time per day on the terminals. This period was increased to 20 minutes after the first 6 weeks, in the last month we allowed students to stay on the terminal 30 to 35 minutes. We wished to find out how well first-grade students would adapt to such long periods of time. They adapt quite well, and next year we plan to use 30-minute periods for all students throughout the year. This may seem like a long session for a first-grader, but our observations suggest that their span of attention is well over a half hour if the instructional sequence is truly responsive to their response inputs. This year's students had a relatively small number of total hours on the system. We hope that by beginning in the early fall and using half-hour periods, we will be able to give each student at least 80 to 90 hours on the terminals next year.

I do not have time to discuss the social-psychological effects of introducing CAI into an actual school setting. However, systematic observations have been made by a trained clinical psychologist, and a report is being prepared. To preview this report, it is fair to say that the students, teachers, and parents were quite favorable to the program.

Nor will time permit a detailed account of the various optimization routines used in the reading curriculum. But since this topic is a major focus of our research effort, it requires some discussion here. As noted earlier, the curriculum incorporates an array of screening and sequencing procedures designed to optimize learning. These optimization schemes vary in terms of the range of curriculum included, and it has been convenient to classify them as either short- or long-term procedures. Short term procedures refer to decision rules that are applicable to specific problem formats and utilize the very recent response history of a subject to determine what instructional materials to present next. Long term optimization procedures are applicable to diverse units of the curriculum and utilize a summarized form of the subject's complete response record to specify his future path through major instructional units.

As an example of a short-term optimization procedure, consider one that follows directly from a learning theoretic analysis of the reading task involved (Groen & Atkinson, 1966). Suppose that a list of m words is to be taught to the child, and it has been decided that instruction is to be carried out using the picture-to-word format described earlier. In essence, this problem format involves a series of discrete trials, where on each trial a picture illustrating the word being taught is presented on the projector screen and three words (including the word illustrated) are presented on the CRT. The student makes a response from among these words, and the trial is terminated by telling him the correct answer. If x trials are allocated for this type of instruction (where x is much larger than m), how should they be used to maximize the amount of learning that will take place? Should the m items be presented an equal number of times and distributed randomly over the x trials, or are there other strategies that take account of idiosyncratic features of a given subject's response record? If it is assumed that the learning process for this task is adequately described by the one-element model of stimulus sampling theory, and there is evidence that this is the case, then the optimal presentation strategy can be prescribed. The optimal strategy is initiated by presenting the m items in any order on the first m trials, and a continuation of this strategy is optimal over the remaining $x - m$ trials if, and only if, it conforms to the following rules:

1. For every item, set the count at 0 at the beginning of trial $m + 1$.

2. Present an item at a given trial if, and only if, its count is *least* among the counts for all items at the beginning of the trial.

3. If several items are eligible under Rule 2, select from these the item that has the smallest number of presentations, if several items are still eligible, select with equal probability from this set.

[3] More details on these and other analyses may be found in Atkinson (1967) and Wilson and Atkinson (1967).

4 Following a trial, increase the count for presented item by 1 if the subject's response was correct, but set it at 0 if the response was incorrect

Even though these decision rules are fairly simple, they would be difficult to implement without the aid of a computer. Data from this year's experiment establish that the above strategy is better than one that presents the items equally often in a fixed order.

This is only one example of the type of short-term optimization strategies that are used in the reading curriculum Some of the other schemes are more complex, involving the application of dynamic programming principles (Groen & Atkinson, 1966), and use information not only about the response history but also the speed of responding. In some cases the optimization schemes can be derived directly from mathematical models of the learning process, whereas others are not tied to theoretical analyses but are based on intuitive considerations that seem promising[4]

Even if short-term optimization strategies can be devised which are effective, a total reading curriculum that is optimal still has not been achieved. It is, of course, possible to optimize performance on each unit of the curriculum while, at the same time, sequencing through the units in an order that is not particularly efficient for learning The most significant aspect of curriculum development is with regard to long-term optimization procedures, where the subject's total response history can be used to determine the best order for branching through major instructional units and also the proper balance between drill and tutorial activities It seems clear that no theory of instruction is likely to use all the information we have on a student to make instructional decisions from one moment to the next Even for the most sophisticated long-term schemes, only a sample of the subject's history is going to be useful In general, the problem of deciding on an appropriate sample of the history is similar to the problem of finding an observable statistic that provides a good estimate of a population parameter. The observable history sample may be regarded as an estimate of the student's state of learning A desirable property for such a history sample would be for it to summarize all information concerning the current learning state of the student so that no elaboration of the history would provide additional information. In the theory of statistical inference, a statistic with an analogous property is called a sufficient statistic. Hence, it seems appropriate to call an observable sample history with this property a "sufficient history."

In the present version of the reading curriculum, several long-term optimization procedures have been introduced with appropriate sufficient histories As yet, the theoretical rationale for these procedures has not been thoroughly worked out, and not enough data have been collected to evaluate their effectiveness. However, an analysis of long-term optimization problems, and what data we do have, has been instructive and has suggested a number of experiments that need to be carried out this year. It is my hope that such analyses, combined with the potential for educational research under the highly controlled conditions offered by CAI, will lay the groundwork for a theory of instruction that is useful to the educator. Such a theory of instruction will have to be based on a model of the learning process that has broad generality and yet yields detailed predictions when applied to specific tasks.

In my view, the development of a viable theory of instruction and the corresponding learning theory will be an interactive enterprise, with advances in each area influencing the concepts and data base in the other. For too long, psychologists studying learning have shown little interest in instructional problems, whereas educators have made only primitive and superficial applications of learning theory. Both fields would have advanced more rapidly if an appropriate interchange of ideas and problems had existed. It is my hope that prospects for CAI, as both a tool for research and a mode of instruction, will act as a catalyst for a rapid evolution of new concepts in learning theory as well as a corresponding theory of instruction.

[4] The learning models and optimization methods that underlie much of the reading curriculum are discussed in Atkinson and Shiffrin (1968), Groen and Atkinson (1966), Rodgers (1967), and Wilson and Atkinson (1967).

REFERENCES

Atkinson, R C., & Hansen, D. N. Computer-assisted instruction in initial reading The Stanford Project. *Reading Research Quarterly*, 1966, 2, 5-25.

Atkinson, R. C., & Shiffrin, R M. Human memory. A proposed system and its control processes. In K. W Spence & J. T. Spence (Eds), *The psychology o learning and motivation. Advances in research an* *theory* Vol. 2. New York. Academic Press, 1968, in press.

FISHMAN, E. J., KELLER, L., & ATKINSON, R. C. Massed vs distributed practice in computerized spelling drills. *Journal of Educational Psychology*, 1968, 59, in press.

GROEN, G. J., & ATKINSON, R. C. Models for optimizing the learning process. *Psychological Bulletin*, 1966, 66, 309–320.

HANSEN, D. N., & RODGERS, T. S. An exploration of psycholinguistic units in initial reading. Technical Report 74, 1965, Stanford University, Institute for Mathematical Studies in the Social Sciences.

RODGERS, T. S. Linguistic considerations in the design of the Stanford computer-based curriculum in initial reading. Technical Report 111, 1967, Stanford University, Institute for Mathematical Studies in the Social Sciences.

SUPPES, P. The uses of computers in education. *Scientific American*, 1966, 215, 206–221.

SUPPES, P., JERMAN, M., & GROEN, G. J. Arithmetic drills and review on a computer-based teletype. *Arithmetic Teacher*, 1966, April, 303–308.

WILSON, H. A., & ATKINSON, R. C. Computer-based instruction in initial reading. A progress report on the Stanford project. Technical Report 119, 1967, Stanford University, Institute for Mathematical Studies in the Social Sciences. (To be published in H. Levin & J. Williams, Eds.), *Basic studies in reading*. New York. Harper and Row.)

TEACHING MACHINES: A REVIEW[1]

CHARLES S. MORRILL[2]

The same forces which have characterized the evolution of general educational practices are inherent to the history of the new science of automated teaching. As a result of the expansion and multiplying complexities of political, economic, and social interests, there developed an ever increasing need for the rapid education of large numbers of people. New educational objectives demanded new methods of instruction, and the history of education is marked by many diverse attempts at establishing more efficient teaching procedures. Once again teaching methods must be reevaluated. Rigid adherence to the principle of personal teacher-student relationships no longer seems feasible—an instructional system more appropriate for present-day needs must be established. It is probable that the use of automated teaching devices can fill this need in the method of education. As Corrigan (1959) has suggested:

> the automated teaching method has grown out of a pressing need. This need has been created by a twofold technical training problem. As advances in science and technology have been made, there has been an ever increasing demand for well trained instructors; at the same time the availability of these trained persons has been diminishing. This situation is aggravated further, by the increased scope and complexity of subjects, and the ever increasing ratio between number of instructors and students (p. 24).

Current Trends in Automated Teaching Machines

Current interest in the area of automated teaching machines is well illustrated by the simple index of frequency-per-year of published teaching machine articles. Fry, Bryan, and Rigney (1960) report that for the years prior to 1948 there are only 6 references, whereas through 1959 there were more than 50 reports published.

The grandfather of automated teaching machines is Sydney L. Pressey (1926, 1927), who designed machines for automated teaching during the mid-1920s. His first device was exhibited and described at the American Psychological Association (APA) meetings in 1924; an improved device was exhibited in 1925 at the APA meetings. Both forms of the apparatus automatically performed simultaneous administration and scoring of a test and taught informational and drill material. Pressey's device, about the size of a portable typewriter, presented material to the subject via a small window. Four keys were located alongside the apparatus. If the student activated a key corresponding to the correct answer, the machine advanced to the next item. If his response was incorrect, the machine scored an error and did not advance to the next item until the correct answer was chosen. The capacity of the drum was 30 two-line typewritten items; the paper on which the questions appeared was carried as in a typewriter.

---

[1] The research reported in this document was supported by the Department of the Air Force under Air Force Contract AF-33-(600)-39852.

[2] The author wishes to acknowledge the valuable editorial assistance of Sylvia Pilsucki.

PSYCHOLOGICAL BULLETIN, 1961, Vol. 58, pp. 363-375.

In 1927, Pressey summarized his efforts as follows:

> The paper reports an effort to develop an apparatus for teaching drill material which (a) should keep each question or problem before the learner until he finds the correct answer, (b) should inform him at once regarding the correctness of each response he makes, (c) should continue to put the subject through the series of questions until the entire lesson has been learned, but (d) should eliminate each question from consideration as the correct answer for it has been mastered (p. 552).

In 1930, Peterson devised a self-scoring, immediate feedback device. The Chemo Card, as this device was later called, utilized the technique of multiple choice. A special ink was used by the student in marking his answer. The mark appeared red if the answer was incorrect; a dark color resulted if the answer was correct. Although Pressey's notions and the Chemo Card might have stimulated an interest in automated teaching techniques in the twenties, educators and researchers obviously were not at that time ready for this advanced concept of teaching. Automated teaching did not take hold.

In 1932, Pressey published an article describing a kind of answer sheet which could be scored by an automatic scoring device. This apparatus recorded errors by item, and thus provided the instructor with clues as to what questions needed further instruction. In 1934, Little experimented with this device as well as with the device originated by Pressey in 1926. His results favored the use of automated devices in contrast to regular classroom techniques.

The next appearance of automated teaching literature came a considerable number of years later. During World War II, the Automatic Rater was used by the Navy for training. This device projected a question on a small screen, the subject's response consisted of pushing one of five buttons.

In 1950, Pressey described a new automated device called the Punchboard. Multiple-choice questions were presented to the student. The key answer sheet inside the Punchboard contained holes opposite the correct answers only. If the answer was correct, the student's pencil penetrated deeply; if incorrect, the pencil did not penetrate the paper significantly. Angell and Troyer in 1948 and Angell in 1949 reported the results of using the Punchboard. Both studies suggested the superiority of this method over traditional classroom procedures.

In 1954, Skinner published "The Science of Learning and the Art of Teaching," which provided the basis for the development of his teaching machines. In this article, he stressed the importance of reinforcement in teaching and suggested teaching machines as a method of providing this needed reinforcement for the learner.

Reports concerning the Subject-Matter Trainer began to appear in 1955 (Besnard, Briggs, Mursch, & Walker, 1955; Besnard, Briggs, & Walker, 1955). This electromechanical device is a large multiple-choice machine used essentially for training and testing in the identification of components and in general verbal subject matter. Extensive research has been done with this device because of its considerable flexibility, i.e., it allows several modes of operation for self-instruction, variety of programed subject matter, drop-out feature after items have been mastered, etc.

The Pull-Tab, used experimentally by Bryan and Rigney in 1956, was a device in which the subject received not only a "right" or "wrong" indication after his choice but also a somewhat detailed explanation of "why" a response was incorrect. In 1949, Briggs had found in experimenting with the Punchboard that learning is significantly enhanced by immediate knowledge of results. Bryan and Rigney's data illustrated that the combination of immediate knowledge of results plus explanation, if the student is in error, produced significantly higher scores on a criterion test than if no explanation had been given. The importance of this research from a historical point of view is that it investigated immediate knowledge of results as a factor existing on a continuum with varying degrees of effect. Up to this point any comparison involving the effectiveness of teaching machines had been one between classroom instruction and the "new" machine under consideration. In Briggs' and in Bryan and Rigney's research, however, we see the beginning of a concern, to become greater in the next few years, with the possible effects of specific variables and their interactions on learning.

The years 1957-58 mark the beginning of the period in which resurgent interest in teaching machines was initiated. Ramo's arguments (1957) reopened the consideration of automated techniques for classroom use. His article served as one of the more forceful attempts to alert educators to the needs and requirements for automated techniques in education. Skinner's continued interest (1958) served as the major catalyst in this area. In his article, he reviewed earlier attempts to stimulate interest in teaching machines and further explained that the learning process was now better understood and that this increased sophistication would be reflected in teaching machine technology. Skinner suggested that the most appropriate teaching machine would be that which permits the student to *compose* his response rather than to select it from a set of alternatives. On the basis of this philosophy and in conjunction with other principles of learning theory to which Skinner adheres, he designed a teaching machine with the following characteristics. The questions, printed on a disk, are presented to the student through a window. The student's response is written on a paper tape, which is advanced under a transparent cover when the student lifts a lever. At this point the correct answer appears in the window. If the student is correct, he activates the lever in one manner, which eliminates the item from the next sequence. If he is incorrect, the lever is activated in a different manner, thus retaining the item in the next sequence.

Holland (1960), a co-worker of Skinner's, has suggested several well-known learning principles that should be applied to teaching machine technology. immediate reinforcement for correct answers is a must, learned behavior is possible only when it is *emitted* and reinforced, gradual progression (i.e., small steps in learning sequences and reducing wrong answers) is necessary to establish complex repertoires, gradual withdrawal (fading or vanishing) of stimulus support is effective, it is necessary to control the student's observing and echoic behavior and to train for discrimination, the student should write his response. The Skinner machine does in fact employ these principles.

Ferster and Sapon (1958) described the Cardboard Mask, a most simple teaching machine which employs the principles which Skinner and Holland outline so clearly. This device is a cardboard folder containing mimeographed material which is presented one line at a time. The student, after writing his response on a separate sheet of paper, advances the paper in the mask, thereby exposing the correct response.

In 1958, a number of investigators interested in teaching machines recommended that the programed material be a function of the student's response. This idea suggests that a "wrong" response may not necessarily be negative reinforcement and that both the "right" and "wrong" responses should modify the program. Rath and Anderson (1958) and Rath, Anderson, and Brainerd (1959) have suggested the use of a digital computer which automatically adjusts problem difficulty as a function of the response. Crowder's (1958, 1959a, 1959b) concept of "intrinsic programming" permits the response to alter the programing sequence.

During the last few years, researchers have been focusing their attention on investigating many of the variables which are pertinent to the design and use of teaching machines. The seemingly simple task of defining a teaching machine has been a serious problem to many authors (Day, 1959; Silberman, 1959; Weimer, 1958). Some definitions have made more extensive demands on teaching devices than others. Learning theorists (Kendler, 1959, Porter, 1958, Skinner, 1957, Spence, 1959, Zeaman, 1959) are now most outspoken concerning the application of theoretical concepts to teaching machine technology. Transfer of training, mediational processes, reinforcement, motivation, conditioning, symbolic processes, and language structure are but a few of these areas of interest.

There are indeed many other variables about which there is a divergence of opinion and about which experimental evidence is completely lacking or controversial. The reports of Skinner (1958), Israel (1958), Coulson and Silberman (1960), Fry (1959), and Stephens (1953) are all focused, at least in part, on questions related to response modes, e.g., multiple choice, construction of the response, responses with reinforcement, etc. Briggs, Plashinski, and Jones (1955) investigated self-paced vs. automatically paced machines. The importance of motivation in connection with teaching machines has been explored by Holland (unpublished), Mayer and Westfield (1958), and Mager (1959).

Essentially, the history of automated teaching is short --it started in the mid-twenties and was strenuously reactivated by the appearance of Skinner's 1958 article. Empirical investigations of many important issues in this field are just now beginning to appear. However, the necessity of developing automated teaching methods has been evident for many years.

General Problem Areas

Definition

As in any new field, the first problem is one of definition. What is a teaching machine? Silberman (1959) says that a teaching device consists of four units: an input unit, an output unit, a storage unit, and a control unit. As such, this definition includes a broad category of devices, from the most simple to the most complex. Weimer (1958) goes beyond the device itself, stating that a teaching machine must present information to the student as well as test the student by means of a controlled feedback loop. Crowder (1960) insists that a teaching machine must in some way incorporate two way communication. That is, the student must respond to the information presented by the machine, and the machine must in turn recognize the nature of the student's response and behave appropriately (p. 12).

Perhaps the most inclusive definition is one given by Day (1959):

A teaching machine is a mechanical device designed to present a particular body of information to the student. . . . Teaching machines differ from all other teaching devices and aids in that they require the active participation of the learner at every step (p. 591).

Although the emphasis in some of the above concepts is different, together they give a rather complete description and, if you will, definition.

Programing

The programing of subject matter for teaching machines is the most extensive and difficult problem in this new technology. Beck (1959) describes specific concepts which he thinks appropriate for programing a Skinner-type machine:

A student's responses may be restricted and guided in a great number of ways. These range from all types of hints . . . to simply presenting the response which it is desired a student acquire (p. 55).

Carr (1959) discusses in some detail the importance of programing in terms of learning efficiency and retention. Much of what he says remains open for empirical verification. Rothkopf (1960) has suggested that the development of programed instruction suffers from two difficulties: a weak rational basis for program writing and inadequate subject-matter knowledge among program writers.

The extent to which any initial program needs revision is perhaps exemplified by the program in Harvard's course Natural Sciences 114. Holland points out that the first program of materials included 48 disks, each containing 29 frames, whereas a revision and extension of the program the following year included 60 disks of 29 frames each. Holland's objective was to extend the program and decrease the number of student errors. Crowder's (1960) programing objectives are different from Holland's. He states:

By means of "intrinsic programming" it [the program] recognizes student errors as they occur and corrects them before they can impede understanding of subsequent material or adversely affect motivation (p. 12).

Crowder considers it almost impossible to write a program which completely avoids error, and therefore he structures the program requirements on the probability of error. When an error is made, the next presentation explains the subject's mistake. Depending on the nature of the error and when it occurs, the subject may either return to the original question or enter a program of correctional material.

Another concept for programing is known as *branching* (Bryan & Rigney, 1959). Through branching, many possible routes are provided through which the subject can proceed, depending on the response. The subjects are allowed to skip certain material if they have demonstrated a knowledge of it. One study (Coulson & Silberman, 1960) suggests that under branching conditions subjects require less training time than under nonbranching conditions, however, results on the criterion test were not significantly different.

For certain kinds of subject matter, *vanishing* is still another concept for programing (Skinner, 1958). A complete or nearly complete stimulus is presented to the subject. Subsequent frames gradually omit part of the stimulus until all of it is removed. The subject is then required to reconstruct the stimulus.

To program verbal learning sequences, Homme and Glaser (1959) suggest the Ruleg. With this method, the written program states a rule and provides examples for this rule. In each case, either the rule or the example is incomplete, requiring the subject to complete it.

In a recent study Silverman (1960b) investigated methods of presenting verbal material for use in teaching machines. He recommended that further research involving the design and use of teaching machines should take into consideration the possible use of context cues as a means of facilitating serial rote learning. At the same time, however, he stated that continuous use of context cues as ancillary prompts should be avoided, since such prompts can interfere with learning.

The optimum size of steps and the organization of the programed material are two formidable problems. Skinner (1958) states:

> Each step must be so small that it can always be taken, yet in taking it the student moves somewhat closer to fully competent behavior (p. 2).

In order to determine the value of steps in a program, Gavurin and Donahue (1961) investigated the effects of the organization of the programed material on retention and rate of learning. They state that the assumption that optimum teaching machine programs are those in which items are presented in a logical sequence has been validated for acquisition but not retention. The results of a study carried out by Coulson and Silberman (1959) indicated that small steps were more time consuming but resulted in statistically significant higher test scores on one of the criterion tests. Pressey (1959) in principle disagrees with Skinner's notions of short and easy steps, and he strongly suggests an experimental investigation of this question. Both rate of learning and retention (recall or recognition) are of critical concern.

The above discussion suggests several areas which are directly applicable to programing and which are under investigation and/or need further experimentation. Indeed, there are a number of unanswered questions in the programing complex, some of which have been suggested by Galanter (1959):

1. What is the correct order of presentation of material?
2. Is there an optimum number of errors that should be made?
3. How far apart (in some sense) should adjacent items be spaced?
4. Is experimentally controlled pacing more effective (in some sense) than self-pacing?
5. Is one program equally effective for all students?
6. What are the effects of using different programing techniques (branching, intrinsic programing, vanishing) in various subject-matter areas?
7. What criteria are most appropriate in the evaluation of student learning?

These questions are but a few of the intriguing and complex problems facing investigators in the new field of programing material for teaching machines. Answers to these questions will help not only the educator but also the engineer who is concerned with writing adequate specifications for the construction of teaching machines.

Response Mode

The kind of response that should be given by a subject has been a controversial question in the teaching machine field. Pressey's original machine (1926) required the subject to press a lever corresponding to his choice of answer. The format of the answers was multiple-choice. Skinner (1958) emphasized the necessity of having the subject *compose* (construct) the response. Skinner states.

> One reason for this is that we want him to recall rather than recognize—to make a response as well as see that it is right. Another reason is that effective multiple-choice material must contain plausible wrong responses, which are out of place in the delicate process of "shaping" behavior because they strengthen unwanted forms (p. 2).

Coulson and Silberman (1960) investigated this question of multiple-choice vs. constructed response by using *simulated* teaching machines—human beings were used instead of automatic control mechanisms. Their results indicated that the multiple-choice response mode required significantly less time than the constructed response mode and that no significant difference was obtained between response modes on the criterion test. Further, they reported that no significant differences were obtained among the experimental groups on the multiple-choice criterion subtest or on the total (multiple-choice plus constructed response) criterion test. Fry (1959) has discussed this response-mode question along with other variables, and he has carried out extensive research concerning constructed vs. multiple-choice response modes. The results of his study favor the use of constructed response when recall is the objective of the learning.

In addition to the basic controversy (which needs much more investigation) between multiple-choice and constructed responses, there are several "variations on the theme" which are evident. Stephens (1953) has recommended that every wrong answer in a multiple choice question appear as a correct choice for another item. He calls this program "inside alternatives.' His data indicate that there was no difference between control and experimental groups on a criterion test using either nonsense syllables or Russian unless each right choice appeared as a wrong alternative for the three subsequent items. The use of prompts in general has been shown to be an effective technique in automated teaching (Cook, 1958; Cook & Kendler, 1956; Cook & Spitzer, 1960).

Using learning booklets. Goldbeck (1960) investigated the effect of response mode and learning material difficulty on automated instruction. The three response modes used were: overt response (the subject was required to construct a written response), covert response (the subject was permitted to think of a response), and implicit response (the subject read the response which was underlined). Goldbeck states:

> Learning efficiency scores, obtained by dividing quiz scores by learning time, showed that the implicit (reading) response condition produced significantly more efficient learning than the overt response condition. The covert response condition fell between the other conditions in learning efficiency (p. 25).

Concerning quiz-score results, the overt response group

> performed significantly poorer than the other response mode groups at the easy level of difficulty. Performance of the overt response group improved significantly at the intermediate difficulty level to the extent that it exceeded the performance of all other groups (pp. 25–26).

Goldbeck concludes that

> doubt is cast upon the assumption that the best learning is achieved by use of easy items and requiring written constructed responses (p. 26).

To the author's knowledge, the use of an oral response in conjunction with the Skinner teaching machine and its effect on learning rate and retention have not been reported in the literature. Furthermore, the importance of response mode as a function of reinforcement must be specified. Israel (1958) has suggested that natural and artificial reinforcement may affect the subjects' learning. A most comprehensive analysis of response-mode and feedback factors has been reported by Goldbeck and Briggs (1960).

The general area of reinforcement suggests problems related to the drop-out feature of teaching machines. Pressey's (1927) original machine dropped items after the correct answer had been given twice. Skinner's machines at the Harvard Psychological Laboratory also have the drop-out feature, although the commercially available machines based on Skinner's design do not incorporate this feature. With reference to a study carried out at Harvard, Holland (unpublished) reported significantly superior performance when the drop-out feature was used.

If items are dropped, the sequence of items is of course changed. How important is the sequence? If items should be dropped, by what criterion of learning can one justify omitting an item from the sequence? If items are not dropped and the criterion for the learning procedure is a complete run (i.e., once through the sequence without error), what is the effect upon retention? Being correct is positive reinforcement, thus, some items under these circumstances will receive a greater amount of positive reinforcement than others. What would be the effect of additional reinforcements with or without drop-out? Again, a plethora of problems and a paucity of answers!

Response time, another important variable, has been investigated by Briggs, Plashinski, and Jones (1955). Their study suggests that there is no difference between self paced and automatically paced programs as determiners of response time. However, the problem of pacing for individual items is still a recent one and needs further research. Another aspect of response time the distribution of practice has been studied extensively since Ebbinghaus' investigation in 1885. For example, Holland (unpublished) states that in an experiment at Harvard "a few students completed all the disks in a small number of long sessions while others worked in many short sessions. . . . Apparently the way practice was distributed made little difference" (p. 4). Nevertheless, the distribution of practice, like the problem of pacing, is yet a subject of controversy, with most investigations favoring some form of distributed practice (Hovland, 1951).

The above section outlines briefly some of the major problems associated with the variables affecting response mode. Although some of the variables have already been investigated, these and others, together with their interactions, need further research.

Knowledge of Results

There are many peripheral problems related to teaching machines, one of which is the effect of immediate knowledge of results on learning. Angell (1949), using a multiple-choice punchboard technique, found that "learning is significantly enhanced by immediate knowledge of results." Briggs (1949), also using the Punchboard, confirmed these results. Bryan and Rigney (1956) noted superior performance when subjects were given knowledge of results, specifically, an explanation if the answer was incorrect. This last study was later expanded by Bryan, Rigney, and Van Horn (1957), who investigated differences between three kinds of explanation given for incorrect response. None of the three types of explanation proved to be superior in teaching the subjects. Because of their controvertible results, the above studies demonstrate that, although immediate knowledge of results appears to be effective in the learning process, this problem contains many facets which need more empirical data.

Motivation

One of the many reasons given for the effectiveness of teaching machines is that the student's motivation is increased. Psychologists and educators have realized for some time that the motivation variable ranks very high among those variables pertinent to learning. In 1958 and 1959, Holland surveyed the use of the teaching machine in classes at Harvard. He found that most students felt that they would have gotten less out of the course if the machines had not been used, that most students preferred to have machines used for part of the course, and finally that most students felt that the teaching machine was used by the instructor "to teach me as much as possible with a given expenditure of my time and effort." During a field tryout of the Subject Matter Trainer in the Semi-automatic Ground Environment System, Mayer and Westfield (1958) observed that "motivation to work with the trainer is high." The supervisory as well as the operational personnel encouraged the use of this training technique.

Mager (1959) suggests that motivation and interest are a function of the percentage of correct responses. He observed that in two young subjects negative feelings for learning mathematics in the usual classroom situation did not transfer to learning mathematics by means of a teaching machine. The cause of this phenomenon is perhaps best explained by the subjects' statement that, because they were able to understand the programed material, it did not seem to be mathematics at all. This interesting relationship between comprehension and motivation needs further investigation.

Equipment

There are many inexpensive models of teaching machines which will soon hit the consumer market. For much of this equipment, there is very little experimental evidence which supports the various designs. As previously pointed out, Holland has collected data which support the efficiency of the drop-out feature in a teaching machine; yet commercial models presently available do not incorporate this feature, presumably because of its high cost. Generally, it seems that production is now and will continue to be out of phase with much of the research which has provided necessary teaching machine specifications. Moreover, because of their expense, it is likely that some very important features will be omitted in manufacture.

The methods of displaying programed material, another unexplored problem area, must be investigated so as to provide the design engineer with requirements based on empirical findings. The display problem is less acute, perhaps, with material for the elementary school than it is with programs designed to teach maintenance procedures and aspects of the biological sciences.

The use of computer controlled

teaching machines has been recommended by many authors (Coulson & Silberman, 1959, Skinner, 1958). Utilizing a central computer, with many programs capable of adapting to individual needs and of providing stimulus materials to 50 or more students simultaneously, is a feasible notion for large-scale training programs. With a computer, the display problem again becomes a major issue Training in pattern recognition, information handling, and display interpretation are but a few appropriate areas which should be studied. The alternate modes of presentation become more extensive as computer capacity increases. In the case of certain kinds of subject matter, a computer generated, pictorial display of information may be a more effective presentation than other display techniques. Future research must solve these problems in equipment design.

Teaching Machines and Other Techniques

The use of automated teaching devices may be optimized, perhaps, if there is a proper balance between this technique and other compatible teaching methods. What percentage of a course should be machine taught? What subject matter is best suited to automated devices? If classroom courses were as carefully and thoughtfully programed as some of the programs currently being prepared for teaching machines, might some of the advantages of machines diminish? Perhaps some of the apparent advantages of teaching machines are no more than methods of illustrating correctable classroom techniques! It might well be that the instructor's enthusiasm and inspiration, a factor supposedly dominant in higher education, is vital in mastering a particular subject-matter area. Will creativity in certain students be harmed by extensive education via the machine? Again, consideration of the use of a teaching machine, the subject matter, the program, the level of education, and the techniques used in combination with the teaching machine provide a fertile field for experimentation. As of now, questions in this area remain unanswered. Silverman (1960a) has presented an excellent, detailed discussion of problems inherent in this new technology of automated teaching and the current trends in the field.

Problems of Application

The most obvious problems in the attempt to use automated teaching techniques have been outlined in the previous section. There is still much of the unknown associated with techniques, machines, programing, etc. to be eliminated before a direct solution to a particular training problem can be specified. Many alternatives exist, the best of which has not yet been determined. In addition to these voids, there is a serious lack of definition in the objectives of many training programs.

What is the objective of a particular automated course or program? From a pragmatic point of view, what are the criteria by which a specific educational program can be evaluated? For example, the objectives might range from the teaching of rote tasks to the presentation of more abstract material. Needless to say, the techniques for both teaching and evaluating learning could be substantially different in each case. The purpose of teaching, the objective of an educational program, must be initially defined. Only then will the concepts *learning* and *teaching* be meaningful in a particular context.

After definition, the next step is to determine what subject matter will provide the student with the necessary information. It is at this point that the major pitfall in education is likely to appear. Even though many training programs do not have a defined objective, their course content is nonetheless prescribed, and the text and/or materials used in previous, nonautomated courses become the prime source of material for an automated teaching program. To program an automated teaching machine with presently available materials might well result only in a more efficient method of teaching the wrong material!

The third step requires decisions in the selection of appropriate teaching techniques. Answers to questions involving programing, choice of teaching machine, learning procedures, pacing, and response modes are still not known.

The fourth and last step requires an evaluation of the selected automated teaching method in terms of the originally established objectives. Conventional methods of instruction should be compared with the innovative methods by means of a specific set of criteria, e.g., in terms of training time, job performance, retention of learned information, etc.

The questions confronting the researcher in teaching machine technology are one example of the broader questions of man-machine interrelation. Data pertinent to the principles of human engineering, the optimum man-machine interaction, the degree to which the machine can perform functions formerly allocated to man, and the appropriate allocation of functions between man and machine will be provided by a research program investigating teaching machines. Inadequate attention to any of the above-mentioned steps will result in failure to provide the needed answers in a field which may increase training effectiveness and reduce training costs.

REFERENCES

Angell, G. W. The effect of immediate knowledge of quiz results on final examination scores in freshman chemistry. *J. educ. Res.*, 1949, 42, 391-394.

Angell, G. W., & Troyer, M. E. A new self scoring test device for improving instruction. *Sch. Soc.*, 1948, 67, 84-85.

Beck, J. On some methods of programming. In E. H. Galanter (Ed.), *Automatic teaching. The state of the art.* New York. Wiley, 1959. Pp. 55-62.

Besnard, G. G., Briggs, L. J., Mursch, G. A., & Walker, L. S. Development of the subject-matter trainer. *USAF Personnel Train. Res. Cent. tech. Memo.*, 1955, No. ASPRL-TM-55-7.

Besnard, G. G., Briggs, L. J., & Walker, E. S. The improved subject-matter trainer. *USAF Personnel Train. Res. Cent. tech. Memo.*, 1955, No. ASPRL-TM-55-1?

Briggs, L. J. The development and appraisal of special procedures for superior students and an analysis of the effects of knowledge of results. *Abstr. Doctoral Dissertations, Ohio State U.*, 1949, No. 58.

Briggs, L. J., Plashinski, D., & Jones, D. L. Self-pacing versus automatic pacing of practice on the subject-matter trainer. *USAF Personnel Train. Res. Cent. lab. Note*, 1955, No. ASPRL-LN-55-8.

Bryan, G. L., & Rigney, J. W. An evaluation of a method for ship-board training in operations knowledge. *U. Sth. Calif. Electronics Personnel Res. Group tech. Rep.*, 1956, No. 18.

Bryan, G. L., & Rigney, J. W. Current trends in automated tutoring and their implications for naval technical training. *U. Sth. Calif. Dept. Psychol. tech. Rep.*, 1959, No. 29.

Bryan, G. L., Rigney, J. W., & Van Horn, C. An evaluation of three types of information for supplementing knowledge of results in a training technique. *U. Sth. Calif., Electronics Personnel Res. Group tech. Rep.*, 1957, No. 19.

Carr, W. J. Self-instructional devices. A review of current concepts. *USAF WADC tech. Rep.*, 1959, No. 59 503.

Cook, J. O. Supplementary report. Processes underlying learning a single paired-associate item. *J. exp. Psychol.*, 1958, 56, 455.

COOK, J. O., & KENDLER, T. S. A theoretical model to explain some paired-associate learning data. In G. Finch & F. Cameron (Eds.), *Symposium on Air Force human engineering, personnel, and training research.* Washington, D.C.. National Academy of Sciences–National Research Council, 1956. Pp. 90–98.

COOK, J. O., & SPITZER, M. E. Supplementary report. Prompting versus confirmation in paired-associate learning. *J. exp. Psychol.*, 1960, 59, 275–276.

CORRIGAN, R. E. Automated teaching methods. *Automated teach. Bull.*, 1959, 1(2), 23–30.

COULSON, J. E., & SILBERMAN, H. F. Proposal for extension of automated teaching projects. *Sys. Develpm. Corp. field Note*, 1959.

COULSON, J. E., & SILBERMAN, H. F. Effects of three variables in a teaching machine. *J. educ. Psychol.*, 1960, 51, 135–143.

CROWDER, N. A. *An automatic tutoring book on number systems.* Vol. 1. Timonium, Md.. Hoover Electronics Co., 1958.

CROWDER, N. A. Automatic tutoring by means of intrinsic programming. In E. H. Galanter (Ed.), *Automatic teaching: The state of the art.* New York. Wiley, 1959. Pp. 109–116. (a)

CROWDER, N. A. The concept of automatic tutoring. *USAF Personnel Train. Res. Cent. organizational Pap.*, 1959. (b)

CROWDER, N. A. The "tutor." *J. Amer. Soc. Train. Dir.*, 1960, 14(5), 12–17.

DAY, J. H. Teaching machines. *J. chem. Educ.*, 1959, 36, 591–595.

FERSTER, C. B., & SAPON, S. M. An application of recent developments in psychology to the teaching of German. *Harv. educ. Rev.*, 1958, 28, 58–69.

FRY, E. B. Teaching machine dichotomy: Skinner versus Pressey. Paper presented at American Psychological Association, Cincinnati, September 1959.

FRY, E. B., BRYAN, G. L., & RIGNEY, J. W. Teaching machines. An annotated bibliography. *Audio-Visual commun. Rev.*, 1960, 8, Suppl. 1, 1–80.

GALANTER, E. H. The ideal teacher. In E. H. Galanter (Ed.), *Automatic teaching. The state of the art.* New York. Wiley, 1959. Pp. 1–11.

GAVURIN, E. I., & DONAHUE, V. M. Logical sequence and random sequence. *Automated teach. Bull.*, 1961, 1(4), 3–9.

GOLDBECK, R. A. The effect of response mode and learning material difficulty on automated instruction. *Amer. Inst. Res. tech. Rep.*, 1960, No. AIR-328-60-IR-124.

GOLDBECK, R. A., & BRIGGS, L. J. An analysis of response mode and feedback factors in automated instruction. *Amer. Inst. Res. tech. Rep.*, 1960, No. AIR-328-60-IR-133.

HOLLAND, J. G. Teaching machines. An application of principles from the laboratory. In, Proceedings of the Educational Testing Service Invitational Conference, October 1959, *The impact of testing on the educational process.* Princeton, N. J.. Educational Testing Service, 1960.

HOMME, L. E., & GLASER, R. Problems in programing verbal learning sequences. Paper presented in symposium on research issues in study of human learning raised by developments in automated teaching methods, American Psychological Association, Cincinnati, September 1959.

HOVLAND, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York. Wiley, 1951. Pp. 613–689.

ISRAEL, M. L. Skinnerian psychology and educational redesign. Paper read in symposium, American Psychological Association, Washington, D. C., September 1958.

KENDLER, H. H. Teaching machines and psychological theory. In E. H. Galanter (Ed.), *Automatic teaching: The state of the art.* New York: Wiley, 1959. Pp. 177–185.

LITTLE, J. K. Results of use of machines for testing and for drill upon learning in educational psychology. *J. exp. Educ.*, 1934, 3, 45–49.

MAGER, R. F. Preliminary studies in automated teaching. *IRE Trans. Educ.*, 1959, E-2, 104–107.

MAYER, S. R., & WESTFIELD, R. L. A field tryout of a teaching machine for training in SAGE operations. *USAF Cambridge Res. Cent. tech. Memo.*, 1958, No. OAL-TM-58-16.

PETERSON, J. C. A new device for teaching, testing, and research in learning. *Trans. Kans. Acad. Sci.*, 1930, 33, 41–47.

PORTER, D. Teaching machines. *Harv. Grad. Sch. Educ. Ass. Bull.*, 1958, 3(1), 1–5.

PRESSEY, S. L. A simple apparatus which gives tests and scores—and teaches. *Sch. Soc.*, 1926, 23, 373–376.

PRESSEY, S. L. A machine for automatic teaching of drill material. *Sch. Soc.*, 1927, 25, 549–552.

PRESSEY, S. L. A third and fourth contribution toward the coming "industrial revolution" in education. *Sch. Soc.*, 1932, 36, 668–672.

PRESSEY, S. L. Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. *J. Psychol.*, 1950, 29, 417 447.

PRESSEY, S. L. Certain major psycho-educational issues appearing in the conference on teaching machines. In E. H. Galanter (Ed.), *Automatic teaching: The state of the art.* New York. Wiley, 1959. Pp. 187-198.

RAMO, S. A new technique of education. *Engng. sci. Mon*, 1957, 21 (October), 17-22.

RATH, G. J., & ANDERSON, NANCY S. The IBM research center teaching machine project: I. The teaching of binary arithmetic. II. The simulation of a binary arithmetic teaching machine on the IBM 650. Paper presented at USAF Office of Scientific Research symposium on teaching machines, University of Pennsylvania, December 8-9, 1958.

RATH, G. J., ANDERSON, NANCY S., & BRAINERD, R. C. The IBM research center teaching machine project. In E. H. Galanter (Ed.), *Automatic teaching: The state of the art.* New York. Wiley, 1959. Pp. 117-130.

ROTHKOPF, E. Z. A do-it-yourself kit for programmed instruction. *Teach. Coll. Rec.*, 1960. 62, 195 201.

SILBERMAN, H. F. Introductory description of teaching machines (physical characteristics). Paper read in symposium on automated teaching, Western Psychological Association, San Diego, California, April 1959. (Abstract)

SILVERMAN, R. E. Automated teaching. A review of theory and research. USN Training and Development Center, 1960. (ASTIA AD-241 283) (a)

SILVERMAN, R. E. The use of context cues in teaching machines. USN Training and Development Center, 1960. (ASTIA AD-238 777) (b)

SKINNER, B. F. The science of learning and the art of teaching. *Harv. educ. Rev.*, 1954, 24, 86-97.

SKINNER, B. F. *Verbal behavior.* New York. Appleton-Century-Crofts, 1957.

SKINNER, B. F. Teaching machines. *Science*, 1958, 128, 969-977.

SPENCE, K. W. The relation of learning theory to the technology of education. *Harv. educ. Rev.*, 1959, 29, 84-95.

STEPHENS, A. L. Certain special factors involved in the law of effect. *Abstr. Doctoral Dissertations, Ohio State U.*, 1953, No. 64.

WEIMER, P. K. A proposed "automatic" teaching device. *IRE Trans. Educ.*, 1958, E-1, 51-53.

ZEAMAN, D. Skinner's theory of teaching machines. In E. H. Galanter (Ed.), *Automatic teaching: The state of the art.* New York. Wiley, 1959. Pp. 167-176.

Instruction and the Conditions of Learning

ROBERT M. GAGNE

INSTRUCTION OF THE YOUNG, AS WELL AS OF THE UNINFORMED OF ANY age, has been considered a worthy enterprise throughout history. It has usually been acknowledged to be a highly complex human activity. For the most part, it must be carefully planned and executed in order to accomplish its objective, which is to bring about learning in another individual. Typically, it requires an intricate web of communication with a learner, and this communication is calculated rather than casual.

The planning and design of instruction nowadays are done both in a long-range and a short-range fashion. Long-range design, for instruction that extends beyond about one day, is done by a number of different agents, including the curriculum designer, the faculty committee, and the textbook writer. Short-range instructional design is usually accomplished by a teacher in the development of a course outline, a lesson plan, or a set of notes.

A specific kind of instructional design, often thought of as the particular province of teacher activity, may be called *extemporaneous* design. It occurs when the teacher decides upon each new communication as a result of what has immediately gone before within the give and take of a classroom, seminar, or tutoring session.

In a broad sense, then, there are many possible agents who may be involved in the design of instruction. The teacher is not the only designer of instruction simply because he is also its practitioner. The principles of design must be highly similar, if not identical, whether instruction is short-range or long-range.

INSTRUCTION, SOME CONTEMPORARY VIEWPOINTS, edited by L. Siegel (San Francisco. Chandler) 1967, pp. 291-313.

There are undoubtedly some other principles pertaining to the execution of instruction, different from those relating to its design. The activity of the teacher, naturally enough, tends to carry out extemporaneous design and also the execution of instruction in rapid succession, so that it is sometimes difficult to make a definite distinction between design and execution. In this chapter, I deal almost entirely with the *design* of instruction, and accordingly with only that part of the teacher's performance pertaining to such design. The main interest is in how instruction may be designed, whether on a short-term or a long-term basis and whether by teachers or by other agents.

Classes of Variables in Learning

The purpose of instruction is to bring about a change within the learner, a change of the sort called *learning*. At time t, it may be observed that the learner is incapable of performing task X. A period of instruction is then instituted. Then, at time $t + 1$, it is noted that the learner is now able to perform task X and that he is also able to perform the same task after an additional interval, say, at time $t + 2$. Basically, these are the observations that lead to the inference that learning has occurred. A change in performance has occurred and is retained over a period of time. The inference is that some internal capability of the individual learner has been altered by the instruction which was given, and it is this change in capability that is called learning.

What are the variables that enter into the situation and affect the phenomenon of learning? There are a number of variables; some of them are appropriately a part of what is called instruction, while some are not. It will be useful at the outset to draw some distinctions among these variables.

Variables within the Learner

This category includes the learner variables *initial capabilities* and *motivation*. If the learner is to be instructed in task X, it is assumed for purposes of this example that this particular capability is not initially present. But there may be other capabilities which are also not present, and if they are not, instruction specifically designed to teach task X will not work. Certain innately determined capabilities may be absent, as in a mentally deficient individual, or certain capabilities ordinarily dependent upon previous learning may be missing. For example, instruction in the solution of problems involving vector resolution of forces may be impossible if representation of trigonometric ratios has not previously been learned. To take a simpler example, instruction in pronouncing printed letter combinations will not work if the child has not previously learned to distinguish the printed letters one from another. The learner's initial capabilities constitute some of the most important variables in the process of learning.

Motivation is another kind of state internal to the learner and prerequisite to effective instruction. Motivational states are usually considered to be alterable by different means than are initial capabilities. However established, it seems fairly clear that a certain kind of motivational state must be present as a precondition for learning (Gagné and Bolles, 1959). It is probably a mistake to think of the necessary state as "motivation for learning." Instead, the essential motivation is something more like "willingness to enter into the learning situation." Obviously, if an individual is determined not to respond to a learning situation, but to escape from it physically or otherwise, instruction cannot be effective. If a child turns his attention from the learning situation, instruction will not operate to bring about a change in his capabilities.

Another motivational state necessary for learning is called *alertness*. According to recent evidence (Lindsley, 1957), the receipt of stimulation by higher levels of the central nervous system is a specific brain function. If this alerting system is functioning in a particular way, environmental stimuli will reach the higher centers; if it is functioning in another way, they will not. Therefore, this system must also be operating properly in order for instruction to affect the individual.

Variables in the Learning Situation

The second broad class of variables affecting learning consists of those located outside the learner and within his immediate environment. Together these variables constitute what is usually called the learning situation. A number of different variables may be distinguished.

There are stimuli which direct the learner's *attention* and thus determine which aspects of the total stimulus situation will be received. Almost any form of stimulation is possible as a director of attention. Initially, any sudden change in stimulation commands attention. But the learner may also have acquired certain observational habits, and because of these his attention may be directed by moderate or even faint stimuli. The first-grade teacher may initially direct her pupils' attention by clapping her hands quite smartly; but the pupils rather quickly learn to pay the same attention to a much less intense handclap or even to a spoken word.

The stimuli which are to be involved in the *performance* being learned constitute another part of the total learning situation. These specific stimuli are of course only a portion of the total stimulation of the learning situation, but they are essential ones. For a science student learning the concept of vaporization, the specific stimulus is a liquid in the process of vaporizing. For a child learning to read, the essential stimuli are printed words. As the individual develops in language ability, it becomes increasingly feasible simply to represent the specific stimuli by means of printed or spoken words. To an adult who "knows the words," it is possible to represent the stimulus of a telephone pole in the verbally stated proposition "Telephone poles are at least one foot in diameter." Even for adults, however, many learning situations need to include the essential stimuli themselves.

Verbal communications which stimulate learner responses constitute a fairly large subclass of variables within the learning situation. In general, these are commands, suggestions, or other statements to which the learner is presumed able to respond correctly. For example, a statement beginning "We have already seen that . . ." suggests the action of recall on the part of the learner. "Now, we shall consider . . ." is another kind of verbal communication in this general class. "See if you can follow . . . ," "Now notice that . . . ," and "It should be possible to find out whether . . ." are still other examples. Very often, such verbal communications are made in the form of questions, like "Can you show that these two statements are equivalent?" In a later section, the attempt is made to classify the functions of verbal communications more precisely. For the moment, it is important to note the following two things about them. (1) They function to *direct the behavior* of the learner within the learning situation, perhaps they might be called "verbal directions" for this reason. (2) They actually *convey no content* to the learner pertaining to the capability to be learned; instead, they are independent of content.

The presentation of stimuli and verbal directions within a learning situation is subject to variation in *sequence*. Thus, the essential stimuli may be presented first, they may follow one or more verbal directions, or they may remain throughout the learning event. The sequence of questions constituting the verbal directions may also be varied. A question which generates recall of some previously learned capability may come first, or it may be preceded by a statement informing the learner about the objective of the learning. Other variations are possible. Sequence of verbal communication is likely to have some demonstrable importance in most kinds of learning situations, derived in large part from the effects of *contiguity* of certain events within the total act of learning (see Underwood, 1952).

Instruction and Learning

There are, then, two broad classes of variables that influence learning. those within the learner, and those in the learning situation. These sets of variables undoubtedly have interactive effects upon learning, as many writers have emphasized. The external variables cannot exert their effects without the presence in the learner of certain states derived from motivation and prior learning and development. Nor can the internal capabilities of themselves generate learning without the stimulation provided by external events. As a problem for research, the learning problem is one of finding the necessary relationships which must obtain among internal and external variables in order for a change in capability to take place.

What is instruction? Instruction may be thought of as the institution and arrangement of the *external* conditions of learning in ways which will optimally interact with the internal capabilities of the learner, so as to bring about a change in these capabilities. Instruction thus deals with the manipulation of the conditions of the learning situation—with commanding attention, with presenting essential stimuli, and with the nature and sequence of verbal directions given to the learner. *The function of instruction is the control of the external conditions of the learning situation.*

Varieties of Change Called Learning

The determination of the nature and sequence of the external conditions of learning is surely the purpose of the great bulk of learning research over a period of many years. Over much of this period of time, too, there has been a guiding assumption that the nature of the change called learning must be in some fundamental sense the same, regardless of what is being learned. Accordingly, for a great many years, theories about the optimal conditions for learning have been dominated by concern with the variables of *contiguity*, *reinforcement*, and *frequency*. Investigators have searched for certain *general laws* relating these obviously important variables to learning outcomes, independently of "what is being learned," that is, of the nature of the change in capability being studied.

Although the verification of general laws is surely a desirable objective, the assumption that the *kind of change* in capability being studied is always somehow "the same" may be unjustified. How much similarity is there, actually, between the kind of change represented by a child learning to say his first word and that represented by a more experienced child learning to read printed English sentences? Or between learning to distinguish triangles from rectangles and learning to demonstrate that the sum of the internal angles of a triangle is the same as a straight angle? How much similarity is there between the learning of new "facts" by a beginning chemistry student from a textbook and the learning of new "facts" by his chemistry professor from a technical journal? All of these are surely examples of learning; that is, they involve a change in capability which can be inferred from a before-and-after comparison of performance. But are they the same kind of change?

Despite the prevailing emphasis on fundamental similarities of process in various learning situations, investigators of learning have always recognized certain "types" of learning. There is "trial-and-error learning," "discrimination learning," "paired-associates learning," "concept learning," "conditioned-response learning," and so on (see, for example, Melton, 1964). But these varieties of learning have tended to be identified with certain kinds of stimulus situations generated by particular equipment or materials, like the bar-pressing apparatus, the memory drum with verbal syllables, or the maze with choice points. The tendency has *not* been for these types of learning to be distinguished in terms of the *kind of change in capability* they imply. The questions that have tended *not* to be answered are of this kind. What sort of capability does the learner have (that he didn't possess before) when he has completed learning a list of verbal paired associates? Or when he has acquired a distinction between responses appropriate to a white card and to a black card? Or when he has learned to identify "food words" and "flower words"?

Differences in Performance Outcomes

An examination of the performances which reflect the outcomes of learning has led me to the conclusion that there are a number of varieties of these activities (Gagné, 1964, 1965). These kinds of performances may be called:

1. Specific responding.
2. Chaining (motor and verbal).
3. Multiple discrimination.
4. Classifying.
5. Rule using.
6. Problem solving.

Each differs from the others, in the sense that it is possible to distinguish the kinds of things that the learner can and cannot do in each instance. The capabilities underlying these performances form a partially ordered set. the acquisition of a more complex capability requires the previous existence of a simpler one, whereas the possession of a simpler capability does not imply that the individual can exhibit a more complex one.

For example, the performance known as rule using (or principle using) implies that the individual can also classify the terms which make up the rule, otherwise, he would not have been able to learn the rule. On the contrary, the individual who can classify the terms contained in a rule does not necessarily know the rule, that is, he cannot necessarily show that he can use it. At one point in his education, a student may know what the concept "borine" means (as a member of a class of chemical elements), what "gas" means, and what "room temperature" means. Yet he may not have acquired the principle (rule) reflected in the statement, "Borine is a gas at room temperature." If he already knows how to classify these terms, learning the principle is easy, if he does not know the terms, acquiring the principle cannot take place until he does.

Differences in What Is Learned

The existence of differentiable performances as outcomes of learning naturally leads to the inference that different kinds of capabilities are established by learning. The neural mechanisms determining these capabilities are presumably different for each one, but this possibility can only be guessed at with present evidence. However, the capabilities inferred from the performances do have different names, more or less familiar to psychologists. It is important to the present discussion to draw a careful distinction between the *performances* made possible by learning and the *capabilities* inferred as underlying these performances.

Table 10-1 contains a list of distinguishable kinds of performance which may be observed as learning outcomes. The second column defines each of these performances in terms which imply the criteria of distinction. The third column contains an example of each performance. The final column provides the name most commonly used to identify the inferred capability corresponding to each performance. It may be noted that, in certain instances, the commonly used name for the type of performance is at least partially the same as that used for the capability. This is unfortunate, but changing language customs cannot be attempted within the scope of this chapter. Successive rows of the table, reading down, refer to increasingly complex kinds of performance, and the more complex performances are hypothesized to imply the preexistence of less complex capabilities (that is, classifying implies that discrimination capability already exists in the individual, and so on).

TABLE 10-1. TYPES OF PERFORMANCE WHICH ARE OUTCOMES OF LEARNING, WITH DEFINITIONS, EXAMPLES, AND CORRESPONDING INFERRED CAPABILITIES.

Performance type	*Definition*	*Example*	*Inferred capability*
Specific responding [1]	Making a specific response to a specified stimulus	Child saying "doll" when mother says "doll"	Connection
Chaining:			
Motor	Exhibiting a chain of responses each member of which is linked to each subsequent member	Unlocking a door with a key	Chain
Verbal	Exhibiting a chain of verbal responses linked by implicit codes	Giving French equivalents of English words; saying "A stitch in time saves nine"	Verbal association; verbal sequence
Multiple discrimination	Making different (chained) responses to two or more physically different stimuli	Naming a specific set of object colors	Discrimination
Classifying	Assigning objects of different physical appearance to classes of like function	Distinguishing various objects as "plant" or "animal"	Concept
Rule using	Performing an action in conformity with a rule represented by a statement containing terms which are concepts	Placing *i* before *e* except after *c* in spelling various English words	Principle (or rule)
Problem solving	Solving a novel problem by combining rules	Raising an automobile without using a jack	Principles plus "problem-solving ability"

[1] In commonly accepted terminology, this is the *instrumental response.*

The account of various kinds of *performance* has been given because it seems highly probable that there are corresponding varieties of *performance change*. In other words, the identification of these different kinds of performance, together with the different kinds of capability they imply, suggests that there may be at least as many different kinds of learning. And if this is so, it may be supposed that there exist an equal number of *conditions of effective learning* to correspond with each variety. A theory of instruction, then, cannot be maximally useful if it concerns itself with only those conditions that are general to all classes of learning. Instead, such a theory must concern itself in an individual manner with each of the types of learning.

Kinds of Learning

The idea that there are as many as seven kinds of learning is not entirely a novel one. A number of writers (Woodworth, 1958; Tolman, 1949, Mowrer, 1960) have emphasized the importance of giving separate consideration to several learning varieties. Skinner (1938, 1957) also describes different conditions for establishing connections (operants), chains, discriminations, concepts, and several types of verbal sequences, without necessarily using these names for them. Actually, the kinds of performance listed in Table 10-1 are distinguished by most educational psychologists when they find themselves in the position of being required to survey the range of naturally occurring events called learning.

These seven varieties of learning may be established by different sorts of conditions for learning, that is, by different kinds of *instruction*. Although the requisite conditions for the simpler types are fairly well known, those for the more complex types are not well known and have not yet been thoroughly established by careful experimentation. It is possible, however, to describe what does appear to be known about the conditions for bringing about these varieties of performance, based upon existing experimental evidence liberally supplemented by ordinary observation. A summary of the learning conditions appropriate to each of these varieties is given in Table 10-2 (see Gagné, 1965). The table indicates, for each kind of performance learned, the internal

TABLE 10-2. SUMMARY OF CONDITIONS CONSIDERED NECESSARY FOR SEVEN KINDS OF LEARNING.

Performance established by learning	*Internal (learner) conditions*	*External conditions*
Specific responding	Certain learned and innate capabilities	Presentation of stimulus under conditions commanding *attention*, occurrence of a response *contiguous* in time; *reinforcement*
Chaining:		
Motor	Previously learned individual connections	Presenting a *sequence* of external cues, effecting a sequence of specific responses *contiguous* in time, repetition to achieve selection of response-produced stimuli
Verbal	Previously learned individual connections, including implicit "coding" connections	Presenting a *sequence* of external verbal cues, effecting a sequence of verbal responses *contiguous* in time
Multiple discrimination	Previously learned chains, motor or verbal	Practice providing *contrast of correct and incorrect stimuli*
Classifying	Previously learned multiple discriminations	Reinstating discriminated response chain contiguously with *a variety of stimuli* differing in appearance, but belonging to a single class
Rule using	Previously learned concepts	Using external cues (usually verbal), effecting the recall of previously learned concepts contiguously in a suitable sequence; specific applications of the rule
Problem solving	Previously learned rules	Self-arousal and selection of previously learned rules to effect a novel combination

conditions presumed to be necessary for learning and the external conditions which may be used to bring about the learning. It will be noted that each more complex performance is considered to require simpler capabilities as prerequisites. The external conditions, which should provide direct implications for instruction, vary considerably with the kind of performance the learning is expected to make possible in each case.

Some examples, which may be helpful in the interpretation of Table 10–2, are given in the following sections.

VERBAL CHAINING

Example: Learning that the English words "the foot" are a translation of the printed French words *le pied.*

First, the internal conditions listed in the table tell us that some previously learned individual connections must be recalled. These are (1) saying "foot," a capability which can be assumed; (2) observing the printed word *pied*, which can be checked by asking the learner to pick it out or to say it; and (3) a coding connection such as *pedal*, which will form the sequence *pied*–(*pedal*)–"foot." This last connection may be supplied as part of the instructions, or the learner may be encouraged to supply his own. The internal conditions having been satisfied, the external ones are easy. The stimulus *pied* is presented; verbal cues such as the printed words *pedal* and *foot* are presented at the same time (that is, contiguously), and the learner says the word "foot." Assuming now that the learner knows that this is the kind of performance expected of him, it may be said that the verbal sequence *pied*–"foot" has been learned.

MULTIPLE DISCRIMINATION

Example: Learning a number of different English words for a number of different French words.

Suppose now that the learner is expected to acquire not only *pied*-"foot" but also *main*-"hand," *doigt*-"finger," and perhaps several others as well. It is reasonable to think that each of them can be learned under the same kind of internal and external conditions as was the first. But another difficulty arises if he attempts to learn them all. there will be *interference* among these chains. Unless some additional set of conditions is added, the words will tend to be forgotten almost as fast as they are learned. The kind of condition used is practice. Again and again, these previously learned individual chains are recalled, using external verbal cues as before and differentiating the right from the wrong stimuli to be associated with each response by telling the learner when his response is right or wrong. Of course, the amount of practice necessary will depend on how many of these chains the learner attempts to acquire as a total set. But practice is the major emphasis of the external learning conditions when multiple discriminations are to be acquired.

CLASSIFYING

Example: Learning the botanical classification *tuber* and the verbal classification "tuber."

The table indicates that previously learned discriminations must be recalled. In the example at hand, these discriminations pertain to the physical appearance of plant stems on the one hand and of roots on the other. Following this recall, a *variety of examples* is presented, illustrating the appearance of tubers as enlarged stems. In each case, there is reinstatement of a chain such as *potato*(or pictured potato)-*tuber*-"tuber." In order to emphasize the distinction, a similar set of examples may be presented for *root*.

RULE USING

Example: Learning the rule "Pronouns which are the subjects of sentences are in the nominative case."

According to the table, the first concern is with the recall of previously learned concepts. External conditions must be arranged to supply verbal cues to stimulate such recall. Verbal cues may be in the form of questions, such as "Which of the following are pronouns? Which of these words is the subject?" These concepts having been aroused, it is then a fairly simple matter to use their names as verbal cues for the rule to be learned; in other words, to state the rule: "Pronoun subjects are nominative." But it is also important at this point to insure that such a statement is not learned merely as a verbal sequence. The next step in arranging external conditions, then, is to require a number of specific applications of the rule, such as "Supply the first-person pronoun for the following sentence. '—— students like football games.'" A suitable variety of examples of rule application, of course, needs to be provided.

Functions of the Instructor

The viewpoint presented here is that the form taken by instruction needs to be tailored to the particular objective which represents the kind of performance change to be brought about. Instruction is used to establish the necessary conditions for learning, and instruction differs in accordance with what is to be learned. The kinds of communication required for establishing a verbal chain are quite different from those required for establishing a principle, and they are likewise different for all the other categories of performance change listed in Tables 10–1 and 10–2.

The major implication of this approach to the problem of instruction is surely that the instructor cannot be guided by a simple set of rules that apply to all cases. The important aspects of the instructor's behavior do not lie in the fact that he uses a *general principle* to control learning (such as reinforcement or contiguity), but rather in the fact that he employs *different techniques* for different kinds of learning.

The functions performed by an instructor in helping to bring about these various kinds of change are discussed in the following paragraphs. For convenience, these functions are classified under four headings. (1) functions performed early in the instructional sequence, (2) functions designed to evoke and guide learning, (3) functions encouraging generalization, and (4) functions permitting an evaluation of outcomes.

Performing Early Instructional Functions

PRESENTING THE STIMULUS

This important function for the instructor can be implemented by pictures and other audiovisual aids. Selecting the proper stimulus requires making some good decisions. For example, if motor chains are to be learned, a portion of the necessary stimuli ought to come from muscular movement, presenting these stimuli becomes a matter of permitting repetition of the motor acts. If a child is learning to name a specific object (a verbal chain), the object itself must be presented; it cannot be represented by a word, as may be done with an adult, who has already acquired such a link. If a classification is being learned, a sufficiently great variety of stimulus objects must be presented to represent the class, otherwise, the concept acquired will be inadequate. If a rule is being learned, the verbal statement of the rule is not always a sufficient stimulus, since it may merely lead to learning the verbalization itself. Instead, situations requiring application of the rule must be presented as stimuli.

CONTROLLING ATTENTION

Directing the learner's attention to essential aspects of the stimulus situation is another function the instructor performs in controlling the conditions of learning. This is done in various ways, perhaps by introducing extra stimuli into the situation (arrows, pointers) and often by verbal commands. It is obviously important to all forms of learning that the relevant stimuli be perceived. For example, the attainment of the biological concept *cell* may require that attention be paid to such characteristics as membrane, nucleus, and cytoplasm since these are the major identifications to be recalled.

INFORMING THE LEARNER OF OBJECTIVES

A requirement of instruction which may transcend the conditions of learning previously discussed is that the learner be informed about the nature of the performance expected when learning is finished. Presumably, this procedure establishes a continuing *set* which facilitates learning, perhaps by making possible reinforcement at several points in a rather lengthy sequence of activities (see Gagné, 1965, Chap. 9). Providing information about expected performance may also be justified as a way of establishing a conceptual structure under which ideas to be learned are subsumable (see Ausubel, 1963).

Suppose that a new rule is to be learned, such as "adding the numerators of fractions with identical denominators to obtain a sum." The instructor communicates to the learner what is meant by obtaining a sum and, particularly, what a sum may look like. He might do this, for instance, by saying, "Now you need to learn how to obtain the sum of two fractions having common denominators, such as $\frac{3}{5}$ and $\frac{4}{5}$. The sum of these, you will see, is $\frac{7}{5}$." It appears probable that a sequence of instruction which begins by communicating the objective of learning is more effective than one which does not begin this way. However, additional experimental evidence on this point is needed.

STIMULATING RECALL OF PREVIOUSLY LEARNED CAPABILITIES

I have already indicated that stimulation of recall is a very important part of instruction, applicable to all of the varieties of learning.

Accordingly, it is one of the essential functions of an instructor. As described in Table 10–2, the necessary internal conditions for learning always include some previously learned capability of a simpler sort. Recalling these capabilities is usually an early step in instruction.

Recall is often stimulated by means of verbal communication, although it need not be. A question may be "Do you remember that we learned to distinguish between a stem and a leaf?" In another form, a question of similar intent might be "What are the main parts of a plant above ground?" The first of these questions requires only a recognition response on the part of the learner, while the second requires him to reinstate a previously acquired performance. It is not possible to find a simple rule governing the relative effectiveness of these two forms of question, the important thing is that they both have the function of stimulating *recall*. The necessity for having these previously learned capabilities "lively" in memory provides still another example of the importance of *contiguity* in learning. Not only must external events be contiguous for effective learning, but also recalled memories must be contiguous with these events.

Evoking Performance

Presumably, the activities of presenting the essential stimuli, controlling attention, informing the learner of objectives, and stimulating recall constitute the early events of instruction and need to occur approximately in the order given. The next step, broadly speaking, is one of evoking the required performance that is being learned. Thus, the application of a particular sequence of events to stimulate the learner is another function which may be carried out by the instructor.

DETERMINING A SEQUENCE

There are probably effective and ineffective sequences depending upon the category of performance dealt with. For example, as is well known, a relatively short verbal sequence (such as *wollen Sie bitte*) may readily be learned by being presented all at once, whereas a long one (like the stanza of a poem) may well require a more elaborate sequence of presentation, as is implied by the "progressive-part method" (McGeoch and Irion, 1952). Similarly, 4 objects requiring multiple discrimination can often be effectively presented in a single "trial," whereas 20 objects typically imply the need for a series of interspersed learning and recall trials (Gagné, 1965, Chap. 5). Each of the varieties of learning requires a different decision about a proper sequence of presentation, moreover, the length and nature of the content to be learned are also likely to affect this decision.

The need for prompting as a part of the learning situation is quite frequent for many types of performance change. For example, evoking the correct verbal chain which provides the French word for "horse" requires that the instructor supply not only the stimulus *horse* but also the printed or spoken word *cheval*, in order that the learner for the first time institute the performance of saying "cheval." An additional prompting word or phrase may also be used for the coding response, as when the presentation of *horse*–(*chevalier*)–"cheval" is made to the learner. External verbal cues (prompts), such as the verbal name for a concept, are commonly used in connection with learning of the classifying sort. Thus, the presentation of each of several varieties of cell is often accompanied by the prompting verbal sentence "This is a cell." (Note that the learner in this situation is not learning to say the word "cell"—he already knows that, prompting him to say it in this situation is simply a means to the end of learning the correct classification for several different stimuli.)

A somewhat more elaborate kind of prompting, which may be called *guiding learning*, may be employed by the instructor when relatively complex rules are being learned or when the learner is being encouraged to discover new rules for himself, as in problem solving. In such cases, a sequence of verbal statements or questions may be employed in order to aid the learner's progress through several steps of reasoning to the principle being acquired. Learning to employ a rule and discovering how to combine principles in novel ways are both examples of complex intellectual activity, most of which occurs internally. But the direction of such activity can be influenced by external stimulation, usually in question form, and the course of learning can accordingly be shortened and made more efficient. Guiding learning in this sense is generally considered an important and frequent instructor function.

Encouraging Generalization

An account of instructor functions would be quite incomplete without mention of the activities having the purpose of encouraging the generalization of what is learned. This function is particularly applicable to the learning of concepts and rules and to problem solving. Generalization of the concept is encouraged by the procedure of presenting a suitable variety of stimuli to be classified; if the variety is too constricted, the concept is likely to be inadequate. Similarly, the diversity of examples used in illustrating specific problems to which rules are to be applied can determine the amount of generalization to be expected as a result of the learning. Actually, a variety of specific techniques may be employed by the instructor to encourage generalization (or transfer of learning, as it is sometimes called). In general, these techniques stimulate the learner to apply his recently acquired capabilities within widely differing contexts.

Assessing Outcomes

Naturally enough, the instructor must be concerned with assessing the outcomes of learning. Regardless of the type of performance, there are differential criteria to indicate whether or not learning has actually taken place. These are suggested, although not specifically described, by the final column of Table 10–2. For example, if multiple discrimination of 15 stimuli has been learned, then the learner should be able to identify each of these stimuli without error when they are presented all together (or one by one, for that matter). If a concept has been acquired, the presentation of an example not previously encountered should result in its being correctly classified. If a principle has been learned, an assessment of the learning should be possible by requiring the learner to apply it to a problem or situation novel to him.

Assessing learning outcomes may readily be seen to be a very frequent requirement of the instructional process. Each new kind of learned performance ought to be assessed in some manner or other, and the results made known to the learner. Such testing is often done informally by the teacher. Other kinds of testing of a more formal sort may, of course, also be needed at periodic intervals.

It may readily be realized from the preceding discussion that instruction is not a simple task. It cannot be reduced to simple formulas, such as "applying reinforcement," "communicating with the learner," or "stimulating discovery." Although all of these functions may be involved in any given instance, they are not the sole requirements for learning.

Controlling attention, presenting the essential stimuli, guiding learning, encouraging generalization, assessing outcomes—any or all of these functions may constitute the task of the instructor for any given act of learning. In making decisions about each one of them, the instructor must be guided by three principles of primary importance. (1) What is sought is a change in capability, represented by a change in performance. (2) One of the major determinants of this change is the initial capability of the learner, originating in turn with previous learning (see Glaser, 1962). (3) Different kinds of change (learnings) have different instructional requirements.

Viewed as a whole, the task of instruction is a tremendously complex one and therefore likely to be quite difficult. The skilled practitioner of this art, who is able to select and manipulate the variables comprising the external conditions of learning, of whatever type, is surely worthy of admiration. There is reason to wonder whether there are very many people who can really do a good job of instruction, particularly if they have been able to practice it for only a few years. Might it be possible to raise the standards of achievement in this field by somehow reducing the complexity of the instructional task?

Certain practical measures can be taken to simplify the instructional task for a teacher. Audiovisual aids are perhaps a *partial* answer to this problem. They are, for example, excellent for such functions as controlling attention and presenting the essential stimuli (see Gagné, 1965, Chap. 10). A more comprehensive approach to the problem, however, is provided by the notion of *predesigning* instruction, so that the entire set of decisions which must be made by an instructor is reduced in number and simplified. Informing the learner of objectives, stimulating recall, determining a sequence of presentation, guiding and prompting, and assessing outcomes are all functions which can be predesigned in order to generate effective learning conditions. As they are currently constructed, textbooks make no real attempt to accomplish predesign, and neither do educational films and television programs. Teaching-machine programs do make an attempt at predesigning, and some of them are quite successful at it. If their purposes and potentialities can be viewed with sufficient breadth, perhaps these programs will show the way to the accomplishment of such a goal.

Instruction and the Control of Behavior

It is interesting to consider the question of just what instruction is doing, as far as the behavior of the learner is concerned. How can the effects of instruction be conceived, in a general sense? Does instruction stimulate, communicate, or restrict behavior (for example, in its prompting and guiding function)? Does it exert a direct control over the process of learning?

One relatively prevalent conception of how instruction works is that it creates a learning situation which in a sense "captures control" of the nervous system of the individual so that he inevitably learns. Provided the external conditions of learning are optimally chosen and provided they are made consistent with the individual's previously learned capabilities (that is, the internal conditions), the learning situation may be conceived as one which "takes hold" of the learner and brings about the desired learning as a consequence. Actually, this is not a bad way of conceiving of what should be true of the learning situation. In an experimental study of learning, when both internal and external conditions are independently measured and controlled, the experimenter may well gain the impression that he has an unusual degree of control over what is happening.

However, such a view of the effects of instruction on learning has another side to it, which may be at least equally illuminating of the process of instruction. Regardless of how much "control" is exerted by external conditions, the internal processes of the learner are likely to make a crucial contribution to the events of learning. Furthermore, this crucial internal process contributed by the learner's nervous system is, so far as is now known, highly idiosyncratic. At present, it cannot be predicted or even described adequately, and it can be "controlled" only in a probabilistic sense.

Even the simpler kinds of learning described in Table 10-1 are greatly affected by idiosyncratic internal events. Motor chains require practice, it has long been believed, because the learner must "discover" the proper kinesthetic stimuli which are controlling a smoothly organized response. The training of a motor skill, like driving a golf ball, is often frustrating to both learner and instructor because these crucial internal kinesthetic stimuli cannot be directly controlled, they must be courted during periods of practice. A number of studies of verbal associates have shown that the patterns of associations previously acquired by the individual enter into the establishment of verbal chains as crucial mediating links (Bousfield, 1961; Jenkins, 1963). A study of French-English word association (referred to by Gagné, 1965, p. 100) found that the most effective learning of such associates results from a condition in which learners were encouraged to supply their own "codes," rather than having codes suggested in pictures or words.

In a consideration of the more complex varieties of learning, the contribution of events internal to the learner becomes even more prominent. Seven-year-old children can learn to "choose the opposite" card of a black-white or large-small set very quickly and presumably by means of a mediating concept; whereas four-year-olds learn such a reversal task slowly, by "trial and error" (Kendler and Kendler, 1961). But whatever set of internal events is represented by "choosing the opposite," it is doubtful that it is a matter of using these words, and likely that the events are highly idiosyncratic. A number of studies of concept learning have verified Hull's (1920) finding that conceptually based performance can be exhibited at the same time that the learner cannot "name" the concept or otherwise describe it.

The individualistic nature of internal events is perhaps even more strongly apparent in rule using and problem solving. Katona (1940) found that a demonstration illustrating the effects of rule application was often more effective in solving matchstick problems than was the supplying of a verbalized "arithmetic rule." A study requiring learners to verbalize while learning to solve a problem (Gagné and Smith, 1962) was unsuccessful in turning up consistent patterns of verbalization, again the suggestion is that the processes contributed by the individual are highly idiosyncratic. Many students of problem-solving behavior would probably take this idea for granted, a judgment based in part upon outcomes of earlier studies which have attempted to get the learner to "think aloud" while solving a problem (Woodworth, 1938, pp. 768–800). In addition to the specific rules which may be involved in solving a problem, some writers, notably Bruner (1961), have emphasized the importance of approaches to problem solving, called *strategies*. Although such self-directed principles may perhaps be categorized (Bruner, Goodnow, and Austin, 1956), there seems little doubt that they too are unique to the individual.

From its simple to its complex forms, then, an act of learning appears to be characterized by some internal events which have so far resisted precise description and analysis. There are some uniquely individual processes contributed by the learner's nervous system. Furthermore, these internal events are not properly viewed as artifacts of the learning process, but instead must be considered as essential components, without which no learning can occur. It is noteworthy that even the learning of a connection (see Table 10–1) is considered by many modern theorists to require a contribution of internal "feedback stimulation" for its establishment (Mowrer, 1960, Chap. 7). (This discussion of learning types does not include the classical conditioned response, which may be even simpler than the connection, or instrumental response, with which Table 10–1 begins.)

It seems evident, then, that in controlling all of the aspects of external learning conditions described previously, instruction nevertheless can only make the occurrence of this crucial, internal, idiosyncratic event more probable. So far as is now known, instruction cannot directly control this internal event. The careful design of instruction can surely increase its probability and, by so doing, can make the entire process of learning more sure, more predictable, and more efficient. But the individual nervous system must still make its own individual contribution. The nature of that contribution is, of course, what defines the need for the study of individual differences.

Conclusion

The practical implications of this discussion suggest a goal for instruction which seems quite consistent with what is known of the educative process as a totality. One of the major goals of instruction might well be to bring about progressively the elimination of the need for instruction directed at the individual. Doesn't this suggest, after all, the major difference between a fourth-grader and a graduate student? The former needs all the help he can get from the external conditions established by properly designed instruction. The latter needs almost none of this help because he performs the various functions himself. In other words, to a very large extent he carries out self-instruction. What he does when he "studies" is first to analyze and organize the content to be learned in such a way that it can in fact be readily learned.

It may be that additional emphasis needs to be given, in a very broad and comprehensive sense, to the notion of *self-instruction*, not simply as an immediately applicable technique, but as a goal. This may be the direction of practical development which could best exploit the unique internal contribution of the individual learner.

MASTERY LEARNING AND MASTERY TESTING

Samuel T. Mayo

Mastery learning is not a new idea but it has not always gone under that name. The word "mastery" is very common in educational parlance. It connotes having learned something well as promised in the adage, "Practice makes perfect." Mastery usually comes easily when there is a very limited skill or parcel of knowledge to be learned and one has the opportunity for abundant practice. Moreover, with mastery comes a feeling of pleasure and self-confidence from a job well done.

In the study of human learning, educational psychologists long ago discovered two important principles: (1) Given meaningfulness, learning is retained easily when there is abundant practice; and (2) Meaningful learning is easily transferred. "Meaningful" here means bearing a relationship to previous learning. It also implies that the goals to be obtained are obvious. Transfer, in essence, means that one is able to use previous learning by applying it to solution of problems or decision making.

Until ten or fifteen years ago prevailing practices of instruction and evaluation of instruction promoted unsound effects on learners. Individual differences were often neglected in "lockstep in struction." In a sense, instructional time was held constant while the amount of material learned varied. The normal curve was being overused and misused in evaluation processes. Mastery learning differs in that, in a sense, material is held constant, while study time is allowed to vary.

The picture has been changing rapidly. Entire curriculums, particularly in mathematics and physics, have been revolutionized. Programmed learning has had impact on almost every school system. Independent study and individualized instruction, either with or without machines, are

NCME - MEASUREMENT IN EDUCATION, March, 1970, Vol. 1, No. 3.

being tried in more and more schools every year. Such innovations in the learning environment call for innovation in testing, measuring, and evaluating techniques. Useful references for further reading would be those by Bloom (1968), Bruner (1960), and Carroll (1963).

Until recently, most people were convinced that mental abilities were somehow tied very tightly to academic achievement. This point of view was valid under traditional instruction when one constructed achievement tests in such a way as to assure a normal distribution of the scores. It should be noted that mental abilities, as measured, have also been distributed like a normal curve.

The expectation of the distribution of achievement for traditional treatment is a normal curve, or something similar to Curve A of Figure 1, with central tendency in the midregion. By contrast, mastery treatment is expected to shift central tendency closer to perfection and probably to show skewne as well, as in Curves B and C of Figure 1.

In programmed instruction questions in the frames are made so easy that the resulting error rate is only 5 to 10 percent. For such cases a different kind of proficiency test is called for -- a test for mastery in which the score distribution is very different from that of the normal curve. Instead of being piled up in the middle and tapering gently in each direction, scores are bunched at the high end of the scale The test items for the new type of proficiency test obviously must be of a different type from those which have been recommended for traditional achievement tests. Mastery test items are tied intimately to the stated objectives of instr ction which are specific for a relatively short period of time in a school course (even for a day's lesson). But programmed learning is only one of several innovations in what may be called mastery learning.

THE MASTERY MODEL

The term "mastery model" is used here to summarize the thinking of some educators who have tried out certain innovations to improve school learning.

Rather than thinking of aptitude as a kind of ceiling. Carroll (1963) suggested that aptitude may be related to the amount of time necessary to achieve mastery. Bloom (1968) feels that if students are normally distributed with respect to aptitude and if the kind and quality of instruction and the amount of time available for learning are made appropriate to the characteristics and needs of each student, a large majority of the students can be expected to achieve mastery.

Briefly characterized, the model calls for such strategies as these:

1. Inform students about course expectations, even lesson expectations or unit expectations, so that they view learning as a cooperative rather than as a competitive enterprise.
2. Set standards of mastery in advance; use prevailing standards or set new ones and assign grades in terms of performance rather than relative ranking
3. Use short diagnostic progress tests for each unit of instruction.
4. Prescribe additional learning for those who do not demonstrate initial mastery.
5. Attempt to provide additional time for learning for those persons who seem to need it.

TEST DEVELOPMENT -- A MASTERY APPROACH

Test experts and authors of textbooks on tests and measurement have been telling teachers for many years to construct their achievement tests in traditional ways. What are these ways? The answer to this question may be expressed as a series of steps: (1) Define the objectives to be measured; (2) write items to sample content and behavior domains of the objectives; (3) adjust item characteristics with average item difficulty around 50%-60% and maximum discrimination against the internal criterion of the total test scores; (4) interpret performance against a norm (i.e., peer) group.

Under the mastery model, the first two steps would remain somewhat as they were. The third step would be replaced by eliminating the necessity of discriminating power; that is, average difficulty would be shifted to possibly 85% or higher.

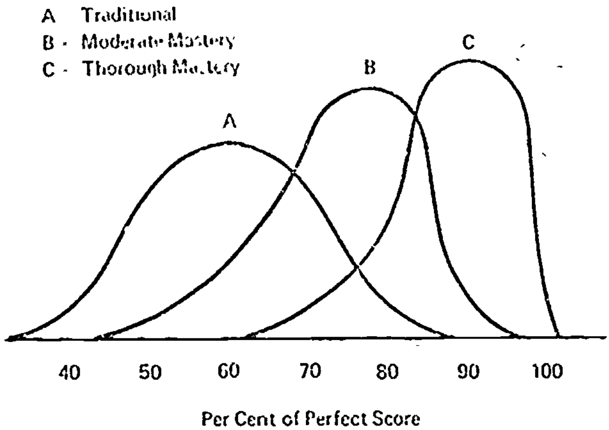


Figure 1. Distributions of Achievement Test Scores Under Traditional Treatment and Mastery Treatment (Hypothetical)

The kind of instruction under the model also differs from traditional, and this affects the kind of testing required. In mastery it is assumed that the response that one learns (such as giving the answer to a question or making an overt muscular response) can be made confidently upon cue. Tests constructed to measure such learning may appear easy to one accustomed to very difficult tests, especially those written by persons prone to use "trick" questions or obscure language. Mastery tests may be conceived as operating on a "go-no-go" basis. Most students are expected to pass an item. The few who fail the item show a clear deficit, and this feedback indicates need for additional remedial learning sessions and repeated testing until items are passed.

Clear Objectives

Objectives are made clear to the student under mastery learning. We have had years of experience under the alternative in which objectives have not been made clear and in which achievement was on a sink-or-swim basis. Under a mastery approach, one plan which can be used effectively is a periodic (even daily) sheet which spells out the immediate objectives. The student has the feeling of clearly understanding the goals and knowing that they may be reached easily in a short time.

Mastery learning and mastery testing seem to promise the elimination of some of the fear which plagues many students in testing situations, especially in quantitative courses such as those in mathematics, science, and statistics. Steps that teachers can take toward overcoming fears include (1) announcing that daily quizzes will not be counted on the final grade but rather be used for diagnosis; (Such quizzes can easily be given back for students to keep in their files) (2) announcing that major tests (such as midsemester) will be repeated with an alternate form, counting the higher of the two grades; (3) holding a lengthy, comprehensive review session for the final examination in which the information fed to students is as close to examination content as feasible and which is perceived by students as the limited set of content topics and behavior which will actually be included on the examination.

Criterion - Referenced

Interpretation of test results from mastery tests differs in kind from interpretation of "traditional" tests showing a normal curve distribution of total scores. Interpretation here is *relative*, that is in relation to peers. Recently, the term "norm referenced" has been applied to this kind of test. Under mastery theory a test score in a sense may be considered *absolute*, since one need not compare a score with a peer. One only judges whether a sufficient number of items have been passed to give evidence of mastery of some limited segment of an entire score. In contrast to "norm referenced tests," the term "criterion referenced tests" has been applied to those used in mastery learning. With criterion-referencing a new operation is brought into the picture one which may have far reaching social implications. This will be the interpretation of a test score in terms of describing the specific behaviors which a student can now perform. Thus, it will be much easier to match such a repertory of skills to a forthcoming job or training situation than before when we only knew how a student ranked with his peers but not what he could do.

Frequent quizzes may be used effectively to identify aspects of a course where revision is desirable and to improve the course while it still is fluid. This is an example of the trend toward adapting the course to the student rather than adapting the student to the course.

It should be recognized that while mastery learning theory promises much in educational improvement, very little definite research has yet been done. Therefore, we know little, as yet, about its applicability in education. Meanwhile, it appears worth trying on the part of individual teachers as more persons recognize the weaknesses of traditional instructional and testing practices and try to improve them in accordance with newer methods.

SOME SUGGESTIONS

Teachers are invited to put into practice the theory briefly described in this paper. Each teacher's situation is unique, and there is no guarantee that the same innovation will work everywhere. Run a short-term study of an actual innovation and observe its effect upon teachers, pupils and administrators. Several suggested innovations are:

1. *Give alternate forms of a quiz or examination until students improve to a predetermined level of mastery.* Since the knowledge and skills represented in the test will already have been judged to be important, missing particular items will pinpoint the kind of remedial instruction necessary to bring a student to the desired level.

2. *Involve students as representatives on a committee along with faculty to review a curriculum and to set objectives.* Perhaps the best candidates for such student representatives would be juniors or seniors in high school to set objectives for the freshman cl s.

3. *Have a student committee make up a final examination in order to show what mastery is required of all students.* Obviously, to do this would drastically change the security precautions usually taken and the grading system employed. However, this would be "experimental," and perhaps it would be found that more benefits would be obtained than the disadvantages associated with security and grading.

4. *Have a student-faculty committee critically review the grading system.* There has never

been a perfect grading system. Even some so called "new" systems seem unsatisfactory.

5. *If your school has a new program in individualized instruction and/or independent study, you may wish to strengthen the evaluation you have planned for it.* Your own research director or a consultant from a local university can be invaluable here. Several references can be of help, also. Among these are Gleason (1967), Webb (1966), Bloom (1956), and Krathwohl (1968).

6. *Run a full-fledged experiment in mastery learning and teaching with the help of a nearby university.* Most universities would welcome such an invitation from a local school system. Many of them have graduate students who are looking for an agency in which they may spend time on an internship or on a course project or thesis.

REFERENCES

Bloom, B. S. Learning for mastery. *Evaluation Comment,* 1968, 1 (No. 2).

Bloom, B. S. *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain.* New York: David McKay Company, Inc., 1956.

Bruner, J.S *The Process of Education.* Cambridge: Harvard University Press, 1960.

Carroll, J. B. A model of school learning. *Teachers College Record,* 1963, 64, 723-33.

Gleason, G. T. *The Theory and Nature of Independent Learning.* Scranton, Penn.: International Textbook Company, 1967.

Krathwohl, D. R. *Taxonomy of Educational Objectives, Handbook: II. Affective Domain.* New York: David McKay Company, Inc., 1968.

Webb, E. J. Campbell, D. T., Schwartz, D., Sechrest, L. *Unobtrusive Measures: Nonreactive Research in the Social Sciences.* Chicago: Rand McNally & Company 1966.

Student Evaluation and Promotion of Learning

James H. Block

A critical problem haunting educators throughout the world is how to effectively, efficiently, and economically increase all students' learning. Nowhere is this problem felt more keenly than in the public schools of major cities. In this chapter, several methods are proposed whereby student evaluation techniques might be used to help solve this problem.

Impairment of Learning

Traditionally, student evaluation has been conceived as a process for judging changes relative to a set of educational objectives as the result of some instruction (Tyler, 1950). While in recent years some evaluators (for instance, Bloom, Hastings, and Madaus, 1971; Dressel, 1965) have expanded this conception considerably, many

RETHINKING URBAN EDUCATION, edited by H. Walberg and A. Kopan (San Francisco. Jossey-Bass) 1972, pp. 61 76.

public school personnel continue to view student evaluation as a *judgmental* process for identifying and certifying performance differences among learners.

Data documenting the negative effects of this view on student learning, especially urban student learning, are amazingly sparse. From the anecdotal (for example, Black, 1963) and empirical data available, however, it is possible to outline at least three ways in which this narrow conception can contribute to poor pupil learning.

First, this view leads to the use of student evaluation primarily for grading purposes. Unfortunately, present grading practices doom most students to receive average or below-average marks regardless of how well they might learn. For example, the use of a "normal curve" ensures that roughly 70 per cent of the students receive a C or lower grade. Since grades are one of the major types of evidence by which a student gauges his academic and general personal competence (Torshen, 1969), many students are bombarded with negative indications of their worth. These students eventually develop negative interest in, attitudes toward, and anxiety about further learning (Bloom, 1971; Sarason et al., 1960). Some develop negative academic and even negative general self-concepts (Purkey, 1970; Torshen, 1969), and a significant number acquire mental health problems (Stringer and Glidewell, 1967; Torshen, 1969). These affective traits, in turn, encourage the students to cease striving for learning excellence (see Stanwyck and Felker, 1971) and even undermine their capacity to learn the things they are capable of learning (Feather, 1966; Purkey, 1970).

Second, this conception of student evaluation fosters the predominant use of "summative, norm-referenced" evaluation instruments in the classroom. These instruments are *summative* in that they are administered at the end of a long instructional segment (several textbook chapters, a quarter, a term, or a course) to sum up each student's learning progress (see Bloom, Hastings, and Madaus, 1971). They are *norm-referenced* in that each student's progress is determined by comparing his score with the scores of his classmates or some other norm group (Glaser and Nitko, 1971).

The use of these instruments can contribute to poor learning for two reasons. First, they typically are employed for grading purposes and thus can provide the negative feedback that eventually extinguishes some pupils' desire to learn and corrodes their learning capacity. Second, they usually provide too little information too late to help a student surmount past learning difficulties (Airasian, 1969). Specifically, these instruments ordinarily do not consistently identify *what* objectives he failed to learn or *where* in the segment he encountered difficulties. Hence, they usually supply few specific cues to *why* he failed to attain certain objectives. Without answers to the questions of what, where, and why, the teacher cannot reliably design his instruction to best facilitate a student's future learning, nor can the student best organize his subsequent study and review. Even in those rare cases where the instruments do provide the necessary information, they do not encourage the student to complete prior learning before attempting new. As Airasian (1971) points out: "Mastering certain points missed in a summative evaluation may give a student satisfaction, but it does not raise his grade" (p. 79).

Third, the traditional view of student evaluation encourages teachers and other school personnel, such as counselors (Cicourel and Kitsuse, 1963), to use evaluation results for labeling a student's academic potential. Examples of common labels are "smart-dumb," "fast-slow," "overachiever-underachiever," "gifted-educably mentally retarded." All such labels carry an implicit set of performance expectations for the pupils to whom they are overtly or covertly applied. Some studies suggest that, especially in the crucial early grades, these expectations may create and perpetuate a self-fulfilling prophecy in student learning (Baker and Crist, 1972). The teacher subconsciously structures his instruction so that each student is constrained to learn roughly in accordance with the teacher's original expectations. Thus, should the teacher attach a label connoting low performance expectations, as is often the case in urban, especially ghetto, education (Clark, 1965; Gottlieb, 1966; Harlem Youth Opportunities, 1964; Wiles, 1970), then he unwittingly lowers the quality of his instruction and thereby produces the poor student learning expected.

Besides influencing the quality of a student's instruction, labels, like grades, can also affect his willingness to learn and his actual learning ability. Through a myriad of subtle and not so subtle means, including the grading process, the teacher will convince some learners to adopt a particular label as reflecting their true academic potential. Henceforth these students are likely to perform in accordance with the expectations they believe to be implicit in their label (Lecky, 1945). Thus, should they adopt a label connoting their inability to learn, they are likely to exhibit both an unwillingness and an inability to learn, even though they possess the requisite abilities.

Promotion of Learning

Why has student evaluation generally been used in such negative and unproductive ways? One explanation is that most public school teachers, counselors, and administrators have had little formal training in evaluation theory and practice (Brim et al., 1965; Goslin, 1967); hence, they are either unfamiliar with alternative uses (such as diagnosing student learning difficulties, assessing learning readiness, and defining and clarifying instructional objectives) or are unskilled in defining and clarifying instructional objectives) or are unskilled in their application. An obvious key to the more salutary utilization of student evaluation in urban education, therefore, would be to provide more and better evaluation training for *all* pre-service and in-service personnel. Whether it occurs in public school-based clinics and workshops or in university and teacher-training courses, this preparation should stress the following. the variety of possible classroom roles for student evaluation; awareness of the positive and negative effects of each role on students, teachers, the school, and even the community (Bloom, 1969), the fallibility of evaluation results and the sources of fallibility. on-the-job training and practice in the application of various techniques and interpretation and use of results (Ebel, 1967). A second possibility would be for each urban school system, perhaps each school, to hire a cadre of full-time consultants with special competence in both testing and teaching (Ebel, 1967).

But increased training or expert aid, even if exceptionally good, will not be enough. Unless it is possible to convince urban school personnel that the primary goal of student evaluation is *not* the judgment of student learning, then some individuals will continue to use evaluation in the ways described above. To emphasize its many alternative classroom uses, therefore, educators must consider new goals for student evaluation. One such goal is the maximal promotion of each student's learning.

Model for Optimal Quality of Classroom Instruction

Let us begin by sketching a model for classroom instruction of optimal quality. The model derives from research on group and tutorial instruction (Moore, 1971; Rosenshine, 1970; Thelen et al., 1968) and from writings on models of learning (for example, Carroll, 1963; Hilgard, 1965) and instruction (for example, Gage, 1964; Wallen and Travers, 1963). This literature suggests that the quality of instruction for any student can be defined by three sets of highly complex variables. (1) the clarity and appropriateness of the instructional *cues;* that is, "the degree to which the presentation, explanation, and ordering of elements of the learning task approach the optimum for a given learner" (Carroll, 1963, as quoted in Bloom, 1968). (2) the amount of active *participation* in and *practice* of the learning allowed; (3) the amount and types of *reinforcements* or rewards provided.

Under ideal instructional conditions, such as one well trained tutor per student, each of these variables is manipulated almost automatically to ensure optimal quality of instruction for each learner. Classroom conditions, however, are often far from ideal—usually a single type of instruction, which employs primarily verbal cues; a limited amount of learning time; and a high student-teacher ratio prevail. Under such circumstances, it is unlikely that a teacher's instruction can be of optimal quality for each student no matter how good the teacher may be. Some students may require nonverbal rather than verbal cues; more time to practice; more chance to interact with the teacher; or more than a smile, verbal praise, or the teacher's attention. The problem faced by any model for high-quality classroom instruction, therefore, is how to recreate, under existing conditions, the optimal quality of instruction usually found only under ideal conditions.

Research carried out both here and abroad (for instance, Block, 1971) suggests that the model in Figure 1 will solve this problem. As drawn, the model proposes that feedback and correction techniques are the catalysts required to transform ordinary group-based instruction into instruction of optimal quality for each learner. These techniques supplement the group instruction in such a way that the teacher can make the same kinds of modifications in each of his student's instruction as would a tutor for his tutee. The feedback devices detect gaps in each student's learning at selected intervals in the group instruction. Then, a number of group and/or individual instructional correctives are used systematically to vary the cues and/or the participation-practice and/or the reinforcements until each learner's problems have been surmounted. For example, reteaching is used to overcome learning difficulties shared by a group of students. Small-group study sessions, individual tutoring, and alternative instructional materials (textbooks, workbooks, programmed instruction, audio-visual methods, academic games and puzzles) are employed to handle individual learning errors.

Three major roles for student-evaluation techniques in the promotion of learning follow readily from this model. Each centers on the development of improved classroom feedback/correction techniques.

Feedback. One obvious role is the provision of the necessary feedback information. Here "formative, criterion-referenced" classroom evaluation instruments are required. The instruments should be *formative* in the sense that they are designed to be an integral part of the teaching-learning process and thus to shape the outcomes of the process. They should be short, ungraded, administered as frequently

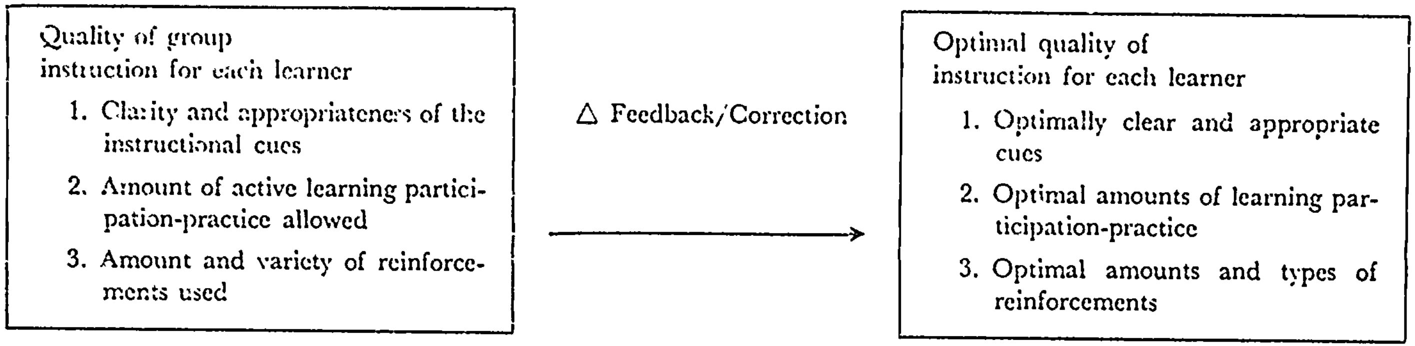


FIGURE 1. A MODEL FOR OPTIMAL INSTRUCTION

as necessary, and diagnostic. Useful formative instruments must provide a continuous flow of information to *both* the teacher and the learner regarding the ongoing effectiveness of the teaching and learning (Airasian, 1969; Bloom, Hastings, and Madaus, 1971). The instruments should be *criterion-referenced* in two senses. First, they should provide an in-depth picture of what objectives the student did not learn in a given instructional segment and where he encountered learning problems. This information will provide clues to why particular objectives were not attained. Second, they should assess the adequacy-inadequacy of the student's learning with respect to *what* objectives he attained, rather than *how many* he reached relative to his peers (Block, 1971; Bormuth, 1970; Glaser and Nitko, 1971).

Correction process. A second role for student evaluation is the streamlining of the correction process. Here the term *streamlining* does not mean the assemblage of the most efficient and effective, yet parsimonious and inexpensive, corrective battery possible. The construction of such a battery primarily requires the techniques of curricular, not student, evaluation. Rather, it means the development of efficient procedures for translating the learning difficulties evidenced by a student on a feedback instrument into a set of correctives which will most likely overcome these difficulties.

This streamlining will require the creation of what might be called a "corrective contingency management system." One component of such a system must be a battery of evaluation instruments compatible with the classroom instruction in which the student will engage (Cronbach, 1970). These instruments should delimit the cues to be used, the participation-practice to be allowed, and the reinforcements to be provided, and then ascertain each student's status relative to these characteristics. The second component must be a battery of evaluation instruments designed to identify the particular types of cues, the amounts of participation-practice, and the types and amounts of reinforcement each student requires for optimal quality of instruction over the given material.

From information regarding the classroom instruction's major characteristics and each student's status relative to these characteristics, general areas can be identified where the quality of the group instruction will not be sufficient for a particular student. Contingency plans, based upon each student's requirements for optimal instruction, can then be developed. These plans will specify a particular set of correctives likely to remove the incongruence between the quality of instruction provided and the optimal quality of instruction required. Should the student actually encounter learning problems over any segment of the group instruction, the contingency plans would be set in operation.

Consider this example. Suppose the amounts of participation-practice allowed and the amounts and types of reinforcement to be provided by some classroom instruction are satisfactory for a given student. But suppose the cues are to be presented in a mode (perhaps verbally) that the student cannot understand. The instructional cues, then, might be a potential source of learning difficulty for the student. Hence, a set of contingency plans would be drawn up based on information regarding the particular cues this student requires for optimal instruction. Should he actually encounter learning difficulties, the contingency plans would be implemented to present the unlearned material in a cue mode (perhaps nonverbally) he could most easily understand.

Standards. A final role for student evaluation techniques is the setting of performance standards against which the adequacy or inadequacy of each student's learning over a given instructional segment can be judged. Bormuth (1969, 1971) has demonstrated the possibility of empirically setting more objective and realistic standards than those used in the past. Block (1970), in turn, has demonstrated the possibility of empirically setting standards whose maintenance throughout an instructional sequence maximizes selected end-of-sequence learning outcomes. Taken together, therefore, this research suggests the possibility of using performance standards as major instructional variables in the promotion of student learning. If the teacher or administrator can select a particular set of desired learning outcomes, then he should be able to set a standard whose maintenance throughout an instructional sequence will maximize each student's attainments relative to these outcomes. Once set this standard would act as a benchmark for monitoring either the "health" or equivalently what Farquhar (1968) terms the "pathology" of each urban student's learning. At any time the student's learning was substandard, correctives would be applied to return it to par.

To illustrate these ideas, consider the following example. The basic subjects in most elementary school curricula are learned sequentially. Thus, the student's failure to learn adequately one task in a sequence may condemn him to learning failure in all subsequent tasks. The tragedy of this situation is graphically illustrated in a longitudinal study of school achievement conducted by Payne (1963). She found that a student's success or failure in sixth-grade reading or arithmetic could be highly predicted given knowledge of only his first- and perhaps second-grade performance. That is, the student's performance in grades one and two seemed to predetermine his success or failure later in the sequence. As Bloom (1970) points out, had short-term performance standards ("sequential norms") been available in grades one and two, then some students' subsequent long-term learning failure might have been averted. Urban schools should be able to use already existing data to set such standards for all required, sequentially learned subjects. Initially the standards might be set to indicate simply "how much" the student should have learned at each stage in the sequence if his learning is to meet certain end-of-sequence criteria. Later, they might be set in terms of "what" he should have learned.

Implications for Urban Education

Student benefits. Urban students will benefit in at least four ways from the use of student-evaluation techniques to promote learning. First, especially in sequentially learned subjects, most students will learn more effectively. Research (Block, 1971) suggests that classroom use of sound feedback/correction techniques typically can enable 75 per cent of the students to learn selected subjects as well as do only the top 25 per cent at present. Virtually all students learning with the assistance of such techniques can achieve well enough to earn C or higher marks as judged by the same grading criteria used for students learning without such assistance.

Second, most students' learning can be made more efficient. Especially for sequentially learned subjects or topics, there seem to be particular performance standards whose attainment increases the amount of subsequent material students learn in a given time (Block, 1970; Merrill, Barton, and Wood, 1970). The reason is clear. The student who is helped to learn adequately the material in one task in a sequence can concentrate on the new material in the next task. The student who is not helped must spend time learning both the former and the latter material. The further he is allowed to proceed unassisted in the sequence, the more time he must devote to learning old rather than new material.

Third, most students will be partially insulated from the negative affective consequences of present instructional and evaluational practices. On the one hand, they will gain the repeated positive school learning experiences which foster positive interests in and attitudes toward learning, promote the growth of a positive academic and probably general self-concept, and immunize against mental illness (Bloom, 1971; Purkey, 1970; Stringer and Glidewell, 1967; Torshen, 1969). On the other hand, they will obtain new perspectives on the meaning of student evaluation which should reduce its anxiety-provoking qualities and exaggerated importance in student life. In particular, students will discover that evaluation is an integral part of the teaching-learning process rather than the ultimate goal of the process and consequently that grading and evaluation need not be synonymous.

Finally, many students will be freed from the negative scholastic effects of the concept of individual differences. Few ideas have had greater positive impact on urban education than the notion of individual differences as many chapters in this volume attest. But the idea has also provided some urban educators a convenient series of excuses whereby responsibility for poor student learning is shifted from the educational system to the student's genetic and/or environmental background. As these educators have absolved themselves from responsibility for student learning, they have become increasingly lax in providing the conditions under which almost all students can learn.

The use of student-evaluation techniques to develop improved classroom feedback/correction processes should place the concept of individual differences in proper perspective in urban education. Research using relatively crude systems to improve the quality of instruction for suburban, lower-middle class students (Block, 1970) and urban students (Kersh, 1970; Kim et al., 1969) suggests that the commonly observed relationship between individual differences and student achievement may often be a sheer *artifact* of our present instructional practices. These studies show that the degree to which individual differences in such variables as aptitude, socioeconomic status, and IQ are reflected in student achievement is some negative function of the quality of instruction provided each learner. The higher the quality of instruction, the less the effects of these differences. Thus, if these findings are replicated as improved feedback/correction systems evolve, they should caution urban educators against using any student's background as a scapegoat for instructional ineffectiveness.

Let me emphasize that I am not proposing that individual differences may not condition student learning in some cases. Rather, I am suggesting that perhaps the concept of individual differences has been uncritically overused in urban education to explain our inability to teach some students or our ability to teach some better than others.

Teacher benefits. Urban teachers also will receive certain important benefits from the use of student evaluation as proposed. First, they can reap returns of increased learning for most, rather than just a few, students from their instruction. These returns will be greater each time a particular topic, course, or subject is retaught, provided the teachers continue to use the techniques described above to refine and improve their classroom feedback/correction procedures (see, for example, Bloom, 1968).

Second, especially for sequentially learned material, the teachers can make their instruction more efficient. Limited experimental research suggests that the teacher, with the assistance of a good feedback/correction system, may be able to ensure high levels of learning for most students while spending no more, perhaps even less, instructional time than is usually spent helping only some students learn well (Block, 1970; Merrill, Barton, and Wood, 1970).

Third, urban teachers will gain some tools by which to discharge growing public demands for accountability. To date, while teachers have been deluged with demands that they assure greater responsibility for their students' learning, almost without exception they have not been provided with any effective means for reaching this end. The development of improved classroom feedback/correction techniques should provide teachers with the vehicles they have lacked.

Fourth, urban teachers may be able to alleviate some of their classroom disorder problems. In particular, the use of student evaluation in the ways proposed above may help the teacher reduce the hostility and frustration generated by previous instructional and evaluational practices. The teachers can help most students attain consistent academic success and thus both the short- and long-term benefits of such success. High levels of student learning, in turn, should discourage teachers from underestimating their students' capacity for intellectual growth.

School benefits. Urban schools will gain a set of benefits equally as important as those accruing to their students and teachers. First, they will acquire the lever required to increase pupil learning substantially. Hence, like the individual teacher, the schools also can most fully capitalize on their tremendous investment of resources in student instruction. Sound feedback/correction systems have proved to be both an effective and efficient means of individualizing instruction even in classrooms with up to 70 to 1 student-teacher ratios and for samples of thousands and tens of thousands of students (Kim et al., 1969, 1970). So far, development costs have been minimal. For example, teachers working together for an hour or two a day have constructed sound feedback/correction systems for an entire year's work in courses such as mathematics and biology in fewer than ten weeks (Airasian, 1969). Once set up, of course, such systems can be used repeatedly.

Second, the schools may be able to reclaim experienced teachers and to save the inexperienced from the indifference toward teaching and learning excellence prevalent in some urban schools (Clark,

BEST COPY AVAILABLE

1965, Silberman, 1970). Doubtless such indifference may derive from poor working conditions; for the most part, however, it probably results from many teachers' inability to realize, regardless of their efforts, any tangible, positive student learning results. If the use of student-evaluation techniques to build improved classroom feedback/correction systems can enable each urban teacher significantly to improve his students' learning, then this indifference can be stemmed.

Third, the schools, within their present organizational and curricular framework, will be able to promote the fullest development of most students. They will be able to provide almost every learner with both those cognitive skills and competencies and those affective traits required for successful entry and residence in the adult world of work and leisure. From this cognitive and affective base, most students will be able to realize any vocational or avocational goals *they choose* instead of only those goals *chosen for them*.

This last argument may be unacceptable to the reader familiar with either Cronbach's (1971) or Ebel's (1970) distinction between training and education. Cronbach, for example, argues that training involves only the shaping of student behaviors in some basic areas while education entails training, the development of analytic and problem-solving skills, and the fostering of creativity and self-expression. On the basis of this distinction, the uses of student evaluation developed here might be misconstrued as fostering the fullest development of the individual with respect only to training and not to educational goals. Remember, however, that the proposed uses are meant for implementation within the framework of present urban educational systems. In this framework, students are typically required to spend ten to twelve years locked for the most part into a training curriculum consisting of long learning sequences for a few required subjects. The teaching of analytic and problem-solving skills and the promotion of creativity and self-expression usually are, in practice, either predicated upon or secondary to mastery of these subjects. Hence, a student's failure to learn these basic subjects and the repeated frustration, humiliation, and despair engendered thereby can make him unwilling or unable or both to participate in "educational" efforts to develop high-level cognitive skills or to nurture creativity and self-expression. Within such a framework, therefore, both training and educational goals are fostered by any set of procedures that will ensure mastery of the required subjects for most students.

Admittedly, the view taken here of student evaluation's latent potential for improving urban education is optimistic. But perhaps the best argument for this optimism is Pascal's argument for the suvival of the rational man: "One argument for believing in heaven is that if you are wrong, little is lost, but if you are right, much may be gained."

References

AIRASIAN, P. W. "Formative Evaluation Instruments. A Construction and Validation of Tests to Evaluate Learning over Short-Time Periods." Unpublished doctoral dissertation, University of Chicago, 1969.

AIRASIAN, P. W. "The Role of Evaluation in Mastery Learning." In J. H. Block (Ed.), *Mastery Learning: Theory and Practice*. New York: Holt, Rinehart, and Winston, 1971. Pp. 77–88.

BAKER, J. P., AND CRIST, J. "Teacher Expectancies: A Review of the Literature." In J. D. Elashoff and R. E. Snow (Eds.), *Pygmalion Reconsidered*. Worthington, Ohio: Jones, 1972. Pp. 48–64.

BLACK, H. *They Shall Not Pass*. New York: William Morrow, 1963.

BLOCK, J. H. "The Effects of Various Levels of Performance on Selected Cognitive, Affective and Time Variables." Unpublished doctoral dissertation, University of Chicago, 1970.

BLOCK, J. H. (Ed.). *Mastery Learning: Theory and Practice*. New York: Holt, Rinehart, and Winston, 1971.

BLOOM, B. S. "Learning for Mastery." *UCLA Evaluation Comment*, 1968, *1*, 1–12.

BLOOM, B. S. "Some Theoretical Issues Relating to Educational Evaluation." In *Educational Evaluation: New Roles, New Means*. Sixty-Eighth NSSE Yearbook, Part II. Chicago: University of Chicago Press, 1969. Pp. 26–50.

BLOOM, B. S. "Toward a Theory of Testing Which Includes Measurement-Evaluation-Assessment." In M. C. Wittrock and D. E. Wiley (Eds.), *The Evaluation of Instruction: Issues and Problems*. New York: Holt, Rinehart, and Winston, 1970. Pp. 25–50.

BLOOM, B. S. "Individual Differences in School Achievement: A Vanishing Point?" In *Education at Chicago*, Newsletter of the Department of Education, University of Chicago, Winter 1971, pp. 4–14.

BLOOM, B. S., HASTINGS, J. T., AND MADAUS, G. F. *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill, 1971.

BORMUTH, J. R. *Development of Readability Analysis*. Final report, USOE, Contract No. OEC-3-7-070052-0326, Department of Education, University of Chicago, 1969.

BORMUTH, J. R. *On the Theory of Achievement Test Items*. Chicago: University of Chicago Press, 1970.

BORMUTH, J. R. *Development of Standards of Readability: Toward a Rational Criterion of Passage Performance*. Final report, USOE, Project No. 9-0237, Department of Education, University of Chicago, 1971.

BRIM, O. G., JR., et al. *The Use of Standardized Ability Tests in American Secondary Schools and Their Impact on Students, Teachers, and Administrators*. Technical Report No. 3 on the Social Consequences of Testing. New York: Russell Sage Foundation, 1965.

CARROLL, J. B. "A Model of School Learning." *Teachers College Record*, 1963, *64*, 723–733.

CICOUREL, A. V., AND KITSUSE, J. I. *The Educational Decision Makers*. New York: Bobbs-Merrill, 1963.

CLARK, K. *Dark Ghetto: Dilemmas of Social Power*. New York: Harper & Row, 1965.

CRONBACH, L. J. *Essentials of Psychological Testing*. (2nd Ed.), New York: Harper & Row, 1960.

CRONBACH, L. J. "Mental Tests and the Creation of Opportunity." Paper presented at the annual meeting of the American Philosophical Society, April 1970.

CRONBACH, L. J. "Comments on 'Mastery Learning and Its Implications for Curriculum Development.'" In E. W. Eisner (Ed.), *Confronting Curriculum Reform*. Boston: Little, Brown, 1971. Pp. 49–55.

DRESSEL, P. L. "The Role of Evaluation in Teaching and Learning." In *Evaluation in Social Studies*. Thirty-Fifth Yearbook of the National Council for Social Studies. Washington, D.C.: National Education Association, 1965. Pp. 1–20.

EBEL, R. L. "Improving the Competence of Teachers in Educational Measurement." In J. T. Flynn and H. Garber (Eds.), *Assessing Behavior: Readings in Educational and Psychological Measurement*. Reading, Mass.: Addison-Wesley, 1967. Pp. 263–269.

EBEL, R. L. "Behavioral Objectives: A Close Look." *Phi Delta Kappan*, 1970, *52*, 171–173.

FARQUHAR, W. W. "Academic Motivation and Inner-City Schools." In H. C. Rudman and R. L. Featherstone (Eds.), *Urban Schooling*. New York: Harcourt, Brace and World, 1968.

FEATHER, N. T. "Effects of Prior Success and Failure on Expectations of Success and Subsequent Performance." *Journal of Personality and Social Psychology*, 1966, *3*, 187–198.

GAGE, N. L. "Theories of Teaching." In E. R. Hilgard (Ed.), *Theories of*

Learning and Instruction, sixty-third yearbook of the National Society for the Study of Education. Part 1. Chicago: National Society for the Study of Education, 1964. Pp. 268 285.

GLASER, R., AND NITKO. A. "Measurement in Learning and Instruction." In R. L. Thorndike (Ed.), *Educational Measurement*. Washington, D.C.: American Council on Education, 1971. Pp. 625-670.

GOSLIN. D. A. *Teachers and Testing*. New York: Russell Sage Foundation, 1967.

GOTTLIEB, D. "Teaching and Students. The Views of Negro and White Teachers." In S. W. Webster (Ed.), *The Disadvantaged Learner*. San Francisco: Chandler, 1966.

Harlem Youth Opportunities Unlimited. *Youth in the Ghetto*. New York: Harlem Youth Opportunities, 1964.

HILGARD, E. R. *Theories of Learning*. New York: Appleton-Century-Crofts, 1956.

KERSH, M. E. "A Strategy for Mastery Learning in Fifth-Grade Arithmetic." Unpublished doctoral dissertation, University of Chicago, 1970.

KIM, HOGWON, et al. *A Study of the Bloom Strategies for Mastery Learning*. Seoul: Korean Institute for Research in the Behavioral Sciences. 1969 (In Korean.)

KIM, HOGWON, et al. *The Mastery Learning Project in the Middle Schools*. Seoul: Korean Institute for Research in the Behavioral Sciences, 1970. (In Korean.)

LECKY, P. *Self-Consistency*. New York: Island Press, 1945.

MERRILL, M. D., BARTON, K., AND WOOD, L. E. "Specific Review in Learning a Hierarchical Imaginary Science." *Journal of Educational Psychology*, 1970, *61*. 102 109.

MOORE, C. A. "Tutoring Dyadic Interaction and Interpersonal Processes." Unpublished bibliography, Center for Research and Development in Teaching, Stanford University, 1971.

PAYNE, M. A. "The Use of Data in Curriculum Decisions." Unpublished doctoral dissertation, University of Chicago, 1963.

PURKEY, W. W. *Self Concept and School Achievement*. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

ROSENSHINE, B. "Teaching Behaviors and Student Achievement: A Review of Research." Stockholm. International Association for Evaluation of Educational Achievement. 1970.

SARASON, S. B., et al. *Anxiety in Elementary School Children*. New York: Wiley, 1960.

SILBERMAN, C. E. *Crisis in the Classroom*. New York: Random House, 1970.

STANWYCK, D. J., AND FELKER, D. W. "Intellectual Achievement Responsibility and Anxiety as Functions of Self Concept of Third- to Sixth-Grade Boys and Girls." Paper presented at the annual meeting of the American Educational Research Association, New York, 1971.

STRINGER, L. A., AND GLIDEWELL, J. C. *Early Detection of Emotional Illnesses in School Children.* Final report. St. Louis, Missouri: Division of Research and Development, St. Louis County Health Department, 1967.

THELEN, H., et al. *Learning by Teaching.* Chicago: Stone-Brandel Center, 1968.

TORSHEN, K. P. "The Relation of Classroom Evaluation to Students' Self Concepts and Mental Health." Unpublished doctoral dissertation, University of Chicago, 1969.

TYLER, R. W. *Basic Principles of Curriculum and Instruction.* Chicago: University of Chicago Press, 1950.

WALLEN, N. E., AND TRAVERS, R. M. W. "Analysis and Investigation of Teaching Methods." In N. L. Gage (Ed.), *Handbook of Research on Teaching.* Chicago: Rand McNally, 1963. Pp. 448–505.

WILES, D. K. "The Mosaic Composition of Urban School Teachers." *Urban Education,* 1970, *5,* 141-151.

Behavioral Objectives: A Close Look

By Robert L. Ebel

Most teachers have heard of behavioral objectives. They have read books and articles which urge them to state their own instructional objectives in behavioral terms.[1] Some of them have tried to do so, and lacking clear success may feel some guilt. A few teachers actually do have statements of behavioral objectives for their courses and build their teaching efforts around them. But the number of these is small. Ammons, in fact, found *no* behavioral objectives in the 300 school systems she surveyed.[2] Some educators are not greatly concerned with this state of affairs. They see limited value in behavioral objectives and some potential danger in making behavior, rather than cognitive processes, the target of our educational efforts.

The Origin and History of the Concept

Although the phrase "behavioral objectives" has not been widely used until recent times, every program of training does in fact have behavioral objectives, whether they are stated explicitly or not. The purpose of training for a specific task is to develop the capability for the behavior required by the task. But the broader usage of behavioral objectives in connection with *educational* programs is probably attributable largely to Ralph Tyler.[3] While at Ohio State University, he developed a systematic program for the specification, in behavioral terms, of the desired outcomes of a course. Usually these outcomes were a limited number of fairly specific cognitive abilities. Their emphasis was, in part, a

PHI DELTA KAPPAN, 1970, Vol. 52, pp. 171-173.

reaction to the overemphasis on factual information in many current objective tests of achievement.

With the advent of teaching machines and programmed instruction, suggested first by Pressey[4] and popularized by Skinner,[5] the usefulness of behavioral objectives became more apparent, especially to the programmers. Then the cutting edge of innovation moved on to more complex models of systematic instruction. With computers prescribing individualized instruction[6] and "mastery" replacing "as much as possible" as the goal, behavioral objectives remained an essential feature of innovation.

The net effect of both Tyler's leadership and recent developments has been to convince many teachers that they ought to state their objectives in behavioral terms. "Help stamp out nonbehavioral objectives" is their only half facetious slogan.

Justifications for Behavioral Objectives

In the case of programmed instruction and the more complex learning systems, the need for specific, detailed instructional objectives is obvious. Some of these systems may be too complex to be generally feasible, and too impersonal or too inflexible to be generally effective. But where they can be used they require and make good use of behavioral objectives.

But why should the ordinary nonmachine like teacher state his objectives in behavioral terms? Two justifications have been offered. The first, more basic and far-reaching, is that since the general purpose of all education is to change behaviors, course objectives should be stated in terms of the behaviors expected to result from the course. The second is quite different. It justifies the use of behavioral descriptions of objectives on the ground that such descriptions are more meaningful.[7]

When the purpose of instruction is to provide training for a particular task, the first justification can hardly be questioned. Even when the purpose is to provide more general, liberal education, one can argue that it is only justified if it affects behavior somehow, sometime. It may not be possible to foresee all the ways in which learning might affect future behavior, but surely some of the more probable and more important can be anticipated. On the other hand, it is quite clear that such behavioral consequences are not the real objectives of instruction. Those objectives are, rather, the knowledge and understanding, the attitudes and values which induced the behavior or made it possible. To stress behavior as the objective is somewhat inaccurate and misleading.

What of the second justification? Do behavioral objectives have clearer, more definite meaning than nonbehavioral objectives in conventional classroom instruction? In one sense they do because behavior is overt and observable, whereas knowledge, understanding, ability, etc., are hidden inside their possessors. We can assess these internal qualities only by eliciting behavior that is dependent on them. But here again the overt behavior is not the real objective. It is simply a useful indicator. To refer to it as the objective is more apt to confuse than to clarify thinking about educational goals.

Problems with Behavioral Objectives

In view of the widespread endorsement of behavioral objectives, one might expect to find many examples of their effective use. That this is not the case suggests that practical application of the concept may involve some difficulties.

One of these is the difficulty of knowing precisely what the concept means. Some use it as if the behavior in which they are interested is that of the student while he is learning, or even that of his teacher. Others use it to refer to the student's behavior on special tasks designed to show whether or not he has learned something. Still others have in mind the student's use in life, or on the job, of what he has learned in school. While these three meanings are more closely related in some subjects of study than in others, they are distinctly different. One cannot speak or even think clearly about behavioral objectives without defining which type of behavior he has in mind.

Another difficulty is that the behavior specified in these definitions is seldom the real objective of the instruction. When the behavior is that of the learner while learning, it is clearly a means to an end, not the end itself. Nor is test behavior the real objective except in those rare cases where the test is a performance test in a natural setting. Only in the third sense of on-the-job performance can behavior be the real objective. The situations in which such behavioral objectives are appropriate appear to be limited to instruction which aims at the cultivation of particular skills. Behavioral objectives seem quite inappropriate to instructional efforts whose aim is to enable the student to respond adaptively and effectively to unique future problem situations, to equip him to make, independently but responsibly, the kind of individual choices and decisions which are the essence of human freedom.

A useful distinction can be made between training, for which behavioral objectives are often quite appropriate, and education, for which they are seldom appropriate. Educational development is little concerned with the establishment of predetermined responses to recurring problem situations. Rather, it is concerned with the student's understanding, his resources of useful and available knowledge, his intellectual self sufficiency. It sees him not as a puppet on strings controlled by his teachers, but as one who needs and wants the help of his teachers and others as he tackles the difficult problems of designing and building a life of his own.

A third problem is that of specifying the behavioral objective in sufficient detail. Any significant behavioral act, such as the construction of an achievement test for a course, consists of myriads of contributory acts. Often these are not easy to identify as separate elements in the total matrix of behavior. Often they vary from situation to situation. To identify and specify all of them may be an impossible task. But to the extent that these elements are not specified the behavior is left undefined.

A fourth problem is that of specifying an appropriate level of skill or competence in the behavior. Most significant acts of behavior cannot be said to be either present or absent, available or unavailable. They occur more or less often when appropriate, and are handled more or less well. To define them as educational objectives requires us to say not only what they are, but how well they are handled. This task also is difficult, and frequently seems to be more trouble than it is likely to be worth.

Some Limitations of Stated Objectives

There are problems in making effective use of any statement of objectives.

One is the problem of validity. Simply stating that something is an objective does not make it a desirable one. True, one must think about his objectives in order to state them, and thinking is one of the best ways of working to improve them. But then one must also think about objectives when doing anything rational about educating – when developing materials, planning procedures, or preparing for evaluations. There is no reason to believe that better thinking will go into the statement of objectives than into plans for attaining them.

Another is the problem of flexible adaptability. There is always danger that stated objectives may impose a rigid formality on teaching. Stated objectives may describe what a teacher plans to do, but they should seldom prescribe what he ought to do. On Tuesday he may perceive a more important objective than he wrote into his statement on Monday. The notion that there is no further need for creative thought about objectives once they have been stated is an enemy of dynamic teaching.[8]

Finally there is the problem of effective use. What do you do with a statement of objectives once you have it? If it is a good brief summary of your general objectives you may discuss it with your students. You may refer to it from time to time to keep your teaching on course, or to keep your evaluations relevant. But if it is a highly detailed statement of specific objectives, the chances are that it will be filed "for possible future reference" It will add little of value to your own cognitive resources, to the materials you use in instruction, or to your planning of instructional procedures. If you value creative teaching, you will not try to follow it step by step.

Conclusion

Teaching is purposeful activity. Part of a teacher's effectiveness depends on his having the right purposes. Hence it is important for the curriculum builder, the textbook writer, the teacher, and the student to think hard about their purposes, about the objectives they seek to achieve.

These considerations support the belief that objectives are important. They do not suggest that objectives need to be stated explicitly or in detail. The pedagogical issues that divide teachers, the inadequacies that limit their effectiveness, cannot be disposed of by statements of objectives. Little that is wrong with any teacher's educational efforts today can be cured by getting him to define his objectives more fully and precisely. We ought not to ask teachers to spend much of their limited time in writing elaborate statements of their objectives.

Nor should we insist that the statements be in behavioral terms. Our main business as teachers is developing the cognitive resources of our pupils, not shaping their behavior. The great majority of teachers at all levels who feel no urgent need to write out their objectives in detail, and in terms of behavior, are probably wiser on this matter than those who have exhorted them to change their ways. Too much of the current reverence for behavioral objectives is a consequence of not looking closely enough at their limitations.

---

[1] Robert F. Mager, *Preparing Instructional Objectives* Palo Alto, Calif.: Fearon, 1962; C. M. Lindvall, ed., *Defining Educational Objectives*. Pittsburgh, Pa.. University of Pittsburgh Press, 1964. and David R. Krathwohl, "Stating Objectives Appropriately for Program, for Curriculum, and for Instructional Material Development." *Jour-*

nal of Teacher Education, March, 19[illegible]5, pp. 83-92.

[2]Margaret Ammons, "An Empirical Study of Process and Product in Curriculum Development," *Journal of Educational Research*, May-June, 1964, pp. 451-57.

[3]Ralph W. Tyler, "A Generalized Technique for Constructing Achievement Tests," in *Constructing Achievement Tests*. Columbus, O.: Bureau of Educational Research, 1934.

[4]S. L. Pressey, "A Simple Apparatus Which Gives Tests and Scores and Teaches," *School and Society*, March 20, 1926, pp. 373-77.

[5]B. F. Skinner, "Science of Learning and the Art of Teaching," *Harvard Educational Review*, Spring, 1954, pp. 86-97.

[6]C. M. Lindvall and John O. Bolvin, "Programed Instruction in the Schools: An Application of Programming Principles in Individually Prescribed Instruction," in Phil C. Lange, ed., *Programed Instruction*, Sixty-sixth Yearbook, Part II, National Society for the Study of Education. Chicago: University of Chicago Press, 1967, pp. 217-54.

[7]Ralph H. Ojemann, "Should Educational Objectives Be Stated in Behavioral Terms?, Parts I, II, and III," *Elementary School Journal*, February, 1968, pp. 223-31; February, 1969, pp. 229-35; February, 1970, pp. 271-78.

[8]Elliot W. Eisner, "Educational Objectives: Help or Hindrance," *School Review*, Autumn, 1967, pp. 250-60.

Why Behavioral Objectives?

Martin R. Wong

Teaching is often described as a profession in which the rewards are measured in terms of satisfaction in helping students learn and in pleasure in watching students grow in learning and achievement. These are two excellent reasons for the use of behavioral objectives in teaching - so that both the teacher and the learner can have concrete evidence of growth and learning.

The more specific reasons for using behavioral objectives fall into four basic areas: goal setting; task analyzing; progress monitoring; and achievement assessing.

The use of behavioral objectives in goal setting clarifies goals precisely and unambiguously. The student knows what standard he is aiming for. The requirements, the behaviors, and the achievement that will result in reward are clear to the student in advance. He is not pitted against others so that maximizing his achievement means beating other people, but instead he competes with himself for the achievement of a goal.

Students can be involved in goal setting and the obvious advantages accruing from students working toward goals that they have helped to set for themselves are brought in to play. Thus objectives can be based on some kind of justi-

ORIGINAL MANUSCRIPT, 1974

fiable needs of the student instead of imposed on the student. In addition, goals that are unambiguously stated can be reviewed and evaluated so that they can remain current, relative to the needs of the students, and relevant to the times.

Task analyzing is another of those obvious tasks of good teaching that is more often overlooked than adhered to. It is only when a goal is set that the task of achieving the goal can be analyzed so as to find out what is required to reach it. Most learning tasks are complex and made up of many smaller tasks. Teaching the task as a whole, without specifically teaching some of the sub-tasks inherent in the whole, is often a slow process dependent on the learner picking up many of the subtasks for himself even though he may not know what they are. It is only when a complex goal has been analyzed into its many subtasks that the steps toward the goal--the things that need to be learned in order to achieve it--are clear and can be appropriately sequenced in order to maximally facilitate learning. Then, the teacher and the student can decide how best to work toward the goal, what order will be employed, what methods, tools, aids, learning experiences, and practices will be best.

It is often assumed and research seems to indicate that learners setting their own goals and seeing themselves making progress toward achieving goals is reinforcing. It seems hardly necessary to point out that before students and teachers can monitor progress toward goals, those goals need to be set in some unambiguous way. Then and only then can the teacher reinforce progress and the student gain the satisfaction of monitoring his own progress toward goals. When objectives are set and the task is analyzed and sequenced, at any given point in the progression toward a goal, both the teacher and the student will know what comes next in the steps toward achieving the goal.

Assessing achievement and assigning grades that represent different levels of achievement often places the teacher in a defensive quandry. If goals are not precisely set, it is impossible for the teacher to know how far along the student has gone toward achieving the goals. It is only when they are precisely stated--in behavioral terms--that concrete evidence can be available to indicate whether or not the goals have been met. Achievement can then be measured objectively, and accomplishment can be rewarded with high grades. Both the student and the teacher can rightly feel a sense of accomplishment and success in teaching and achieving goals.

Then grades take on meaning; a particular grade represents the achievement of specific goals, and these are spelled out in terms of what the learner can do when he has achieved them. The present most usual alternative is to assign grades on a curve in which each grade is only meaningful in relation to what other people in the class did. Students are pitted against each other for the grades instead of concerning themselves with their own achievement goals.

If the rewards for being in the teaching profession really are satisfaction and a continuing sense of accomplishment, then unambiguously stated goals--behavioral objectives--must be principal to the teaching practice. Only then can learning be maximally facilitated, progress monitored and rewarded, and true satisfaction for achievement and accomplishment deserved.

HEREDITY, ENVIRONMENT, AND THE QUESTION "HOW?"[1]

ANNE ANASTASI

Two or three decades ago, the so-called heredity-environment question was the center of lively controversy. Today, on the other hand, many psychologists look upon it as a dead issue. It is now generally conceded that both hereditary and environmental factors enter into all behavior. The reacting organism is a product of its genes and its past environment, while present environment provides the immediate stimulus for current behavior. To be sure, it can be argued that, although a given trait may result from the combined influence of hereditary and environmental factors, a specific difference in this trait between individuals or between groups may be traceable to either hereditary or environmental factors alone. The design of most traditional investigations undertaken to identify such factors, however, has been such as to yield inconclusive answers. The same set of data has frequently led to opposite conclusions in the hands of psychologists with different orientations.

Nor have efforts to determine the proportional contribution of hereditary and environmental factors to observed individual differences in given traits met with any greater success. Apart from difficulties in controlling conditions, such investigations have usually been based upon the implicit assumption that hereditary and environmental factors combine in an additive fashion. Both geneticists and psychologists have repeatedly demonstrated, however, that a more tenable hypothesis is that of interaction (15, 22, 28, 40). In other words, the nature and extent of the influence of each type of factor depend upon the contribution of the other. Thus the proportional contribution of heredity to the variance of a given trait, rather than being a constant, will vary under different environmental conditions. Similarly, under different hereditary conditions, the relative contribution of environment will differ. Studies designed to estimate the proportional contribution of heredity and environment, however, have rarely included measures of such interaction. The only possible conclusion from such research would thus seem to be that both heredity and environment contribute to all behavior traits and that the extent of their respective contributions cannot be specified for any trait. Small wonder that some psychologists regard the heredity-environment question as unworthy of further consideration!

But is this really all we can find out about the operation of heredity and environment in the etiology of behavior? Perhaps we have simply been asking the wrong questions. The traditional questions about heredity and environment may be intrinsically unanswerable. Psychologists began by asking *which* type of factor, hereditary or environmental, is responsible for individual differences in a given trait. Later, they tried to discover *how much* of the variance was attributable to heredity and how much to environment. It is the primary contention of this paper that a more fruitful approach is to be found in the question *"How?"* There is still much to be learned about the specific *modus operandi* of hereditary and environmental factors in the development of behavioral

[1] Address of the President, Division of General Psychology, American Psychological Association, September 4, 1957.

PSYCHOLOGICAL REVIEW, 1958, Vol. 65, pp. 197-208.

differences. And there are several current lines of research which offer promising techniques for answering the question "How?"

Variety of Interaction Mechanisms

Hereditary factors. If we examine some of the specific ways in which hereditary factors may influence behavior, we cannot fail but be impressed by their wide diversity. At one extreme, we find such conditions as phenylpyruvic amentia and amaurotic idiocy. In these cases, certain essential physical prerequisites for normal intellectual development are lacking as a result of hereditary metabolic disorders. In our present state of knowledge, there is no environmental factor which can completely counteract this hereditary deficit. The individual will be mentally defective, regardless of the type of environmental conditions under which he is reared.

A somewhat different situation is illustrated by hereditary deafness, which may lead to intellectual retardation through interference with normal social interaction, language development, and schooling. In such a case, however, the hereditary handicap can be offset by appropriate adaptations of training procedures. It has been said, in fact, that the degree of intellectual backwardness of the deaf is an index of the state of development of special instructional facilities. As the latter improve, the intellectual retardation associated with deafness is correspondingly reduced.

A third example is provided by inherited susceptibility to certain physical diseases, with consequent protracted ill health. If environmental conditions are such that illness does in fact develop, a number of different behavioral effects may follow. Intellectually, the individual may be handicapped by his inability to attend school regularly. On the other hand, depending upon age of onset, home conditions, parental status, and similar factors, poor health may have the effect of concentrating the individual's energies upon intellectual pursuits. The curtailment of participation in athletics and social functions may serve to strengthen interest in reading and other sedentary activities. Concomitant circumstances would also determine the influence of such illness upon personality development. And it is well known that the latter effects could run the gamut from a deepening of human sympathy to psychiatric breakdown.

Finally, heredity may influence behavior through the mechanism of social stereotypes. A wide variety of inherited physical characteristics have served as the visible cues for identifying such stereotypes. These cues thus lead to behavioral restrictions or opportunities and at a more subtle level —to social attitudes and expectancies. The individual's own self concept tends gradually to reflect such expectancies All of these influences eventually leave their mark upon his abilities and inabilities, his emotional reactions, goals, ambitions, and outlook on life.

The geneticist Dobzhansky illustrates this type of mechanism by means of a dramatic hypothetical situation. He points out that, if there were a culture in which the carriers of blood group AB were considered aristocrats and those of blood group O laborers, then the blood-group genes would become important hereditary determiners of behavior (12, p. 147). Obviously the association between blood group and behavior would be specific to that culture. But such specificity is an essential property of the causal mechanism under consideration.

More realistic examples are not hard to find. The most familiar instances occur in connection with constitutional types, sex, and race. Sex and skin pigmentation obviously depend upon heredity. General body build is strongly influenced by hereditary components, although also susceptible to environmental modification. That all these physical characteristics may exert a pronounced effect upon behavior within a given culture is well known. It is equally apparent, of course, that in different cultures the behavioral correlates of such hereditary physical traits may be quite unlike. A specific physical cue may be completely unrelated to individual differences in psychological traits in one culture, while closely correlated with them in another. Or it may be associated with totally dissimilar behavior characteristics in two different cultures.

It might be objected that some of the illustrations which have been cited do not properly exemplify the operation of hereditary mechanisms in behavior development, since hereditary factors enter only indirectly into the behavior in question. Closer examination, however, shows this distinction to be untenable. First it may be noted that the influence of heredity upon behavior is always indirect. No psychological trait is ever inherited as such. All we can ever say directly from behavioral observations is that a given trait shows evidence of being influenced by certain "inheritable unknowns." This merely defines a problem for genetic research; it does not provide a causal explanation. Unlike the blood groups, which are close to the level of primary gene products, psychological traits are related to genes by highly indirect and devious routes. Even the mental deficiency associated with phenylketonuria is several steps removed from the chemically defective genes that represent its hereditary basis. Moreover, hereditary influences cannot be dichotomized into the more direct and the less direct. Rather do they represent a whole "continuum of indirectness," along which are found all degrees of remoteness of causal links. The examples already cited illustrate a few of the points on this continuum.

It should be noted that as we proceed along the continuum of indirectness, the range of variation of possible outcomes of hereditary factors expands rapidly. At each step in the causal chain, there is fresh opportunity for interaction with other hereditary factors as well as with environmental factors. And since each interaction in turn determines the direction of subsequent interactions, there is an ever-widening network of possible outcomes. If we visualize a simple sequential grid with only two alternatives at each point, it is obvious that there are two possible outcomes in the one-stage situation, four outcomes at the second stage, eight at the third, and so on in geometric progression. The actual situation is undoubtedly much more complex, since there will usually be more than two alternatives at any one point.

In the case of the blood groups, the relation to specific genes is so close that no other concomitant hereditary or environmental conditions can alter the outcome. If the organism survives at all, it will have the blood group determined by its genes. Among psychological traits, on the other hand, some variation in outcome is always possible as a result of concurrent circumstances. Even in cases of phenylketonuria, intellectual development will exhibit some relationship with the type of care and training available to the individual. That behavioral outcomes show progressive diversification as we proceed along the continuum of indirectness is brought out by the other examples which were cited. Chronic illness *can* lead to scholarly renown or to intellectual immaturity; a mesomorphic physique *can* be a contributing factor in juvenile delinquency or in the attainment of a college presidency! Fablished data on Sheldon somatotypes provide some support for both of the latter outcomes.

Parenthetically, it may be noted that geneticists have sometimes used the term "norm of reaction" to designate the range of variation of possible outcomes of gene properties (cf. 13, p. 161). Thus heredity sets the "norm" or limits within which environmental differences determine the eventual outcome. In the case of some traits, such as blood groups or eye color, this norm is much narrower than in the case of other traits. Owing to the rather different psychological connotations of both the words "norm" and "reaction," however, it seems less confusing to speak of the "range of variation" in this context.

A large portion of the continuum of hereditary influences which we have described coincides with the domain of somatopsychological relations, as defined by Barker et al. (6). Under this heading, Barker includes "variations in physique that affect the psychological situation of a person by influencing the effectiveness of his body as a tool for actions or by serving as a stimulus to himself or others" (6, p. 1). Relatively direct neurological influences on behavior, which have been the traditional concern of physiological psychology, are excluded from this definition, Barker being primarily concerned with what he calls the "social psychology of physique." Of the examples cited in the present paper, deafness, severe illness, and the physical characteristics associated with social stereotypes would meet the specifications of somatopsychological factors.

The somatic factors to which Barker refers, however, are not limited to those of hereditary origin. Bodily conditions attributable to environmental causes operate in the same sorts of somatopsychological relations as those traceable to heredity. In fact, heredity-environment distinctions play a minor part in Barker's approach.

Environmental factors: organic. Turning now to an analysis of the role of environmental factors in behavior, we find the same etiological mechanisms which were observed in the case of hereditary factors. First, however, we must differentiate between two classes of environmental influences. (*a*) those producing organic effects which may in turn influence behavior and (*b*) those serving as direct stimuli for psychological reactions. The former may be illustrated by food intake or by exposure to bacterial infection, the latter, by tribal initiation ceremonies or by a course in algebra. There are no completely satisfactory names by which to designate these two classes of influences. In an earlier paper by Anastasi and Foley (4), the terms "structural" and "functional" were employed. However, "organic" and "behavioral" have the advantage of greater familiarity in this context and may be less open to misinterpretation. Accordingly, these terms will be used in the present paper.

Like hereditary factors, environmental influences of an organic nature can also be ordered along a continuum of indirectness with regard to their relation to behavior. This continuum closely parallels that of hereditary factors. One end is typified by such conditions as mental deficiency resulting from cerebral birth injury or from prenatal nutritional inadequacies. A more indirect etiological mechanism is illustrated by severe motor disorder—as in certain cases of cerebral palsy—*without* accompanying injury to higher neurological centers. In such instances, intellectual retardation may occur as an indirect result of the motor handicap, through the curtailment of educational and social activities. Obviously this causal mechanism corresponds closely to that of hereditary deafness cited earlier in the paper.

Finally, we may consider an environmental parallel to the previously discussed social stereotypes which were mediated by hereditary physical cues. Let us suppose that a young woman with mousy brown hair becomes transformed into a dazzling golden blonde through environmental techniques currently available in our culture. It is highly probable that this metamorphosis will alter, not only the reactions of her associates toward her, but also her own self concept and subsequent behavior. The effects could range all the way from a rise in social poise to a drop in clerical accuracy!

Among the examples of environmentally determined organic influences which have been described, all but the first two fit Barker's definition of somatopsychological factors. With the exception of birth injuries and nutritional deficiencies, all fall within the social psychology of physique. Nevertheless, the individual factors exhibit wide diversity in their specific *modus operandi* a diversity which has important practical as well as theoretical implications.

Environmental factors: behavioral. The second major class of environmental factors the behavioral as contrasted to the organic are by definition direct influences. The immediate effect of such environmental factors is always a behavioral change. To be sure, some of the initial behavioral effects may themselves indirectly affect the individual's later behavior. But this relationship can perhaps be best conceptualized in terms of breadth and permanence of effects. Thus it could be said that we are now dealing, not with a continuum of indirectness, as in the case of hereditary and organic-environmental factors, but rather with a continuum of breadth.

Social class membership may serve as an illustration of a relatively broad, pervasive, and enduring environmental factor. Its influence upon behavior development may operate through many channels. Thus social level may determine the range and nature of intellectual stimulation provided by home and community through books, music, art, play activities, and the like. Even more far-reaching may be the effects upon interests and motivation, as illustrated by the desire to perform abstract intellectual tasks, to surpass others in competitive situations, to succeed in school, or to gain social approval. Emotional and social traits may likewise be influenced by the nature of interpersonal relations characterizing homes at different socioeconomic levels. Somewhat more restricted in scope than social class, although still exerting a relatively broad influence, is amount of formal schooling which the individual is able to obtain.

A factor which may be wide or narrow in its effects, depending upon concomitant circumstances, is language handicap. Thus the bilingualism of an adult who moves to a foreign country with inadequate mastery of the new language represents a relatively limited handicap which can be readily overcome in most cases. At most, the difficulty is one of communication. On the other hand, some kinds of bilingualism in childhood may exert a retarding influence upon intellectual development and may under certain conditions affect personality development adversely (2, 5, 10). A common pattern in the homes of immigrants is that the child speaks one language at home and another in school, so that his knowledge of each language is limited to certain types of situations. Inadequate facility with the language of the school interferes with the acquisition of basic concepts, intellectual skills, and information. The frustration engendered by scholastic difficulties may in turn lead to discouragement and general dislike of school. Such reactions can be found, for example, among a number of Puerto Rican children in New York City schools (3). In the case of certain groups, moreover, the child's foreign language background may be perceived by himself and his associates as a symbol of minority group status and may thereby augment any emotional maladjustment arising from such status (34).

A highly restricted environmental influence is to be found in the opportunity to acquire specific items of information occurring in a particular intelligence test. The fact that such opportunities may vary with culture, social class, or individual experiential background is at the basis of the test user's concern with the problem of coaching and with "culture-free" or "culture fair" tests (cf. 1, 2). If the advantage or disadvantage which such experiential differences confer upon certain individuals is strictly confined to performance on the given test, it will obviously reduce the validity of the test and should be eliminated.

In this connection, however, it is essential to know the breadth of the environmental influence in question. A fallacy inherent in many attempts to develop culture fair tests is that the breadth of cultural differentials is not taken into account. Failure to consider breadth of effect likewise characterizes certain discussions of coaching. If, in coaching a student for a college admission test, we can improve his knowledge of verbal concepts and his reading comprehension, he will be better equipped to succeed in college courses. His performance level will thus be raised, not only on the test, but also on the criterion which the test is intended to predict. To try to devise a test which is not susceptible to such coaching would merely reduce the effectiveness of the test. Similarly, efforts to rule out cultural differentials from test items so as to make them equally "fair" to subjects in different social classes or in different cultures may merely limit the usefulness of the test, since the same cultural differentials may operate within the broader area of behavior which the test is designed to sample.

Methodological Approaches

The examples considered so far should suffice to highlight the wide variety of ways in which hereditary and environmental factors may interact in the course of behavior development. There is clearly a need for identifying explicitly the etiological mechanism whereby any given hereditary or environmental condition ultimately leads to a behavioral characteristic in other words, the "how" of heredity and environment. Accordingly, we may now take a quick look at some promising methodological approaches to the question "how."

Within the past decade, an increasing number of studies have been designed to trace the connection between specific factors in the hereditary backgrounds or in the reactional biographies of individuals and their observed behavioral characteristics. There has been a definite shift away from the predominantly descriptive and correlational approach of the earlier decades toward more deliberate attempts to verify explanatory hypotheses. Similarly, the cataloguing of group differences in psychological traits has been giving way gradually to research on *changes* in group characteristics following altered conditions.

Among recent methodological developments, we have chosen seven as being particularly relevant to the analysis of etiological mechanisms. The first represents an extension of selective breeding investigations to permit the identification of specific hereditary conditions underlying the observed behavioral differences. When early selective breeding investigations such as those of Tryon

(36) on rats indicated that "maze learning ability" was inherited, we were still a long way from knowing what was actually being transmitted by the genes. It was obviously not "maze learning ability" as such. Twenty—or even ten years ago, some psychologists would have suggested that it was probably general intelligence. And a few might even have drawn a parallel with the inheritance of human intelligence.

But today investigators have been asking: Just what makes one group of rats learn mazes more quickly than the other? Is it differences in motivation, emotionality, speed of running, general activity level? If so, are these behavioral characteristics in turn dependent upon group differences in glandular development, body weight, brain size, bio chemical factors, or some other organic conditions? A number of recent and ongoing investigations indicate that attempts are being made to trace, at least part of the way, the steps whereby certain chemical properties of the genes may ultimately lead to specified behavior characteristics.

An example of such a study is provided by Searle's (31) follow-up of Tryon's research. Working with the strains of maze-bright and maze-dull rats developed by Tryon, Searle demonstrated that the two strains differed in a number of emotional and motivational factors, rather than in ability. Thus the strain differences were traced one step further, although many links still remain to be found between maze learning and genes A promising methodological development within the same general area is to be found in the recent research of Hirsch and Tryon (18). Utilizing a specially devised technique for measuring individual differences in behavior among lower organisms, these investigators launched a series of studies on selective breeding for behavioral characteristics in the fruit fly, *Drosophila*. Such research can capitalize on the mass of available genetic knowledge regarding the morphology of *Drosophila*, as well as on other advantages of using such an organism in genetic studies.

Further evidence of current interest in the specific hereditary factors which influence behavior is to be found in an extensive research program in progress at the Jackson Memorial Laboratory, under the direction of Scott and Fuller (30). In general, the project is concerned with the behavioral characteristics of various breeds and cross-breeds of dogs. Analyses of some of the data gathered to date again suggest that "differences in performance are produced by differences in emotional, motivational, and peripheral processes, and that genetically caused differences in central processes may be either slight or non-existent" (29, p. 225). In other parts of the same project, breed differences in physiological characteristics, which may in turn be related to behavioral differences, have been established.

A second line of attack is the exploration of possible relationships between behavioral characteristics and physiological variables which may in turn be traceable to hereditary factors. Research on EEG, autonomic balance, metabolic processes, and biochemical factors illustrates this approach. A lucid demonstration of the process of tracing a psychological condition to genetic factors is provided by the identification and subsequent investigation of phenylpyruvic amentia. In this case, the causal chain from defective gene, through metabolic disorder and consequent cerebral malfunctioning, to feeblemindedness and other overt symptoms can be described step by step (cf. ?2; 33, pp. 389-391). Also relevant are the recent researches on neurological and biochemical correlates of schizophrenia (9). Owing to inadequate methodological controls, however, most of the findings of the latter studies must be regarded as tentative (19).

Prenatal environmental factors provide a third avenue of fruitful investigation. Especially noteworthy is the recent work of Pasamanick and his associates (27), which demonstrated a tie up between socioeconomic level, complications of pregnancy and parturition, and psychological disorders of the offspring. In a series of studies on large samples of whites and Negroes in Baltimore, these investigators showed that various prenatal and paranatal disorders are significantly related to the occurrence of mental defect and psychiatric disorders in the child. An important source of such irregularities in the process of childbearing and birth is to be found in deficiencies of maternal diet and in other conditions associated with low socioeconomic st 'us. An analysis of the data did in fact reveal a much higher frequency of all such medical complications in lower than in higher socioeconomic levels, and a higher frequency among Negroes than among whites.

Direct evidence of the influence of prenatal nutritional factors upon subsequent intellectual development is to be found in a recent, well controlled experiment by Harrell et al. (16). The subjects were pregnant women in low income groups, whose normal diets were generally quite deficient. A dietary supplement was administered to some of these women during pregnancy and lactation, while an equated control group received placebos. When tested at the ages of three and four years. the offspring of the experimental group obtained a significantly higher mean IQ than did the offspring of the controls.

Mention should also be made of animal experiments on the effects of such factors as prenatal radiation and neonatal asphyxia upon cerebral anomalies as well as upon subsequent behavior development. These experimental studies merge imperceptibly into the fourth approach to be considered, namely, the investigation of the influence of early experience upon the eventual behavioral characteristics of animals. Research in this area has been accumulating at a rapid rate. In 1954, Beach and Jaynes (8) surveyed this literature for the *Psychological Bulletin,* listing over 130 references. Several new studies have appeared since that date (e.g., 14, 21, 24, 25, 35). The variety of factors covered ranges from the type and quantity of available food to the extent of contact with human culture. A large number of experiments have been concerned with various forms of sensory deprivation and with diminished opportunities for motor exercise. Effects have been observed in many kinds of animals and in almost all aspects of behavior, including perceptual responses, motor activity, learning, emotionality, and social reactions.

In their review, Beach and Jaynes pointed out that research in this area has been stimulated by at least four distinct theoretical interests. Some studies were motivated by the traditional concern with the relative contribution of maturation and learning to behavior development. Others were designed in an effort to test certain psychoanalytic theories regarding infantile experiences, as illustrated by studies which limited the feeding responses of young animals. A third relevant influence is to be found in the work of the European biologist Lorenz (23) on early social stimulation of birds, and in particular on the special type of learning for which the term "imprinting" has been coined. A relatively large number of recent studies have centered around Hebb's (17) theory regarding the importance of early perceptual experiences upon subsequent performance in learning situations. All this research represents a rapidly growing and promising attack on the *modus operandi* of specific environmental factors.

The human counterpart of these animal studies may be found in the comparative investigation of child rearing practices in different cultures and subcultures. This represents the fifth approach in our list. An outstanding example of such a study is that by Whiting and Child (38), published in 1953. Utilizing data on 75 primitive societies from the Cross Cultural Files of the Yale Institute of Human Relations, these investigators set out to test a number of hypotheses regarding the relationships between child rearing practices and personality development. This analysis was followed up by field observations in five cultures, the results of which have not yet been reported (cf. 37).

Within our own culture, similar surveys have been concerned with the diverse psychological environments provided by different social classes (11). Of particular interest are the study by Williams and Scott (39) on the association between socioeconomic level, permissiveness, and motor development among Negro children, and the exploratory research by Milner (26) on the relationship between reading readiness in first grade children and patterns of parent child interaction. Milner found that upon school entrance the lower-class child seems to lack chiefly two advantages enjoyed by the middle class child. The first is described as "a warm positive family atmosphere or adult relationship pattern which is more and more being recognized as a motivational prerequisite of any kind of adult controlled learning." The lower-class children in Milner's study perceived adults as predominantly hostile. The second advantage is an extensive opportunity to interact verbally with adults in the family. The latter point is illustrated by parental attitudes toward mealtime conversation, lower-class parents tending to inhibit and discourage such conversation, while middle-class parents encourage it.

Most traditional studies on child-rearing practices have been designed in terms of a psychoanalytic orientation. There is need for more data pertaining to other types of hypotheses. Findings such as those of Milner on opportunities for verbalization and the resulting effects upon reading readiness represent a step in this direction. Another possible source of future data is the application of the intensive observational techniques of psychological ecology developed by Barker and Wright (7) to widely diverse socioeconomic groups.

A sixth major approach involves research on the previously cited somatopsychological relationships (6). To date, little direct information is available on the precise operation of this class of factors in psychological development. The multiplicity of ways in which physical traits—whether hereditary or environmental in origin—may influence behavior thus offers a relatively unexplored field for future study.

The seventh and final approach to be considered represents an adaptation of traditional twin studies. From the standpoint of the question "How?" there is need for closer coordination between the usual data on twin resemblance and observations of the family interactions of twins. Available data already suggest, for example, that closeness of contact and extent of environmental similarity are greater in the case of monozygotic than in the case of dizygotic twins (cf. 2). Information on the social reactions of twins toward each other and the specialization of roles is likewise of interest (2). Especially useful would be longitudinal studies of twins, beginning in early infancy and following the subjects through school age. The operation of differential environmental pressures, the development of specialized roles, and other environmental influences could thus be more clearly identified and correlated with intellectual and personality changes in the growing twins.

Parenthetically, I should like to add a remark about the traditional applications of the twin method, in which persons in different degrees of hereditary and environmental relationships to each other are simply compared for behavioral similarity. In these studies, attention has been focused principally upon the amount of resemblance of monozygotic as contrasted to dizygotic twins. Yet such a comparison is particularly difficult to interpret because of the many subtle differences in the environmental situations of the two types of twins. A more fruitful comparison would seem to be that between dizygotic twins and siblings, for whom the hereditary similarity is known to be the same. In Kallmann's monumental research on psychiatric disorders among twins (20), for example, one of the most convincing bits of evidence for the operation of hereditary factors in schizophrenia is the fact that the degrees of concordance for dizygotic twins and for siblings were practically identical. In contrast, it will be recalled that in intelligence test scores dizygotic twins resemble each other much more closely than do siblings a finding which reveals the influence of environmental factors in intellectual development.

Summary

The heredity environment problem is still very much alive. Its viability is assured by the gradual replacement of the questions, "Which one?" and "How much?" by the more basic and appropriate question, "How?" Hereditary influences as well as environmental factors of an organic nature vary along a "continuum of indirectness." The more indirect their connection with behavior, the wider will be the range of variation of possible outcomes. One extreme of the continuum of indirectness may be illustrated by brain damage leading to mental deficiency, the other extreme, by physical characteristics associated with social stereotypes. Examples of factors falling at intermediate points include deafness, physical diseases, and motor disorders. Those environmental factors which act directly upon behavior can be ordered along a continuum of breadth or permanence of effect, as exemplified by social class membership, amount of formal schooling, language handicap, and familiarity with specific test items.

Several current lines of research offer promising techniques for exploring the *modus operandi* of hereditary and environmental factors. Outstanding among them are investigations of. (*a*) hereditary conditions which underlie behavioral differences between selectively bred groups of animals, (*b*) relations between physiological variables and individual differences in behavior, especially in the case of pathological deviations, (*c*) role of prenatal physiological factors in behavior development, (*d*) influence of early experience upon eventual behavioral characteristics, (*e*) cultural differences in child rearing practices in relation to intellectual and emotional development, (*f*) mechanisms of somatopsychological relationships, and (*g*) psychological development of twins from infancy to maturity, together with observations of their social environment. Such approaches are extremely varied with regard to subjects employed, nature of psychological functions studied, and specific experimental procedures followed. But it is just such heterogeneity of methodology that is demanded by the wide diversity of ways in which he-

reditary and environmental factors interact in behavior development.

REFERENCES

1. ANASTASI, ANNE. *Psychological testing.* New York: Macmillan, 1954.
2. ANASTASI, ANNE. *Differential psychology.* (3rd ed.) New York: Macmillan, 1958.
3. ANASTASI, ANNE, & CORDOVA, F. A. Some effects of bilingualism upon the intelligence test performance of Puerto Rican children in New York City. *J. educ. Psychol.*, 1953, 44, 1-19.
4. ANASTASI, ANNE, & FOLEY, J. P., JR. A proposed reorientation in the heredity-environment controversy. *Psychol. Rev.*, 1948, 55, 239-249.
5. ARSENIAN, S. Bilingualism in the post-war world. *Psychol. Bull.*, 1945, 42, 65-86.
6. BARKER, R. G., WRIGHT, BEATRICE A., MYERSON, L., & GONICK, MOLLIE R. Adjustment to physical handicap and illness. A survey of the social psychology of physique and disability. *Soc. Sci. Res. Coun. Bull.*, 1953, No. 55 (Rev.).
7. BARKER, R. G., & WRIGHT, H. F. *Midwest and its children: The psychological ecology of an American town.* Evanston, Ill.: Row, Peterson, 1955.
8. BEACH, F. A., & JAYNES, J. Effects of early experience upon the behavior of animals. *Psychol. Bull.*, 1954, 51, 239-263.
9. BRACKBILL, G. A. Studies of brain dysfunction in schizophrenia. *Psychol. Bull.*, 1956, 53, 210-226.
10. DARCY, NATALIE T. A review of the literature on the effects of bilingualism upon the measurement of intelligence. *J. genet. Psychol.*, 1953, 82, 21-57.
11. DAVIS, A., & HAVIGHURST, R. J. Social class and color differences in child rearing. *Amer. sociol. Rev.*, 1946, 11, 698-710.
12. DOBZHANSKY, T. The genetic nature of differences among men. In S. Persons (Ed.), *Evolutionary thought in America.* New Haven: Yale Univer. Press, 1950. Pp. 86-155.
13. DOBZHANSKY, T. Heredity, environment, and evolution. *Science,* 1950, 111, 161-156.
14. FORGUS, R. H. The effect of early perceptual learning on the behavioral organization of adult rats. *J. comp. physiol. Psychol*, 1954, 47, 331-336.
15. HALDANE, J. B. S. *Heredity and politics.* New York: Norton, 1938.
16. HARRELL, RUTH F., WOODYARD, ELLA, & GATES, A. I. *The effect of mothers' diets on the intelligence of the offspring.* New York: Bur. Publ., Teach. Coll., Columbia Univer., 1955.
17. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.
18. HIRSCH, J., & TRYON, R. C. Mass screening and reliable individual measurement in the experimental behavior genetics of lower organisms. *Psychol. Bull.*, 1956, 53, 402-410.
19. HORWITT, M. K. Fact and artifact in the biology of schizophrenia. *Science*, 1956, 124, 429-430.
20. KALLMANN, F. J. *Heredity in health and mental disorder, Principles of psychiatric genetics in the light of comparative twin studies.* New York: Norton, 1953.
21. KING, J. A., & GURNEY, NANCY L. Effect of early social experience on adult aggressive behavior in C57BL10 mice. *J. comp. physiol. Psychol.*, 1954, 47, 326-330.
22. LOEVINGER, JANE. On the proportional contributions of differences in nature and in nurture to differences in intelligence. *Psychol. Bull.*, 1943, 40, 725-756.
23. LORENZ, K. Der Kumpan in der Umwelt des Vogels. Der Artgenosse als auslösendes Moment sozialer Verhaltungsweisen. *J. Orn., Lpz.*, 1935, 83, 137-213; 289-413.
24. LUCHINS, A. S., & FORGUS, R. H. The effect of differential postweaning environment on the rigidity of an animal's behavior. *J. genet. Psychol.*, 1955, 86, 51-58.
25. MELZACK, R. The genesis of emotional behavior. An experimental study of the dog. *J. comp. physiol. Psychol.*, 1954, 47, 166-168.
26. MILNER, ESTHER A. A study of the relationships between reading readiness in grade one school children and patterns of parent-child interaction. *Child Develpm.*, 1951, 22, 95-112.
27. PASAMANICK, B., KNOBLOCH, HILDA, & LILIENFELD, A. M. Socioeconomic status and some precursors of neuropsychiatric disorder. *Amer. J. Orthopsychiat.*, 1956, 26, 594-601.
28. SCHWESINGER, GLADYS C. *Heredity and environment.* New York: Macmillan, 1933.

29. Scott, J. P., & Charles, Margaret S. Some problems of heredity and social behavior. *J. gen. Psychol.*, 1953, 48, 209–230.

30. Scott, J. P., & Fuller, J. L. Research on genetics and social behavior at the Roscoe B. Jackson Memorial Laboratory, 1946–1951—A progress report. *J. Hered.*, 1951, 42, 191–197.

31. Searle, L. V. The organization of hereditary maze-brightness and maze dullness. *Genet. Psychol. Monogr.*, 1949, 39, 279–325.

32. Snyder, L. H. The genetic approach to human individuality. *Sci. Mon., N. Y.*, 1949, 68, 165–171

33. Snyder, L. H., & David, P. R. *The principles of heredity.* (5th ed.) Boston: Heath, 1957.

34. Spoerl, Dorothy T. Bilinguality and emotional adjustment. *J. abnorm. soc Psychol.*, 1943, 38, 37–57.

35. Thompson, W. R., & Melzack, R. Early environment. *Sci. Amer.*, 1956, 194 (1), 38–42.

36. Tryon, R. C. Genetic differences in maze-learning ability in rats. *Yearb. nat. Soc. Stud. Educ.*, 1940, 39, Part I, 111–119.

37. Whiting, J. W. M., et al. *Field guide for a study of socialization in five societies.* Cambridge, Mass.. Harvard Univer., 1954 (mimeo.).

38. Whiting, J. W. M., & Child, I. L. *Child training and personality: A cross-cultural study.* New Haven. Yale Univer. Press, 1953.

39. Williams, Judith R., & Scott, R. B. Growth and development of Negro infants. IV. Motor development and its relationship to child rearing practices in two groups of Negro infants. *Child Develpm.*, 1953, 24, 103–121.

40. Woodworth, R. S. Heredity and environment. A critical survey of recently published material on twins and foster children. *Soc. Sci. Res. Coun. Bull.*, 1941, No. 47.

B. F. SKINNER

Reflections on a decade of teaching machines

To THE GENERAL PUBLIC, and to many educators as well, the nature and scope of teaching machines are by no means clear. There is an extraordinary need for more and better teaching, and any enterprise which may help to meet it will not be left to develop normally. The demand for information about teaching machines has been excessive. Articles and books have been published and lectures given, symposia have been arranged, and conferences and workshops have been held and courses taught. Those who have had anything useful to say have said it far too often, and those who have had nothing to say have been no more reticent.

Education is big business. Teaching machines were soon heralded as a growth industry, and fantastic predictions of the sales of programed texts were circulated. Devices have been sold as teaching machines which were not well built or designed with any understanding of their function or the practical exigencies of their use. No author was ever more warmly received by a publisher than the author of a programed text. Many programs, to be used either with machines or in textbook form, have been marketed without adequate evaluation.

Teachers and Devices

The "mechanizing of education" has been taken literally in the sense of doing by machine what was formerly done by people. Some of the so-called computer-based teaching machines are designed simply to duplicate the behavior of teachers. To automate education with mechanical teachers is like automating banking with mechanical tellers and bookkeepers. What is needed in both cases is an analysis of the functions to be served, followed by the design of appropriate equipment. Nothing we now know about the learning process calls for very elaborate instrumentation.

Educational specialists have added to the confusion by trying to assimilate the principles upon which teaching machines are based to older theories of learning and teaching.

In the broadest sense, teaching machines are simply devices which make it possible to apply our technical knowledge of human behavior to the practical field of education *(10)*. Teaching is the expediting of learning. Students learn without teaching, but the teacher arranges conditions under which they learn more rapidly and effectively. In recent years, the experimental analysis of behavior has revealed many new facts about relevant conditions. The growing effectiveness of an experimental analysis is still not widely recognized, even within the behavioral sciences themselves, but the implications of some of its achievements

TEACHERS COLLEGE RECORD, 1963, Vol. 65, pp 168-177.

for education can no longer be ignored.

An important condition is the relation between behavior and its consequences, learning occurs when behavior is "reinforced." The power of reinforcement is not easily appreciated by those who have not had firsthand experience in its use or have not at least seen some sort of experimental demonstration. Extensive changes in behavior can be brought about by arranging so-called contingencies of reinforcement. Various kinds of contingencies are concealed in the teacher's discussions with his students, in the books he gives them to read, in the charts and other materials he shows them, in the questions he asks them, and in the comments he makes on their answers. An experimental analysis clarifies these contingencies and suggests many improvements.

Shaping by Program

An important contribution has been the so-called "programing" of knowledge and skills—the construction of carefully arranged sequences of contingencies leading to the terminal performances which are the object of education. The teacher begins with whatever behavior the student brings to the instructional situation, by selective reinforcement, he changes that behavior so that a given terminal performance is more and more closely approximated. Even with lower organisms, quite complex behaviors can be "shaped" in this way with surprising speed, the human organism is presumably far more sensitive. So important is the principle of programing that it is often regarded as the main contribution of the teaching-machine movement, but the experimental analysis of behavior has much more to contribute to a technology of education.

The direct contact which often exists between teacher and student favors the construction of programed sequences, and the teacher who understands the process can profit from the opportunity to improvise programs as he goes. Programs can be constructed in advance, however, which will successfully shape the behavior of most students without local modifications, and many of them can conveniently be mediated by mechanical devices. Laboratory studies have shown that contingencies emphasizing subtle properties of behavior can often be arranged *only* through instrumentation. There are potentially as many different kinds of teaching machines as there are kinds of contingencies of reinforcement.

Teaching machines which present material to the student and differentially reinforce his responses in well constructed programs differ in several ways from self-testing devices and self-scoring test forms, as well as from the training devices which have long been used by industry and the armed services. As Pressey pointed out many years ago (8), a student will learn while taking a multiple-choice test if he is told immediately whether his answers are right or wrong. He learns not to give wrong answers again, and his right answers are strengthened. But testing has traditionally been distinguished from teaching for good reason. Before using a self-testing device, the student must already have studied the subject and, presumably, learned most of what he is to learn about it. Tests usually occupy only a small part of his time. Their main effect is motivational. A poor score induces him to study harder and possibly more effectively. Materials designed to be used in self-testing devices have recently been programed, but the contingencies which prevail during a test are not favorable to the shaping and maintaining of behavior.

Conventional training devices arrange conditions under which students learn, usually by simulating the conditions under which they eventually perform. Their original purpose was to prevent injury or waste during early stages of learning, but attention has recently been given to programing the actual behaviors they are designed to teach. To the extent that they expedite learning, they are teaching machines. Terminal performances have usually been selected for practical reasons, but a more promising possibility is the analysis and programing of basic motor and perceptual skills—a goal which should have an important place in any statement of educational policy.

In arranging contingencies of reinforcement, machines do many of the things teachers do, in that sense, they teach. The resulting instruction is not impersonal, however. A machine presents a program designed by someone who knew what was to be taught and could prepare an appropriate series of contingencies. It is most effective if used by a teacher who knows the student, has followed his progress, and can adapt available machines and materials to his needs. Instrumentation simply makes it possible for programer and teacher to provide conditions which maximally expedite learning. Instrumentation is thus secondary, but it is nevertheless inevitable if what is now known about behavior is to be used in an effective technology.

The New Pedagogy

Any practical application of basic knowledge about teaching and learning is, of course, pedagogy. In the United States at least, the term is now discredited, but by emphasizing an analysis of learning processes, teaching machines and programed instruction have been responsible for some improvement in its status. The significance of the teaching machine movement can be indicated by noting the astonishing lack of interest which other proposals for the improvement of education show in the teaching process.

Find better teachers. In his *Talks to Teachers*, William James insisted that there was nothing wrong with the American school system which could not be corrected by "impregnating it with geniuses" *(7)*. It is an old formula. If you cannot solve a problem, find someone who can. If you do not know how to teach, find someone who knows or can find out for himself. But geniuses are in short supply, and good teachers do not come ready-made. Education would no doubt be improved if, as Conant *(3)* has repeatedly pointed out, good teachers who know and like the subjects they teach could be attracted and retained. But something more is needed. It is not true that "the two essentials of a good teacher are (a) enthusiasm and (b) thorough knowledge of and interest in his subject" *(5)*. A third essential is knowing how to teach.

Emulate model schools. Rickover's criticism of the present American school system is well-known *(9)*. His only important positive suggestion is to set up model schools, staffed by model teachers. The implication is that we already have, or at least can have for the asking, schools which need no improvement and whose methods can be widely copied. This is a dangerous assumption if it discourages further inquiry into instruction.

Simplify what is to be learned. Unsuccessful instruction is often blamed on refractory subject matters. Difficulties in teaching the verbal arts are often attributed to the inconsistencies and unnecessary complexities of a language. The pupil is taught manuscript handwriting because it more closely resembles printed forms. He is taught to spell only those words he is likely to use. Phonetic alphabets are devised to help him learn to read. It may be easier to teach such materials, but teaching itself is not thereby improved. Effective teaching would correct these pessimistic estimates of available instructional power.

Reorganize what is to be learned. The proper structuring of a subject matter is perhaps a part of pedagogy, but it can also serve as a mode of escape. Proposals for improving education by reorganizing what is to be learned usually contain an implicit assumption that students will automatically perceive and remember anything which has "good form"—a doctrine probably traceable to Gestalt psychology. Current revisions of high school curricula often seem to lean heavily on the belief that if what the student is to be taught has been "structured," he cannot help understanding and remembering it *(1)*. Other purposes of such revisions cannot be questioned. Materials should be up to date and well organized. But a high school presentation acceptable to a current physicist is no more easily taught or easily remembered than the out of-date and erroneous material to be found in texts of a decade or more ago. Similarly, the accent of a native speaker encountered in a language laboratory is no more easily learned than a bad accent. No matter how well structured a subject matter may be, it must still be taught.

Improve presentation. Pedagogy can also be avoided if what is to be learned can be made memorable. Audio-visual devices are often recommended for this purpose. Many of their other purposes are easily defended. It is not always easy to bring the student into contact with the things he is to learn about Words are easily imported into the classroom, and books, lectures, and discussions are therefore staples of education, but this is often an unfortunate bias. Audio visual devices can enlarge the student's nonverbal experience. They can also serve to present material clearly and conveniently. Their use in attracting and holding the student's attention and in dramatizing a subject matter in such a way that it is almost automatically remembered must be questioned, however. It is especially tempting to turn to them for these purposes when the teacher does not use punitive methods to "make students study." But the result is not the same. When a student observes or attends to something in order to see it more clearly or remember it more effectively, his behavior must have been shaped and maintained by reinforcement. The temporal order was important. Certain reinforcing events must have occurred *after* the student looked at, read, and perhaps tested himself on the material. But when colored displays, attractive objects, filmed episodes, and other potentially reinforcing materials are used to attract attention, they must occur *before* the student engages in these activities. Nothing can reinforce a student for *paying* attention if it has already been used to *attract* his attention. Material which attracts attention fails to prepare the student to attend to material which is not interesting on its face, and material which is naturally memorable fails to prepare him to study and recall things which are not, in themselves, unforgettable. A well prepared instructional film may appear to be successful in arousing interest in a given subject, and parts of it may be remembered without effort, but it has not taught the student that a subject may *become* interesting when more closely examined or that intensive study of something which is likely to be overlooked may have reinforcing consequences.

Multiply contacts between teacher and students. Audio-visual devices, particularly when adapted to television, are also used to improve education by bringing one teacher into contact with an indefinitely large number of students. This can be done, of course, without analyzing how the teacher teaches, and it emphasizes a mode of communication which has two serious disadvantages. The teacher cannot see the effect he is having on his students, and large numbers of students must proceed at the same pace. Contributions to pedagogy may be made in designing programs for educational television, but the mere multiplication of contacts is not itself an improvement in teaching.

Expand the educational system. Inadequate education may be corrected by building more schools and recruiting more teachers so that the total quantity of education is increased, even though there is no change in efficiency.

Raise standards. Least effective in improving teaching are demands for higher standards. We may agree that students will be better educated when they learn more, but how are they to be induced to do so? Demands for higher standards usually come from critics who have least to offer in improving teaching itself.

The movement symbolized by the teaching machine differs from other proposals in two ways. It emphasizes the direct improvement of teaching on the principle that no enterprise can improve itself to the fullest extent without examining its basic processes. In the second place, it emphasizes the implementation of basic knowledge. If instructional practices violate many basic principles, it is only in part because these principles are not widely known. The teacher cannot put what he knows into practice in the classroom. Teaching machines and programed instruction constitute a direct attack on the problem of implementation. With appropriate administrative changes, they may bridge the gap between an effective pedagogical theory and actual practice.

Educational Goals

An effective technology of teaching calls for a re-examination of educational objectives. What is the teacher's actual assignment? Educational policy is usually stated in traditional terms. The teacher is to "impart knowledge," "improve skills," "develop rational faculties," and so on. That education is best, says Dr. Hutchins *(6)*, which develops "intellectual power." The task of the teacher is to change certain inner processes or states. He is to improve the mind.

The role of the teacher in fostering mental prowess has a certain prestige. It has always been held superior to the role of the trainer of motor skills. And it has the great advantage of being almost invulnerable to criticism. In reply to the complaint that he has not produced observable results, the teacher of the mind can lay claim to invisible achievements. His students may not be able to read, but he has only been trying to make sure they wanted to learn. They may not be able to solve problems, but he has been teaching them simply to think creatively. They may be ignorant of specific facts, but he has been primarily concerned with their general interest in a field.

Traditional specifications of the goals of education have never told the teacher what to do upon a given occasion. No one knows how to alter a mental process or strengthen a mental power, and no one can be sure that he has done so when he has tried. There have been many good teachers who have supposed themselves to be working on the minds of their students, but their actual practices and the results of those practices can be analyzed in other ways. The well educated student is distinguished by certain characteristics. What are they, and how can they be produced? Perhaps we could answer by redefining traditional goals. Instead of imparting knowledge, we could undertake to bring about those changes in behavior which are said to be the conspicuous manifestations of knowledge, or we could set up the behavior which is the mark of a man possessing well developed rational power. But mentalistic formulations are warped by irrelevant historical accidents. The behavior of the educated student is much more effectively analyzed directly as such.

Contrary to frequent assertions, a behavioristic formulation of human behavior is not a crude positivism which rejects mental processes because they are not accessible to the scientific public *(12)*. It does not emphasize the rote learning of verbal responses. It does not neglect the complex systems of verbal behavior which are said to show that a student has had an idea, or developed a concept, or entertained a proposition. It does not ignore the behavior involved in the intellectual and ethical problem solving called "thinking." It does not overlook the value judgments said to be invoked when we decide to teach one thing rather than another or when we defend the time and effort given to education. It is merely an effective formulation of those activities of teacher and student which have always been the concern of educational specialists *(11)*.

Not all behavioristic theories of learning are relevant, however. A distinction is commonly drawn between learning and performance. Learning is said to be a change in some special part of the organism, possibly the nervous system, of which behavior is merely the external and often erratic sign. With modern techniques, however, behavior can be much more successfully studied and manipulated than any such inner system, even when inferences about the latter are drawn from the behavior with the help of sophisticated statistics. An analysis of learning which concentrates on the behavior applies most directly to a technology, for the task of the teacher is to bring about changes in the student's behavior. His methods are equally conspicuous. He makes changes in the environment. A teaching method is simply a way of arranging an environment which expedites learning.

Managing Contingencies

Such a formulation is not easily assimilated to the traditional psychology of learning. The teacher may arrange contingencies of reinforcement to set up new *forms* of response, as in teaching handwriting and speech or nonverbal forms of behavior in the arts, crafts, and sports. He may arrange contingencies to bring responses under new kinds of *stimulus control*, as in teaching the student to read or draw from copy, or to behave effectively upon other kinds of occasions. Current instructional programs designed to fulfill such assignments are mainly verbal, but comparable contingencies generate nonverbal behavior, including perceptual and motor skills and various kinds of intellectual and ethical self-management.

A second kind of programing maintains the student's behavior in strength. The form of the response and the stimulus control may not change, the student is simply more likely to respond. Some relevant methods are traditionally discussed under the heading of motivation. For example, we can strengthen behavior by introducing new reinforcers or making old ones more effective, as in giving the student better reasons for getting an education. The experimental analysis of behavior suggests another important possibility. Schedule available reinforcers more effectively. Appropriate terminal schedules of reinforcement will maintain the student's interest, make him industrious and persevering, stimulate his curiosity, and so on, but less demanding schedules, carefully designed to maintain the behavior at every stage, must come first. The programing of schedules of reinforcement is a promising alternative to the aversive control which, in spite of repeated reforms, still prevails in educational practice.

In neglecting programing, teaching methods have merely followed the lead of the experimental psychology of learning, where the almost universal practice has been to submit an organism immediately to terminal contingencies of reinforcement *(13)*. A maze or a discrimination problem, for example, is learned only if the subject acquires appropriate behavior before the behavior he brings to the experiment has extinguished. The intermediate contingencies are largely accidental. The differences in behavior and in rate of learning which appear under these conditions are often attributed to inherited differences in ability.

In maximizing the student's success, programed instruction differs from so-called trial-and-error learning where the student is said to learn from his mistakes. At best, he learns not to make mistakes again. A successful response may survive, but trial-and-error teaching makes little provision for actually strengthening it. The method seems inevitably committed to aversive control. For the same reason, programed instruction does not closely resemble teaching patterned on everyday communication. It is usually not enough simply to tell the student something or induce him to read a book, he must be told or must read and then be questioned. In this "tell-and-test" pattern, the test is not given to measure what he has learned, but to show him what he has not learned and thus induce him to listen and read more carefully in the future. A similar basically aversive pattern is widespread at the college level, where the instructor assigns material and then examines on it. The student may learn to read carefully, to make notes, to discover for himself how to study, and so on, because in doing so he avoids aversive consequences, but he has not necessarily been taught. Assigning-and-testing is not teaching. The aversive by-products, familiar to everyone in the field of education, can be avoided through the use of programed positive reinforcement.

Many facts and principles derived from the experimental analysis of behavior are relevant to the construction of effective programs leading to terminal contingencies. The facts and principles are often difficult, but they make up an indispensable armamentarium of the effective teacher and educational specialist. We have long since passed the point at which our basic knowledge of human behavior can be applied to education through the use of a few general principles.

Principle and Practice

The difference between general principles and an effective technology can be seen in certain efforts to assimilate the principles of programed instruction to earlier theories. Programed instruction has, for example, been called "Socratic." It is true that Socrates proceeded by small steps and often led his students through an argument with a series of verbal prompts, but the example often cited to illustrate his method suggests that he was unaware of an important detail—namely, that prompts must eventually be "vanished" in order to put the student on his own. In the famous scene in the *Meno*, Socrates demonstrates his theory that learning is simply recollection by leading an uneducated slave boy through Pythagoras's Golden Theorem. The boy responds with the rather timid compliance to be expected under the circumstances and never without help. Although Socrates himself and some of those among his listeners who were already familiar with the theorem may have understood the proof better at the end of the scene, there is no evidence whatsoever that the boy understood it or could reconstruct it. In this example of Socratic instruction, at least, the student almost certainly learned nothing.[1]

A seventeenth-century anticipation of programed instruction has also been found in the work of Comenius, who advocated teaching in small steps, no step being too great for the student who was about to take it. Programing is sometimes described simply as breaking material into a large number of small pieces, arranged in a plausible genetic order. But size of step is not enough. Something must happen to help the student take each step, and something must happen as he takes it. An effective program is usually composed of small steps, but the whole story is not to be found in Comenius's philosophy of education.

Another venerable principle is that the student should not proceed until he has fully understood what he is to learn at a given stage. Several writers have quoted E. L. Thorndike (*15*) to this effect, who wrote in 1912,

> If, by a miracle of mechanical ingenuity, a book could be so arranged that only to him who had done what was directed on page one would page two become visible, and so on, much that now requires personal instruction could be managed by print.

In commenting on this passage, Finn and Perrin (*4*) have written, ". . . Here are the insights of a genius. History can very often teach us a lesson in humility—and it does here. The interesting question is, Why couldn't we see it then?" We might also ask, why couldn't Thorndike see it then? He remained active in education for at least 30 years, but he turned from this extraordinarily promising principle to another and—as it proved—less profitable approach to educational psychology.

It is always tempting to argue that earlier ideas would have been effective if people has only paid attention to them. But a good idea must be more than right It must command attention, it must make its own way because of what it does. Education does not need principles which will improve education as soon as people observe them, it needs a technology so powerful that it cannot be ignored. No matter how insightful the anticipation of modern principles in earlier writers may seem to have been, something was lacking or education would be much further advanced. We are on the threshold of a technology which will be not only right but effective (*14*).

Criteria of Research

A science of behavior makes its principal contribution to a technology of education through the analysis of useful contingencies of reinforcement. It also suggests a new kind of educational research. Thorndike never realized the potentialities of his early work on learning because he turned to the measurement of mental abilities and to matched-group comparisons of teaching practices. He pioneered in a kind of research which, with the encouragement offered by promising new statistical techniques, was to dominate educational psychology for decades. It led to a serious neglect of the process of instruction.

There are practical reasons why we want to know whether a given method

[1] The program of the *Meno* episode constructed by Cohen (*2*) is an improvement in that the student responds with less prompting.

or instruction is successful or whether it is more successful than another. We may want to know what changes it brings about in the student, possibly in addition to those it was designed to effect. The more reliable our answers to such questions, the better. But reliability is not enough. Correlations between test scores and significant differences between group means tell us less about the behavior of the student in the act of learning than results obtained when the investigator can manipulate variables and assess their effects in a manner characteristic of laboratory research. The practices evaluated in studies of groups of students have usually not been suggested by earlier research of a similar nature, but have been drawn from tradition, from the improvisations of skillful teachers, or from suggestions made by theorists working intuitively or with other kinds of facts. No matter how much they may have stimulated the insightful or inventive researcher, the evaluations have seldom led directly to the design of improved practices.

The contrast between statistical evaluation and the experimental analysis of teaching has an illuminating parallel in the field of medicine. Various drugs, regimens, surgical procedures, and so on, must be examined with respect to a very practical question: Does the health of the patient improve? But "health" is only a general description of specific physiological processes, and "improvement" is, so to speak, merely a by-product of the changes in these processes induced by a given treatment. Medicine has reached the point where research on specific processes is a much more fertile source of new kinds of therapy than evaluations in terms of improvement in health. Similarly, in education, no matter how important improvement in the student's performance may be, it remains a by-product of specific changes in behavior resulting from the specific changes in the environment wrought by the teacher. Educational research patterned on an experimental analysis of behavior leads to a much better understanding of these basic processes. Research directed toward the behavior of the individual student has, of course, a long history, but it can still profit greatly from the support supplied by an experimental analysis of behavior.

This distinction explains why those concerned with experimental analyses of learning are not likely to take matched-group evaluations of teaching machines and programed instruction very seriously. It is not possible, of course, to evaluate either machines or programs *in general* because only specific instances can be tested, and available examples by no means represent all the possibilities; but even the evaluation of a given machine or program in the traditional manner may not give an accurate account of its effects. For example, those who are concerned with improvement are likely to test the student's capacity to give right answers. Being right has, of course, practical importance, but it is only one result of instruction. It is a doubtful measure of "knowledge" in any useful sense. We say that a student "knows the answer" if he can select it from an array of choices, but this does not mean that he could have given it without help. The right answer to one question does not imply right answers to all questions said to show the "possession of the same fact." Instructional programs are often criticized as repetitious or redundant when they are actually designed to put the student in possession of a number of different responses "expressing the same proposition." Whether such instruction is successful is not shown by any one right answer.

Correct or Educated?

A preoccupation with correct answers has led to a common misunderstanding of programed materials. Since a sentence with a blank to be filled in by the student resembles a test item, it is often supposed that the response demanded by the blank is what is learned. In that case, a student could not be learning much because he may respond correctly in 19 out of 20 frames and must therefore already have known 95 per cent of the answers. The instruction which occurs as he completes an item comes from having responded to other parts of it. The extent of this instruction cannot be estimated from the fact that he is right 19 out of 20 times, either while pursuing a program *or on a subsequent test.* Nor will this statistic tell us whether other conditions are important. Is it most profitable for the student to execute the response by writing it out, by speaking it aloud, by speaking it silently, or by reading it in some other way? These procedures may or may not have different effects on a selected "right-answer" statistic, but no one statistic will cover all their effects.

Research in teaching must not, of course, lose sight of its main objective—to make education more effective. But improvement as such is a questionable dimension of the behavior of either teacher or student. Dimensions which are much more intimately related to the conditions the teacher arranges to expedite learning must be studied even though they do not contribute to improvement or contribute to it in a way which is not immediately obvious.

The changes in the behavior of the individual student brought about by manipulating the environment are usually immediate and specific. The results of statistical comparisons of group performances usually are not. From his study of the behavior of the individual student, the investigator gains a special kind of confidence. He usually knows what he has done to get one effect and what he must do to get another.

Confidence *in* education is another possible result of an effective technology of teaching. Competition between the various cultures of the world, warlike or friendly, is now an accepted fact, and the role played by education in strengthening and perpetuating a given way of life is clear. No field is in greater need of our most powerful intellectual resources. An effective educational technology based upon an experimental analysis will bring it support commensurate with its importance in the world today.

REFERENCES

1. Bruner, J. S. *The process of education.* Cambridge: Harvard Univer. Press, 1960.
2. Cohen, I. S. Programed learning and the Socratic dialogue. *Amer. Psychologist*, 1962, *17*, 772-775.
3. Conant, J. B. *The education of American teachers.* New York: McGraw-Hill, 1963.
4. Finn, J. D., & Perrin, D. G. *Teaching machines and programmed learning. A survey of the industry, 1962.* Washington, D. C.: US Office of Education, 1962.
5. Helwig, J. Training of college teachers. *Sci.*, 1960, *132*, 845.
6. Hutchins, R. M. *On education.* Santa Barbara: Center for the Study of Democratic Institutions, 1963.
7. James, W. *Talks to teachers.* New York. Holt, 1899.
8. Pressey, S. J. A simple device for teaching, testing, and research in learning. *Sch. & Soc.*, 1926, *23*, 373-376.
9. Rickover, H. G. *Education and freedom.* New York: Dutton, 1959.
10. Skinner, B. F. The science of learning and the art of teaching. *Harvard educ. Rev.*, 1954, *24*, 86-97.
11. Skinner, B. F. Why we need teaching machines. *Harvard educ. Rev.*, 1961, *31*, 377-398.
12. Skinner, B. F. Behaviorism at fifty. *Sci.*, 1963, *140*, 951-958.
13. Skinner, B. F. Operant behavior. *Amer. Psychologist*, 1963, *18*, 503-515.
14. Skinner, B. F. *The technology of teaching* (in preparation).
15. Thorndike, E. L. *Education.* New York: Macmillan, 1912.

The Classroom as a System

Martin R. Wong and Joyce M. Wong

The words "systems," "systems approach" and "systems view" have a scientific and esoteric ring that is almost intimidating to the classroom teacher and to others who do not fully understand their meaning. In fact, however, the concept of a "system" is a relatively simple one.

Systems in General

A system, in the broad sense, is a set of parts united by some form of interaction. Examples of this broad sense would be the solar system, the digestive system and the telephone system. However, this definition is a little too broad because it leaves out what is perhaps the most vital aspect of most systems–especially educational systems, the concept of *purpose*. It is a purpose that defines the content and processes of a system, and it is with reference to purpose that most man made systems are evaluated. Therefore, a more reasonable definition of a system is *a set of parts united by some form of interaction for the attainment of a specific purpose.* The systems approach to design begins with organized common sense-organized so that all factors are taken into consideration, so that the outcomes of the process involved match the designated purposes as nearly as possible.

An example of a relatively simple system is the electric lamp. It has a collection of parts–most commonly a bulb, wire, switch, stand and shade—which are united and perform in interaction for a specific purpose, to provide light. Different kinds of lamps are designed for different purposes

Any particular light-disseminating system is part of a larger system or *suprasystem,* the electrical system, which has many such sub-systems, and has as its central purpose the provision of electric power. The lamp is also a suprasystem in the sense that it is made of sub systems such as the switch, the incandescent bulb, and the wire and plug. Each of these sub-systems also has interacting parts and a purpose.

The overall purpose of the educational system is to facilitate learning. Each sub-system is designed to act in interaction with the other parts of the system to further this purpose. Examples of sub systems are the classroom units, the administrative teams, the custodial teams, the school buildings, etc. Each sub-system has its own specific purposes which should, in interaction, further the overall purpose.

The Classroom System

Any particular classroom can be viewed as a sub-system. Its interacting parts are many: the students, the teacher, the

EDUCATIONAL TECHNOLOGY, May, 1972, pp. 56-57.

materials (blackboards, books, maps, audiovisual aids, etc.) and all the other components that serve to further the learning process.

The classroom system has as its general purpose the facilitation of learning of its individual students and may have a particular learning area that it is responsible for; e.g., seventh grade math, American history, etc. However, this general purpose is also broken up into many sub purposes in the form of specific learning objectives. At any given minute some particular sub-purpose is probably guiding the actions of the teacher in his search to stimulate and guide learning. These sub-purposes may be specifically explicated as behavioral objectives or may exist as momentary objectives in the mind of the teacher or a particular student.

System Components

In addition to its guiding purposes, a system has two other aspects, *process* and *content*. In the systems sense, the term *content* refers specifically to the components of the system. This is clearly different from the usual educational usage. In the case of the lighting system the content was the bulb, wire, switch, stand, etc. In the classroom learning system the content is the teacher, the students, the materials and whatever other aspects of the classroom and community that can be brought to bear in furthering the purpose of the system. Many classroom learning systems reach outside of the classroom and the school to bring in other interacting parts to aid in achieving their purposes. For example, many teachers use visiting guests, field trips, newspapers, television programs and whatever else will further learning. Each, while it is serving the purpose of the system, is a part of the system. Virtually anything that has use in furthering a learning objective is a legitimate potential member of the system.

The term *process* serves to describe all the operations and functions in which component parts of the system are engaged to further the purpose. In the classroom these could be reading, discussing, working problems, viewing films, solving problems, inquiring, building models, etc.

In sum, then, a system is defined by its *purpose*. The reason for analyzing a system is to try to organize it so that all of the components of the content and the processes involved are working together most effectively to further its purpose. In the case of the classroom learning system, the content and processes are many and they work together in many interacting ways to serve the general purpose of facilitating learning. Without some attempt at organization of all these components and processes, it is very likely that the system will be working at less than its best efficiency. It is to provide an organizational framework that we develop the concept of the learning system.

It is hoped that in bringing a greater degree of order to the classroom learning system, two prominent advantages will accrue. 1) the wide and almost limitless range of processes and content that potentially are available to further learning will become more clear, and 2) decisions affecting the learning process will be based on what will best serve the purpose of the system, i.e., furthering learning.

Other Components

In addition to guiding purpose, content and process, three more aspects of the systems model have direct relevance for classroom instruction. The first two are assessment of *input*, and assessment of *output*. Naturally, as with everything else in the systems model, these two assessments are made relative to the expressed purpose of the instruction. The assessment of input to the system is the necessary assessment of the students' entering behavior - their abilities, learned concepts and skills—before instruction takes place. In other words, what do they already know that will help them learn the task at hand, and what further do they need to be taught in order to be able to achieve the expressed goal? After all, it is what lies between this input and planned output that requires teaching. As soon as we find out what it is, we can plan to have it accomplished.

This pre-assessment part of the systems plan is frequently referred to as pre-testing. However, since it can be carried out in numerous ways other than specifically by testing, pre-assessment is probably a better word. The important point is that it should be conducted in a systematic way so that later assessments of output can indicate the actual (not fancied) change that has taken place. Without assessment of input, achievement due to instruction can only be assumed, not proved.

The assessment of output from the system provides either the moment supreme or the echoings of defeat, the teacher gets the glory and feeling of satisfaction derived from achievement or suffers the pangs of failure. In either case, he is better off than before, either he has a good idea about teaching procedures that lead to success and probably has some ideas about making them even better, or he knows specifically what not to do next time.

This brings us to the sixth and last of the aspects of the systems model that we will consider. the concept of modification of instructional processes based on *feedback*. In the case of *input* feedback (or more correctly feed-forward to the learning system), we've already discussed the idea that it shows us the starting point of each student, and the ground that must be covered to get to the goal, i.e., what is to be taught. *Output* feedback ideally tells us how well we've achieved our goal, what we may have done that was good, and what needs to be changed in order to more effectively reach our goal next time.

Teaching is a complex process. The most pervasive truths are variability and change. The input to the system—the students' knowledge, capability, attitudes, etc.—always contains a great deal of variability. This variability and the ever-existing need for improvement demand continual modification of the system to more effectively achieve goals. The complexities of teaching can be left unhandled and achievement can be left to serendipity, or the teacher can plan for and promote change in the desired direction. Planning for and promoting change is what the systems model is all about.

The systems model can help in planning instruction by indicating and organizing the factors that must be taken into consideration for maximally effective instruction to take place. It produces a framework for us to organize and account for the input, content, processes and output of our teaching, and helps us keep in mind how all of these interact to further our purposes. The components of the systems model seen as most relevant to classroom instructional design are purpose, process, content, input, output and system modification based on feedback.

The Systems-Oriented Teacher

What specifically would a teacher, oriented toward the systems approach, do in planning for learning that would be different from what his non-systems-oriented counterpart would do? First of all, his major concern throughout the planning and guiding of learning would be with purpose–the objectives–of instruction. He would first have to know where he was going before he could determine the best way to get there, i.e., the best coordination of the parts of the system to maximally facilitate learning. For example almost any teacher would surely plan differently for a lesson designed to teach the names of the state capitals than he would for a lesson involving the teaching of the concept "democracy." The purpose of the lesson would determine the processes as well as the content–individuals, groups, visuals, texts and other teaching tools–of the learning system.

The guiding purposes for instruction in any particular subject are usually provided for teachers in the form of syllabi, or more often, in the form of broad goals for the unit, quarter or year. However, the more specific day-to-day purposes are usually not provided. Lately there has been a great ballyhoo about the setting of behavioral objectives. Often it appears that the idea of setting objectives is furthered as an end in itself. To the systems-oriented teacher objectives are only valuable when they are based on data he has gathered about his students and when they are used as precisely stated purposes to guide the selection of the most effective learning processes and content.

The systems-oriented teacher is also aware that the purposes–objectives–of his instruction, as elements of the system, are also subject to evaluation and modification based on their usefulness to the individual and to society. Therefore, he evaluates his goals in terms of their relevance to life outside of the classroom. The system within which he is operating is part of the suprasystem called society, and it is this greater system that will examine and assess his output with relevance to its needs.

A second major concern for the systems-oriented teacher revolves around data gathering and analysis. These data are used for design modification of the system. He gathers data about his students, about the subject matter to be taught, and about the processes available to him to teach the subject matter. The first aspect–the assessment of his students–can be accomplished by pre-testing, by systematic investigation of prior test scores, by discussing with previous teachers what has been learned to this point, and by asking the students themselves. The key element is that the gathering be done systematically and accurately, for this information will become the basis for decisions regarding processes, content and even objectives. After this initial assessment, the objectives can be established or modified.

Another form of data gathering refers to the *content* and *processes* of the system. During the initial stages of planning, all sources that may lead to possible help in teaching are consulted and every idea is noted without harsh evaluation. Later, in a closer look, the many alternatives can be evaluated in terms of their likely contribution to furthering the purposes of the system. They can then be included as part of the system or rejected. Again, the decision is made with reference to the purposes of the instruction.

The last two parts of data gathering are most prominently used to evaluate and modify the system. *Ongoing* data gathering and feedback are essential for maintaining the efficiency of the system. In order that it remain maximally effective, the system must remain able to react to the variability of its content by providing flexibility in its processes. In more practical terms, feedback from students is necessary to evaulate how well the instruction is progressing, and to change it to make it better. Notice that the feedback is considered as a source of information for the furthering of objectives of the system, not as a means of checking on the value of the student. In addition, this feedback is used to keep students aware of their progress in relation to the objectives. They too, as components of the system, require feedback.

Output data gathering for feedback is also essential to the assessment of the effectiveness of the system and to the modification of the system to maintain its viability. Thus, for the systems-oriented teacher, post-testing is an important part of the instructional process. Again, the emphasis is on testing for diagnosis of strengths and weaknesses of the system as they are related to changes in the students, not on evaluating the worth of the students. It has been often shown that given appropriate instruction and sufficient time, most students can achieve the level of objectives that are normally expected in our public schools. Output evaluation, then, is to consider how well the system achieved its purposes, and what changes need to be made to improve its success. In a broad sense, this post-testing feedback tells the teacher what changes need to be made in the system to better achieve its purposes. It provides the teacher with proof positive of achievement induced by his efforts.

Summary

The systems-oriented teacher (1) thoroughly assesses the input to his system, (2) explicates his specific purposes unambiguously based on the input feedback, on constraints on his system, and on the needs of the students and the suprasystem, (3) gathers as much data as possible about his subject matter and alternative processes for achieving his purposes, (4) makes decisions concerning processes and content based on the best means of furthering the purposes, (5) activates the system by putting the plan into action, (6) gathers ongoing feedback data systematically and accurately, (7) modifies the system's content and processes based on the feedback, (8) assesses the effectiveness of the system by comparing the output product to the purposes of the system and to the input product, and (9) modifies the system based on all his sources of feedback before it is again activated.

The Behaviorally Engineered Classroom: A Learner-Sensitive Environment

Lanny E. Morreau
Program Coordinator of Instructional Design

Upper Midwest Regional Educational Laboratory

A Systems Approach

A child is an inseparable part of a social system, i.e., an ecological unit, composed of the child, his family, his neighborhood, his school, and his community (Hobbs, 1966). Just as a child affects the social eco-system, each component of this system affects the behavior of the child.

Aside from the home, the system component with the highest probable impact on the child is the school. Therefore, the effects of each variable in the school environment should be assessed and modified to provide an ideal learning environment. Emphasis should be placed on the modification of the school environment rather than the home because the school has a mandated responsibility for effective instruction, it is easier to change, and it can be structured to allow continuous monitoring of the effects of these changes.

Behavioral Management Strategies

When designing and developing the school component of the child's ecosystem, the educator must be familiar with, but cautious in accepting, theoretical, philosophical approaches. A set of empirically derived and

From Morreau, L. E.; Pinney, R. H. and Daley, M. F., INSTRUCTIONAL MANAGEMENT, 1970, Upper Midwest Regional Educational Laboratory, Minneapolis, Minnesota pursuant to U. S. Office of Education Contract 3 7 062870 3069, Department of Health, Education, and Welfare

frequently demonstrated principles can serve as the basis for the design of a functional, educational system leading to desired changes in student behavior. Such an instructional environment, based upon the principles frequently categorized under the rubric "behavior modification," would provide a high density of reinforcement contingent upon specific student responses.

Although the effectiveness of behavior modification strategies is well documented, there are several alternative procedures from which to select in establishing an instructional environment. For example, token systems have frequently been established because tokens: (1) Can often be delivered more immediately and more conveniently than direct reinforcers, (2) can be easily stored, (3) allow for the utilization of reinforcing activities which cannot be made available in the classroom setting, (4) can be employed in a large variety of situations other than the classroom, (5) can be used to manage the behavior of numerous individuals simultaneously.

But, as observed by Skinner (1968), the critical task in most teaching situations is <u>not</u> the incorporation of new reinforcers but the effective utilization of these currently available. Capitalizing on the ongoing behavior of students, the Premack framework of reinforcement offers a viable reinforcement system without using tokens. Premack noted that "...for any pair of responses, the more probable one will reinforce the less probable one (1965, p. 132)." This relationship has direct implications for instruction: Any behavior, at the point in time that it is of higher probability, can be used to reinforce any lower probability behavior. A teacher can manage an instructional environment so that student access to high probability activities (reinforcing events) are contingent on the completion of low probability activities (task behaviors). Through the management of these contingencies, the teacher can increase the frequency of selected task behaviors.

Homme, DeBaca, Devine, Steinhorst, and Rickert (1963) demonstrated the effectiveness of this technique by teaching a large segment of the first-grade repertoire to a group of preschool children in one month. Equally sigificant was the change from a condition of bedlam which existed in the classroom to a situation highly conducive to learning. Further indications of the effectiveness of contingency management techniques were found in the teaching of English literacy to Indian children (Homme, 1965) and in the teaching of arithmetic and reading to mentally retarded children (Daley, Holt, and Vajanasoontorn, 1966).

It has also been established that low probability behaviors can, through repeated pairings with reinforcers (high probability behaviors), become reinforcing in themselves (Daley and Holt, 1969). This data provides a direction for attaining one of the ultimate aims of the ideal educational system: To offer as great a quantity and quality of reinforcement as possible while systematically increasing the frequency of low probability behaviors associated with academic tasks until the behaviors associated with the tasks assume the properties of reinforcing events. This synopsis, in fact, describes the one continuous, operational objective of the behaviorally engineered classroom (Homme, Csanyi, Gonzales, and Rechs, 1969; Morreau and Daley, 1972).

Classroom Design

The behaviorally engineered classroom is structurally defined by three major areas: The task area, the progress check area, and the RE area.

The Task Area

Since the primary goal in the classroom is to increase the frequency and/or accuracy of the students' academic responses, the largest portion of the classroom is designed for academic-task related activities. In the task area, students receive materials selected to assist them in working toward their individual objectives. The size and difficulty of the task are based on the student's entry-level (existing) skill in each specific area. The teacher moves from student to student providing verbal reinforcement for specific behaviors and assistance where needed. The effectiveness of such teacher interaction in reducing disruptive behavior while reinforcing academic responses has been clearly demonstrated (Hall, Lund, and Jackson, 1968). The task area also provides a situation in which the teacher can monitor and record student behavior which ultimately provides part of the data for decision-making related to individual students.

<u>The Progress Check Area</u>

If students are reinforced for simply completing tasks without evaluation of their responses, inaccurate responding may increase in frequency. The progress check area allows for the prompt, systematic evaluation of student tasks so both the teacher and the students know if the criteria specified for a given task have been met. The student, an aide, or a classroom assistant can record the data relevant to task performance, thereby providing the core information required for decisions related to future tasks and contingencies.

<u>The Reinforcing Event Area</u>

In the reinforcing event area, students engage in activities of their choice. Students can create activities or select them from pictures of available activities, a reinforcing event menu (Fig. 1). The tendency to cooperatively engage in group activities indicates that this indirectly-supervised area provides a setting facilitative to the meeting of social

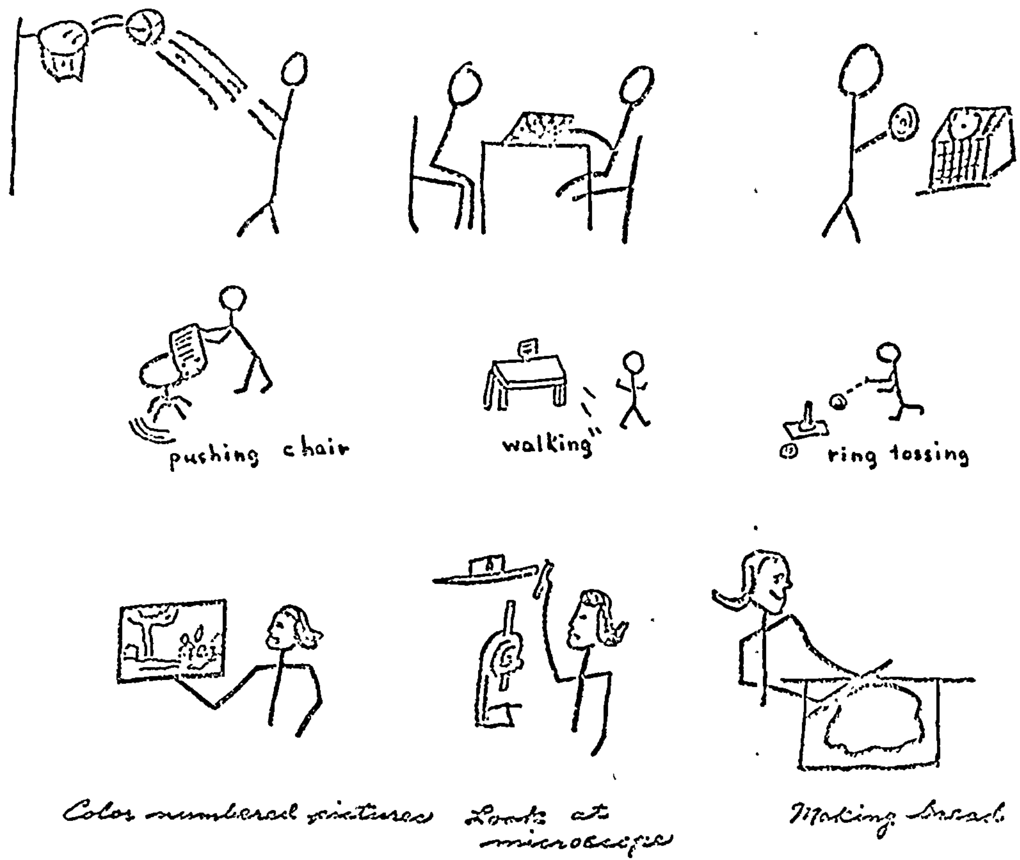


Figure 1. Samples from student-developed RE menus.

objectives.

Classroom Sequence

Students proceed through the following sequence of events in the classroom: The student completes a task unit, a series of small behavioral steps which approximate desired terminal performance, and proceeds to the aide for evaluation. If the student performs successfully, i.e., at the prespecified criterion level, he selects from activities pictured on the RE menu and proceeds to the RE area to engage in the activity of his choice for five to ten minutes. If the task performance does not meet the specified criterion level, the student returns to his desk, repeats only those responses where errors were made, returns for a re-evaluation, and then proceeds to the RE area (Fig. 2).

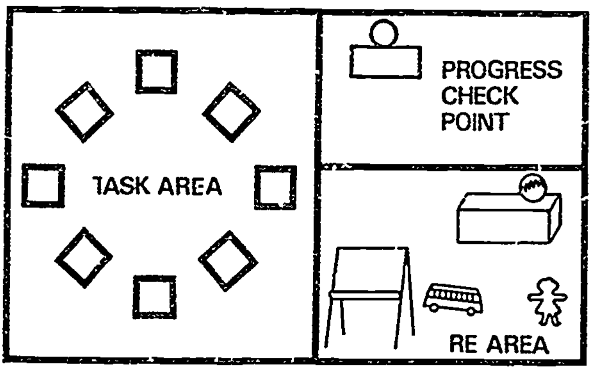


Figure 2. Components of the composite classroom.

The formal agreement between the teacher and the student as to the quantity or quality of task-related material which will be completed and the positive consequences which will be engaged in contingent upon its completion is referred to as a contract. The arrangement of liberal and fair contracts based on individual needs, the clearly delineated expectations implicit in their development, and the subsequence adherence to the contracts by the teacher, provide ideal circumstances for the development of trust.

Environmental Characteristics

It is frequently easier to structure the physical environment than it is to modify behavior through the management of contingencies. For a disruptive student, the teacher might first develop a prosthetic environment and gradually alter it to approximate the task area. For example, a student could complete his task units in a study booth which could be systematically reduced in height as disruptive behaviors decreased in frequency.

Ideally, emphasis is placed on decreasing the frequency of problem behavior by systematically increasing the density of reinforcement for task-related behaviors--a nonpunitive procedure. Punishment often elicits negative responses which can become chronic. The student may: (1) Attempt to escape from the punishment, e.g., withdraw, (2) attack the deliverer of punishment, (3) attack his peers, (4) attempt to destroy the environment.

The nearest approximation to punishment contingencies used in the behaviorally engineered classroom is time-out, i.e., time away from the opportunity to engage in activities leading to reinforcing events. Time-out should be short. Further, a response incompatible to the one for which time-out is presented should be strengthened through systematic reinforcement. The student should receive an explanation of why he is being removed and should be told that he will again be able to engage in reinforcing activities.

Instructional strategies must also be developed which are consistent with the environment. The developmental strategy would be similar to the seven-step sequence described by Bijou (1968):

1. State the desired terminal behavior in objective terms.

2. Assess the student's behavioral repertoire relevant to the task.

3. Arrange stimulus material and the behavioral criteria for reinforcement in a sequence for presentation.

4. Start the student on the unit in the sequence at which he can respond correctly 90% of the time or better.

5. Manage the contingencies of reinforcement to strengthen successive approximations of the terminal behavior.

6. Build conditioned reinforcers that are intrinsic to the task learned.

7. Keep records of the student's progress and modify the program accordingly.

To develop instructional strategies of this type one must first state the objectives for instruction. These objectives can be stated clearly and distinctly through the use of a simplified unit guide (Morreau, 1970). The precision necessary for the evaluation of instructional outcomes is provided by the behaviorally objectives which specify: (1) The learner as differentiated from a class or group of learners, (2) an observable, measurable response which the learner will emit or a product of the response, (3) the stimulus conditions under which the response will occur, (4) the frequency, duration, or quality of the performance (the criterion level) which will be considered successful.

Each of the four parts of the objective provides specific information:

"When presented with a random list of 10 words...

The Condition: The student could be faced with a disaster if all his activities were directed toward a written response to printed material and the teacher measured his progress with an oral check. The conditions indicate the procedures to be used in evaluation and guide the teacher in selecting materials. If mastery of a given skill is to be demonstrated by oral responding, the teacher would record this in the objective and then select activities to direct the student toward oral behavior.

...the learner...

The Learner: "The learner," as contrasted to a class or group of learners, focuses attention on the individual student. It cannot be assumed that a "class" is a homogeneous group of individuals who are progressing at an equal pace and who can all benefit from the same materials.

...will write...

The Action: Behavioral objectives specify what the student can do, not what he "knows." Because the action is observable and measurable, teachers can precisely (objectively) evaluate the effects of materials in assisting the student in mastering the objective. The student's measured performance also serves as a guide for advancing him in a sequence of objectives.

...3 rhyming words for each word with 100% accuracy."

The Measure or Criteria. Stating the precise quantity or quality of responses which indicate mastery of a given skill assures that each student is advanced to more difficult materials only after he has the skills necessary for dealing with them. Teachers are also able to report a student's progress more precisely. He doesn't just "know more;" he can "do" a specific task at a specified level of accuracy.

Objectives should be directed toward actual, life functioning. In other words, the terminal behavior in a given area should occur in a "real-life," stimulus condition rather than in a classroom situation. The student should learn to respond in the presence of a variety of stimuli and to generalize a given response to "real-world" situations. Both the objectives and the curriculum should reflect the need for group activities to meet specific social objectives.

Contracts should be prepared for all curriculum areas such that the student can engage in reinforcing activities for successful task completion both in and outside the classroom, e.g., physical education, art, and a variety of curriculum materials should be made available by which any given objective can be met.

Assessment of the student's present level of performance can be accomplished through the use of curriculum embedded tests or other measures which are consistent with the objectives to be used.

Selecting Objectives and Materials

The sequential arrangement and the storage of objectives, i.e., an objectives bank, would assure that:

1. Students would be placed precisely within a given sequence of objectives.

2. Prerequisite skills for meeting terminal objectives would not be omitted.

3. The pupil would be directed toward increasingly complex discriminations.

4. The required generalizations would occur.

An effective banking system for objectives would meet four specific criteria:

1. All objectives would be sequentially arranged.

2. All objectives would be complete (including student-action-measure/criteria-condition).

3. Each objective would contain instructions for teacher use of the objective itself.

4. Specific units of curriculum material would be tied to each objective.

Following individualized placement within the instructional sequence, the student would receive task units composed of programmed, instructional materials designed to meet a specific objective. Such programmed materials offer several advantages:

1. Each student can proceed at his own rate.
2. The materials allow for self-instruction.
3. Each student can use different materials.
4. Teachers can spend more time with individuals in that students are functioning independently.
5. Teachers need not make all subject-area presentations.
6. The materials are designed to systematically guide student responding.
7. The materials maintain high levels of motivation via reinforcement and feedback.
8. Students can complete units in less time.

Teacher Decision-Making

To design an environment responsive to the specific needs of individual students, the teacher must collect and analyze data on a systematic basis. Although it might appear that all decisions are dictated by the environment, the environment provides only a structure-- the teacher is a crucial part of the classroom, and student progress will be largely determined by the effectiveness of the on-going, decision-making process.

The objectives, the materials, and the classroom are structured to accommodate individual differences. Since these differences may be great, the teacher cannot rely on <u>group</u> information but needs continuous information on <u>each</u> student's performance to successfully assist students on an individual basis. Four basic areas of teacher decision-making can affect individual progress:

Diagnosing the Skill/Knowledge Level

Each student's position on a continuum of behavioral objectives is located by means of an objective-based placement examination. It is possible (and highly probable) that the initial placement for some students will be incorrect. Decisions: Is the diagnosis correct? Should a given student be advanced or otherwise relocated in the objectives sequence?

Evaluating Materials and Activities

If materials and activities are to help students meet their objectives, they must be selected in terms of a specific objective to be met. The materials may appear to be relevant to a given objective, but the only real criterion for their use is the degree to which they will improve the terminal performance by the student. Decisions: Are the materials and activities adequate for meeting specific behavioral objectives? Should they be modified or replaced?

Prescribing to Correct a Skill/Knowledge Deficiency

After the teacher has diagnosed each student's skill/knowledge level, he prepares a prescription, a unit of task material that is neither too large nor too difficult for the individual student. Decisions: Is the task too large or too difficult for the student? Should the task size be reduced? Should the student first use other less difficult materials? These decisions can be facilitated by assigning a small task to each student and monitoring his performance on the task.

Selecting and Evaluating Reinforcing Events

Students will select their favorite activities from those offered in the RE area. But, it can never be assured that, at any given time, a reinforcer is available for each child. Further, the same item used as a reinforcing activity for one student may simultaneously be requested by several other students. Decisions: Is a reinforcing activity available for each child? Should more than one of a given item be present in the RE area?

Since the opportunity to engage in a reinforcing event increases or maintains student behavior, reinforcing activities must be assessed frequently in terms of their effect on student performance. Given continuous performance data on each student, for example, the teacher can decide to add items to the RE area based on the frequency of selection by the students.

To make these decisions, the teacher must record the following information: How many errors does a student make on each task? How much time does it take for the student to complete each task? How many reinforcing events are received by each student? Which reinforcing events are selected by the students? How many students are making a large number of errors on a given unit of material?

Recording and analyzing data allows for effective decision-making by both the teacher and the student. Students can learn the procedures necessary for independent decision-making and self-management of both task and reinforcement selection--certainly a "real-life" situation.

Meeting Learner Needs

The following comparison of the behaviorally engineered classroom with several criteria identified as significant in Project REED (Hobbs, 1966), illustrates the sensitivity of the classroom to learner needs:

1. Life is to be lived now. The classroom is designed to increase the probability of immediate success through the clear specification of objectives for each student and through the use of successive approximations toward a terminal behavior.

2. Time is an ally. Because the classroom is sensitive to individual differences, children functioning at varying levels and having a variety of problem behaviors can be accommodated in the same environment.

3. Trust is essential. The classroom provides for the meeting of contracts by both the adult and the student. What better opportunity for the development of trust?

4. The child must develop competence. Each academic task is designed for a specific student: The task size and the placement in the materials is individualized. The child develops competence while receiving multiple reinforcers for task completion. Further, reinforcing events are related to the real world. Contingencies are very much like those arranged for competent performance in the adult world.

5. Problem behaviors can and should be controlled. Through the reinforcement of task-related responses, there is a high probability that incompatable problem behaviors will decrease in frequency.

6. Cognitive control can be taught. In this classroom, students rapidly learn discriminative stimuli for specific behavior through the differential reinforcement of behavior.

7. Feelings should be nurtured. Students have an opportunity to express anger and hostility in a non-punitive environment. While the environment does not provide positive reinforcement for disruptive behavior, it is not consequated with punishment.

8. The group is important. The RE area, where students join in group activities under reinforcing conditions, provides a functional model for learning group importance. Specific objectives are also prepared and implemented for instructing students on the importance of the group.

9. Communities are important. Through the interaction of the student with his teacher, his family, and his "peer" community, he learns to emit specific terminal behaviors which carry over to his neighborhood, his town, or his city. He is given an opportunity to behave in a manner consistent with the significant social behavior of his community while being reinforced for that behavior.

10. A child should know joy. Joy could be defined as having personal control over access to a large amount of positive reinforcement, experiencing success on a frequent basis in many areas, interacting socially in a "free" setting, and receiving few aversive consequences.

These tenents define the parameters of a learner-sensitive environment; so, too, do they describe the behaviorally engineered environment.

REFERENCES

Bijou, S. W. What psychology has to offer education - now. Paper presented at the 76th Annual Convention of the American Psychological Association, San Francisco, September, 1968.

Daley, M. F. and Holt, G. Systematic probability reversal and control of behavior. Submitted to Journal of Experimental Child Behavior, 1969.

Dal , M. F., Holt, G., and Vajanasoontorn, C. Reinforcement menus in the instruction of mentally retarded children. Paper presented at the Conference on Instructional Methods and Teacher Behavior, Berkeley, November, 1966.

Hall, R. V., Lund, D., and Jackson, D. Effects of teacher attention on study behavior. Journal of Applied Behavior Analysis, 1968, 1, 1-12.

Hobbs, N. Helping disturbed children: Psychological and ecological strategies. American Psychologist, 1966, 1105-1115.

Homme, L. E. A system for teaching English literacy to pre-school Indian children. Westinghouse Research Laboratories Contract 14-20-06500-1506. Bureau of Indian Affairs, Final Report, October 11, 1965, 1-15.

Homme, L., Csanyi, A. P., Gonzales, M. A., and Rechs, J. K. How to use contingency contracting in the classroom. Champaign, Illinois: Research Press, 1969.

Homme, L., DeBaca, P., Devine, J., Steinhorst, R., and Rickert, E. Use of the Premack principle in controlling the behavior of nursery school children. Journal of the Experimental Analysis of Behavior, 1963, 6, 544.

Morreau, L. E. A curriculum for every child. Behaviorally engineered environments - staff papers for AERA symposium. Presented at the meeting of the American Educational Research Association Convention, Minneapolis, March, 1970.

Morreau, L. E. and Daley, M. F. Behavioral management in the classroom. New York: New Century Series, Appleton-Century-Crofts, 1972.

Morreau, L. E., Pinney, R. H., and Daley, M. F. Instructional Management. Minneapolis, Minnesota, November, 1970.

Premack, D. "Reinforcement theory." In David Levine (Ed.) Nebraska Symposium on Motivation. Lincoln, Nebraska: University of Nebraska Press, 1965, 123-188.

Skinner, B. F. The technology of teaching. New York: Appleton-Century-Crofts, 1968.