DOCUMENT RESUME

ED 030 987 24 EA 002 473

By-Messick, Samuel

The Criterion Problem in the Evaluation of Instruction: Assessing Possible Not Just Intended Outcomes.

California Univ., Los Angeles. Center for the Study of Evaluation of Instructional Programs.
Spons Agency-National Inst. of Mental Health (DHEW), Bethesda, Md.; Office of Education (DHEW), Washington, D.C. Bureau of Research.
Report No-CSE-R-22; NIMH-RG-MH-4186
Bureau No-BR-6-1646
Pub Date May 69
Contract-OEC-4-6-061646-1909
Note-38p.; Paper presented at the Symposium on Problems in the Evaluation of Instruction (Los Angeles, December, 1967).
EDRS Price MF-$0.25 HC-$2.00
Descriptors-Affective Behavior, Bibliographies, *Cognitive Development, *Cognitive Processes, Educational Objectives, Educational Research, *Evaluation Criteria, *Individual Differences, *Instructional Programs, Program Evaluation, Values

Cognitive styles--defined as information processing habits--should be considered as a criterion variable in the evaluation of instruction. Research findings identify the characteristics of different cognitive stles. Used in educational practice and evaluation, cognitive styles would be new process variables extending the assessment of mental performance past achievement levels to a concern with patterns of cognitive functioning. Such use would have relevance for student performance in various subject-matter areas, for test construction, and for teaching methods. A possible educational goal would be an attempt to foster alternative modes of cognition and multiple stylistic approaches to problem solving. A bibliography of 48 items is included. A related document is EA 002 406. (MLF)

The CENTER FOR THE STUDY OF EVALUATION OF INSTRUCTIONAL PROGRAMS is engaged in research that will yield new ideas and new tools capable of analyzing and evaluating instruction. Staff members are creating new ways to evaluate content of curricula, methods of teaching and the multiple effects of both on students. The CENTER is unique because of its access to Southern California's elementary, secondary and higher schools of diverse socio-economic levels and cultural backgrounds.

BR-6-1646
PA-24
OE-BR
OEC-4-6-061646-1709

THE CRITERION PROBLEM IN THE EVALUATION OF INSTRUCTION: ASSESSING POSSIBLE NOT JUST INTENDED OUTCOMES

Samuel Messick

Educational Testing Service

From the Proceedings of the

SYMPOSIUM ON PROBLEMS IN THE EVALUATION OF INSTRUCTION

University of California, Los Angeles
December, 1967

M. C. Wittrock, Chairman

Sponsored by the Center for the
Study of Evaluation

The research and development reported here in was performed pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education under the provisions of the Cooperative Research Program.

CSE Report No. 22, May, 1969
University of California, Los Angeles

THE CRITERION PROBLEM IN THE EVALUATION OF INSTRUCTION: ASSESSING POSSIBLE NOT JUST INTENDED OUTCOMES[1]

Samuel Messick

This paper will discuss cognitive styles and affective reactions as two major classes of criterion variables that should be taken into account in the evaluation of instruction. These two types of variables are emphasized because of their bearing upon questions that should be asked in evaluation studies--questions that stem from particular views about the diversity of human performance and about the role of values in educational research.

Individual Differences in Response to Educational Treatments

Traditional questions in education and psychology have frequently spawned answers that are either wrong since they summarize findings "on the average" in situations where a hypothetical "average person" does not exist, or else are seriously lacking in generality because they fail to consider the multiplicity of human differences and their interactions with environmental circumstances.

Consider the kind of "horse race" question typical of much educational research of the past: Is textbook A better than textbook B? Is teacher A better than teacher B? Or, more generally, is treatment A better than treatment B? Such questions are usually

---

[1]The preparation of this paper was supported in part by the National Institute of Mental Health under Research Grant MH-4186.

resolved empirically by comparing average gains in specific achievement for students receiving treatment A with average gains for students receiving treatment B. But suppose treatment A is better for certain kinds of students and treatment B better for other kinds of students. Depending upon the mix of students in the two groups, the two treatments might exhibit negligible differences on the average while producing wildly different effects upon individuals. An entirely different evaluation of the treatments might have resulted if some other questions had been asked, such as "Do these treatments interact with personality and cognitive characteristics of the students or with factors in their educational history or family background to produce differential effects upon achievement? Do certain student characteristics correlate with gains in achievement differently in one treatment than in the other?"

From the vantage point of differential psychology, it would appear that educational researchers frequently fail to take proper account of consistent individual differences. They tend to assess treatment effects on the average, presuming that variations in performance around the average are unstable fluctuations rather than expressions of stable personal characteristics. Developmental psychologists, on the other hand, survey essentially the same arena with their own limited purview. They not only frequently make the same assumption about individual variation but also the obverse

concerning environmental variation. They seek to uncover for the generic human being general laws of learning and cognition--at best a small number of different laws for assorted idealized types of individuals--and to delineate mental development on the average, where the average is taken over all the differential educational experiences and environmental impacts that might interact with current psychological status to moderate change.

To evaluate educational treatments in terms of their effects upon individual students requires not only the assessment of variables directly related to specific treatment goals, such as achievement level, but also the assessment of personal and environmental variables that may moderate the learning. Similarly, to formulate the psychology of the development of cognitive or personality characteristics over a fixed period may require information not only about individual differences in the trait in relation to other traits at different times, but also about the educational treatments and environmental variations accompanying the change. Information about the trait's previous development and the personal, social, and environmental factors associated with prior growth may also be necessary.

If concerns about personal, social, and environmental characteristics were systematically combined with concerns about the effects of educational treatments, a conceptual framework for educational and psychological research would result, stimulating questions about

interactions among these components such as, "What dimensions of educational experience are associated with growth on dimensions of cognitive functioning or with changes in attitude or affective involvement, and what social and environmental factors moderate these effects?" The need for such a multivariate interactional approach derives from the view that in education and psychology we are dealing with a complicated system composed of differentiated subsystems; even in research on presumably circumscribed issues it is important to recognize the interrelatedness of personal, social, environmental, and educational factors. In such a system it is possible that compensating trade-offs among variables will occur under different conditions to produce similar effects, and that particular outcomes will frequently be multiply determined. This is not to say that overall main effects due the specific educational treatments will not occur or that no personal characteristics will prove to be general over situations, but rather that interactions between treatment variables and personal or environmental factors are probable and should be systematically appraised in evaluating treatment effects.

The major thrust of this approach is that evaluations of the significance of changes in performance or attitude over a given time period as a presumed function of a specific instructional program should consider other changes in human characteristics and

environmental influences active at the same time. Educational growth should not be viewed as independent of human growth, and the effects of instructional experiences should not be viewed as independent of other life experiences.

These multiple influences upon behavior should not only be considered at the level of systems analysis, but also at much simpler levels--such as in developing and evaluating a measure of academic achievement--where we sometimes forget that even specific responses are frequently complexly determined and buffeted by many environmental influences. Consider a researcher who attempts to assess quantitative reasoning in a lower class, culturally disadvantaged child by inquiring, "If you had seven apples and I asked you for two, how many would you have left?" The answer comes quickly and triumphantly--"Seven!" Hopefully, of course, we would never use such loose phrasing in our questions, but the example illustrates the point. We often fail to appreciate the extent to which the respondent's affect will be engaged by the content of a question and the extent to which personal, social, and economic factors will focus his attention upon problems quite different from the ones we thought we had posed.

When the efficacy of instruction is evaluated in such a multivariate framework, cognitive styles and affective reactions assume particular interest: (1) as personal characteristics that may interact with treatment variables to moderate learning, retention,

and transfer; (2) as dispositions to be monitored to detect any possibly undesirable side effects of instruction; and (3) as qualities to be fostered either directly as specific objectives of the instructional program or indirectly as by-products of other efforts. This latter possibility of fostering stylistic and affective qualities appears to be consonant with general educational aims and the desirability of developing positive attitudes toward school, learning, subject matter or self. But with respect to cognitive styles there is much less consensus, for we are not sure whether to emphasize particular styles or flexibility in the use of multiple styles, nor are we sure what the options are for changing styles. This problem will be discussed in more detail after we have considered the nature of cognitive styles and some reasons why individual differences in characteristic modes of cognition are relevant to educational practice.

The Role of Values in the Science of Education

To suggest that cognitive styles and affects might serve as additional criteria in the evaluation of instruction is a value judgment. But value judgments abound in the evaluation process, and appear to be made with hesitancy only at the end of the enterprise when a decision about the work of the program is required. Value judgments are usually made explicitly when the specific goals

of the instructional program are outlined and when particular standards of excellence are accepted for judging success. But they are also made, usually implicitly, when criterion instruments are selected to assess the intended outcomes, when additional criterion measures are chosen to appraise side effects, when particular teaching methods, media or materials are scrutinized during the course of instruction, and when certain types of transactions between the student and other persons are observed (Stake, 1967, pp. 68, 523-540)--in short, whenever a subset of the possible alternatives is marked for special attention.

The selection of a subset from the range of possibilities implies priorities--that some things are more important to assess than others. But it is not enough to label such decisions "value judgments" and then proceed with the assessment. If it were, evaluation would be a straightforward affair indeed; we could specify the goals of the instructional program as we intend them and select criterion measures to assess those outcomes that seem directly relevant to the stated objectives. This is what Scriven has called "estimation of goal achievement" in contradistinction to evaluation proper. All appraisal in this case is relative to the stated goals, and the concern is with how well the program achieves its intended objectives. In addition, however, we should inquire to what extent the objectives are worth achieving

and, in general, should endeavor to include in the evaluation process provisions for evaluating the value judgments, especially the goals (Scriven, 1967).

An important step in this direction is to be concerned with possible as well as intended outcomes. Evaluation comprises two major functions--to ascertain the nature and size of the effects of the treatment and to decide whether the observed effects attain acceptable standards of excellence. These two components have been termed "description" and "judgment" by Stake (1967). The point here is that the descriptive phase of evaluation should be as complete as our art and resources allow. In this instance the evaluation specialist should be, in Bruner's words (1966) a "diviner and delineator of the possible"--he should "provide the full range of alternatives to challenge society to choose." This attempt to describe the full range of possible effects of instruction is an important prerequisite for the judgmental phase of evaluation, since it might unearth alternatives that ought to be weighed in reaching the final appraisal. As Henry Dyer (1967 pp. 12-24) has emphasized, "Evaluating the side effects of an educational program may be even more important than evaluating its intended effects." Dyer (1967) also pointed out that such broad assessment of the possible effects of an educational program should contribute to an evaluation of its goals. Inverting the customary

prescription that one must determine the objectives of instruction before developing measures of instructional outcomes, Dyer suggested that it may not be possible to decide what the objectives ought to be until the outcomes are measured.

In practice, of course, evaluation studies rarely approach completeness. We include in any feasible assessment program only a selection of criterion variables--those that reflect our current view of priorities or our attempt to represent several diverse viewpoints. Again it is not enough just to admit that practical considerations demand selectivity. To develop a science of evaluation, we should endeavor to justify these value judgments on rational grounds in terms of the specific objectives of the instructional program in question and of goals of education that transcend the particular course (Scriven, 1967). It is important not only to explicate the separate value judgments implicit in the choice of each criterion variable, but also to consider interrelations among them. Values rarely exist in isolation. They are typically part of ideologies that provide characteristic ways of thinking about man and society. In considering the assortment of variables to be assessed in a particular evaluation study and the goals that the instruction might serve, we should inquire to what extent the possible outcomes reflect divergent value systems that need to be reconciled or compromised and to what extent do they represent

simply different frames of reference for compatible goals" (Proposal for a Research & Development Center, 1965).

Incidentally, the particular teaching methods chosen for an instructional program should also be evaluated for their compatibility with multiple goals and values. Even though two goals are reasonably compatible, the method of instruction may foster one aim and hinder the other. Wallach, for example, is concerned that modern methods of teaching, especially those using programmed materials and teaching machines, so emphasize accuracy of responding that the student is likely to acquire a generalized intolerance of error and consequent decline in his originality of thinking. Some other method or combination of methods might be used to develop facility in the analysis of logical implications without diminishing fluency in the generation of conceptual possibilities (Wallach, 1967, pp. 36-57).

Since educational values derive from broader systems of social values, it is appropriate to evaluate goals and criteria for instruction not only in terms of specific educational implications but also in terms of more general social implications. The suggestion that cognitive styles and affective reactions be used as criterion variables in the evaluation of instruction, for example, should be upheld in precisely such terms, but a consideration of the educational and social implications of these dimensions must await a more detailed discussion of the nature of the variables themselves.

Cognition, Affect, and Personality

In recent years we have seen the isolation of several dimensions of individual differences in the performance of cognitive tasks that appear to reflect consistencies in the manner or form of cognition, as distinct from the content of cognition or the level of skill displayed in the cognitive performance (Thurstone, 1944; Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962; Gardner, Holzman, Klein, Linton, & Spence, 1959; Gardner, Jackson, & Messick, 1960). These dimensions have been conceptualized as cognitive styles, which represent a person's typical modes of perceiving, remembering, thinking, and problem solving. Some examples of these dimensions are:

(1) Field independence vs. field dependence--an analytical, in contrast to a global, way of perceiving (which) entails a tendency to experience items as discrete from their backgrounds and reflects ability to overcome the influence of an embedding context. (Witken et al., 1962).

(2) Scanning--a dimension of individual differences in the extensiveness and intensity of attention deployment, leading to individual variations in the vividness of experience and the span of awareness (Holzman, 1966, pp. 835-844; Schlesinger, 1954, pp. 354-374; Gardner & Long, 1962, pp. 129-140).

(3) Breadth of categorizing--consistent preferences for broad inclusiveness, as opposed to narrow exclusiveness, in establishing the acceptable range for specified categories (Pettigrew, 1958, pp. 532-544; Bruner & Tajfel, 1961, pp. 231-241; Kagan & Wallach, 1964).

(4) Conceptualizing styles--individual differences in the tendency to categorize perceived similarities and differences among stimuli in terms of many differentiated concepts, which is a dimension called conceptual differentiation (Gardner & Schoen, 1962; Messick & Kogan, 1963, pp. 47-51) as well as consistencies in the utilization of particular conceptualizing approaches as bases for forming concepts--such as the routine use in concept formation of thematic or functional relations among stimuli as opposed to the analysis of descriptive attributes or the inference of class membership (Kagan, Moss, & Sigel, 1960, pp. 261-278; Kagan et al., 1963, pp. 73-112).

(5) Cognitive complexity vs. simplicity--individual differences in the tendency to construe the world, and particularly the world of social behavior, in a multidimensional and discriminating way. (Kelly, 1955; Bieri, 1961; Bieri, Atkins, Scott, Leaman, Miller, & Tripodi, 1966; Scott, 1963; Harvey, Hunt, & Schroder, 1961).

(6) Reflectiveness vs. impulsivity--individual consistencies in the speed with which hypotheses are selected and information processed, with impulsive subjects tending to offer the first answer that occurs to them, even though it is frequently incorrect, and

reflective subjects tending to ponder various possibilities before deciding (Kagan, Rosman, Day, Albert, & Phillips, 1964; Kagan, 1965, pp. 609-628).

(7) Leveling vs. sharpening--reliable individual variations in assimilation in memory. Subjects at the leveling extreme tend to blur similar memories and to merge perceived objects or events with similar but not identical events recalled from previous experience. Sharpeners, at the other extreme, are less prone to confuse similar objects and, by contrast, may even judge the present to be less similar to the past then is actually the case (Holzman, 1954, pp. 375-394; Holzman & Klein, 1954, pp. 105-122; Gardner et al., 1959).

(8) Constricted vs. flexible control--individual differences in susceptibility to distraction and cognitive interference (Klein, 1954, pp. 225-274; Gardner et al., 1959).

(9) Tolerance for incongruous or unrealistic experiences--a dimension of differential willingness to accept perceptions at variance with conventional experience (Klein, Gardner, & Schlesinger, 1963, pp. 41-55).

Stylistic consistencies have also been observed in the differential tendencies of individuals to err by omission or by commission on memory tasks (McKenna, 1967). In addition, several dimensions deriving from the work of Thrustone, Cattell, and Guilford, which are usually considered to fall within the purview of intellectual abilities, also reflect such potential exemplars of style or

mode of cognition as speed, flexibility, divergence, convergence, and fluency.

Cognitive styles, for the most part, are information-processing habits. They are characteristic modes of operation which, although not necessarily completely independent of content, tend to function across a variety of content areas. Before considering some possible implications of cognitive styles for educational practice, let us discuss one in more detail to illustrate its generality and breadth of operation. For this purpose the dimension of analytic vs. global attitude offers the best example, since it has been extensively studied in various forms by H. A. Witkin and others.

Witkin's early work emphasized individual differences in the characteristic ways in which people perceive both the world and themselves. One of the test situations used was a tilted room in which the subject, seated in a tilted chair, must adjust his body to the true upright. Reliable individual differences were found in this ability; i.e., some individuals were reliably more susceptible than others to the influence of the surrounding tilted room. In another test, the subject was seated in a completely dark room and confronted with a luminous rod surrounded by a luminous picture frame; his task was to set the rod to the true vertical position while the frame was set aslant. Again, reliable individual differences were found in this ability, and a substantial correlation

was noted between the two tests; the subjects who had difficulty withstanding the influence of the surrounding room while adjusting their body to the upright also had difficulty withstanding the influence of the surrounding frame while adjusting the rod to the upright. These individual differences were initially conceptualized in terms of a differential reliance upon visual cues obtained from the external field as opposed to kinesthetic cues obtained from the subject's own body.

This interpretation of field vs. body orientation was extended to a more general dimension of perceptual analysis, however, when it was found that subjects who had difficulty overcoming the influence of the titlted room and the tilted frame also had difficulty overcoming the influence of superimposed complex designs when asked to find hidden simple forms in an embedded-figures test. This extended conception of the dimension was now termed "field dependence vs. field independence;" the perception of relatively field-dependent subjects is dominated by the overall organization of the field, whereas relatively field-independent subjects readily perceive elements as discrete from their backgrounds. Sex differences have been repeatedly obtained on the measures of this dimension, with females being relatively more field dependent and males relatively more field independent (Witkin, et al., 1954).

Since many correlates for these perceptual scores have been subsequently uncovered in several areas of intellectual and personality functioning, field independence vs. field dependence is now viewed as the perceptual component of the broader dimension of articulated vs. global cognitive style. For example, when the possible relation of field independence to intelligence was investigated, substantial correlations were obtained with some subtests of the Wechsler intelligence scales but not with others. The subtests of the Wechsler scales cluster into three major factors--a verbal dimension composed of the Vocabulary, Information, and Comprehension subtests; an attention-concentration dimension composed of the Digit Span, Arithmetic, and Coding subtests; and an analytic dimension, composed of the Block Design, Object Assembly, and Picture Completion subtests. The measures of field independence were found to correlate substantially with the dimension of analytic intelligence but not with the other two. Thus field-independent subjects exhibited a marked advantage on analytical intelligence tasks, but they could not be characterized as being superior in verbal intelligence or, in a meaningful way, as being superior in general intelligence (Goodenough & Karp, 1961, pp. 241-246; Witkin et al., 1962).

Children with a relatively articulated mode of cognitive functioning have also been found to have relatively articulated body concepts, as inferred from figure drawings; i.e., when asked to

draw human figures, these children display more realistic body proportions, more details, and more sex and role characteristics than children with a relatively global mode of functioning. Global subjects also tend to lack a developed sense of separate identity, as reflected in their relative reliance upon others for guidance and support, the relative instability of their self-view, their suggestibility and their susceptibility to social influence in forming and maintaining attitudes and judgments (Witkin et al., 1962; Linton & Graham, 1959).

Developmental studies have indicated that mode of cognitive functioning become progressively more articulated, and perception more field independent, with age up to late adolescence. At the same time, however, a child's relative level of articulation vis-à-vis his peers is quite stable. From age 10 to 14, the test-retest reliability of the perceptual index score of field independence was .64 for a group of 30 boys and .88 for a group of 30 girls, and from age 14 to 17 it was .87 for the boys and .94 for the girls (Witkin et al., 1962; Witkin, Goodenough, & Karp, 1967, pp. 291-300).

In an effort to uncover the possible origins of this cognitive style, Witkin and his colleagues studied patterns of maternal child-rearing practices and mother-child relations. On the basis of interivew data, the mothers were classified into two groups: those who fostered the child's differentiation from herself and who helped

him develop a sense of separate identify, and those who did not. In general, this classification of the mothers was found to be significantly related to the performance scores of the children, with the children of the mothers judged to have fostered differentiation being more field independent and cognitively articulated (Dyk & Witkin, 1965, pp. 21-35).

Differences have been noted in the type of defense mechanisms likely to be adopted by subjects at the two extremes of articulated and global cognitive sytle when confronted by conflict and stress. Articulated subjects are more likely to utilize specialized defenses, such as intellectualization and isolation, while global subjects are more likely to utilize primitive defenses, such as denial and repression. No general relation has been found, however, between the degree of articulation of the cognitive style and the degree of personal adjustment or psychopathology. Rather, as with the defenses, when psychological disturbances occur, there are differences in the kinds of pathology that are likely to develop at the two extremes of the style. Psychopathology in articulated persons is more likely to involve problems of overcontrol, overideation, and isolation; in severe pathological states, delusions are more likely to develop. Pathology in global persons, on the other hand, is more likely to involve problems of dependence, with symptoms such as alcoholism,

obesity, ulcers, and asthma; in severe states hallucinations are more likely to develop (Witkin, 1952, pp. 317-336). Such findings highlight the fact that styles of intellectual and perceptual functioning are part of the total personality and are intimately interwoven with affective, temperamental, and motivational structures. In some cases for example, "The general style of thinking may be considered a matrix...that determines the shape or form of symptom, defense mechanism, and adaptive trait" (Shapiro, 1965). In other cases the form-determining matrix may not be a mode of cognition but perhaps a type temperament or character structure or neurosis--the cognitive style would then be more derivative and would reflect but one component of a broader personality structure that permeates several areas of psychological functioning.

Although in most of this discussion one probably gets the impression that articulated, field-independent subjects have the advantage over their field-dependent peers, situations do exist where a more dependent reliance upon the external field, and particularly a reliance upon social stimuli for guidance and support, is profitable in the accrual of incidental information. Field-dependent subjects have been found to be significantly better than field-independent subjects, for example, in their memory for faces and social words, even though their incidental memory for nonsocial stimuli is not generally superior (Messick & Damarin, 1964, pp. 313-318; Fitzgibbon, Goldberger, & Eagle, 1965, pp. 743-749). The fact

that certain types of problem situations and certain types of subject matter favor field-dependent subjects over field-independent subjects and vice versa (just as other types of problems might favor broad categorizers over narrow categorizers or levelers over sharpeners, and vice versa) is extremely important, since it highlights the relativity of value of the opposing extremes of each cognitive style. Unlike conventional ability dimensions, one end of these stylistic dimensions is not uniformly more adaptive than the other.

The perceptual and intellectual consistencies just discussed have been interpreted in stylistic terms, which implies, for example, that an individual spontaneously and habitually applies his particular degree of analytic or articulated field approach to a wide variety of situations. Even though a relatively global individual may appear typically global in most situations, it is conceivable that when confronted with a situation that patently demands analysis he might be able to analyze with acceptable skill. Yet in the measurement of this cognitive style, it is usually presumed that subjects who characteristically display an analytic approach will in fact perform better on tasks requiring analysis (such as finding a simple figure in a complicated one) than will subjects who characteristically display a more global approach. Accordingly, most measures of analytic attitude are cast in an ability or maximum performance framework; if a subject does well at the task, he is assumed to have performed

more globally (or to be inadequately applying an unfamiliar, atypical analytic approach). In order to buttress the stylistic interpretation, it would be of interest to relate such maximum performance scores to measures of the spontaneous tendency to articulate the field in a task that ostensibly does not demand analysis.

In one attempt to develop such a task, subjects were required to learn to identify by name (a nonsense syllable) ten complex visual designs, each consisting of a large dominant figure, composed of elements, against a patterned background. In learning to identify these designs, the subject does not have to articulate the component parts, although the instructions do encourage analysis. The subjects are then told that each design was a member of a family of similar designs and that the names they had learned were family names. They are now presented with variations of the original designs (such as the element alone, the form alone, and the form composed of different elements) and asked to identify them in terms of the appropriate family name. In this strategy of test design, it was assumed that subjects who spontaneously articulated the designs during the learning process would be able to identify more variations than subjects who learned to identify the designs in a more global fashion. The total number of variations correctly identified, however, did not correlate significantly with the embedded-figures test. But this was because individuals differed consistently not only in the degree to which they

articulated the original designs but in the type of figural component articulated, and the articulation of only one of these components was associated with embedded-figures performance. A factor analysis of variation scores uncovered two major dimensions representing two distinct modes of stimulus analysis, one emphasizing the articulation of discrete elements and the other of figural forms. A third mode reflecting the utilization of background information was substantially correlated with the other two. A significant relation was obtained between embedded-figures performance and the element articulation factor but not the form articulation factor. Although on the one hand element and form articulation are distinct dimensions of stimulus analysis and exhibit different personality correlates, on the other hand they are significantly correlated with each other and combine, along with the background information factor, to form a second-order dimension (Messick & Fritzky, 1963, pp. 346-370).

These findings underscore the fact that the generality of the articulated vs. global cognitive style appears as a higher-order level in the factor-analytic sense. Another illustration of this point occurs in a study that attempted to extend Thurstone's perceptual closure factors into the verbal and semantic domains. Thurstone's factor of flexibility of perceptual closure, which is measured by tests like embedded figures, deals with the ability to break one closure in order to perceive a different one and thereby

depends upon the capacity to analyze a highly organized perceptual field. Thurstone's factor of speed of perceptual closure deals with the ability to assemble discrete parts into an integrated, meaningful whole and thereby reflects the capacity to structure a relatively unorganized perceptual field (Thurstone, 1944). The concept of an articulated mode of perception implies facility in both analysis and structuring (Dyk & Witkin, 1965) thereby requiring that the two closure factors be correlated, which usually tends to be the case. When several experimental closure tests were constructed using single words and meaningful discourse as the stimulus fields, factors were also uncovered for both speed and flexibility of verbal closure and for both speed and flexibility of semantic closure, in addition to the two perceptual closure factors. The concept of a general articulated vs. global cognitive style requires that all of these closure factors be mutually intercorrelated, which also tends to be the case, although the level of correlation is certainly not uniform. Indeed, some limitation on the generality of the style appeared in a second-order factor analysis, which revealed two relatively independent articulation dimensions, one involving the analysis and structuring of figural materials and the other the analysis and structuring of symbolic materials. In addition, a separate second-order factor of general analytical reasoning was also obtained (Messick & French, 1967).

Studies of other cognitive styles, particularly scanning and breadth of categorizing, have revealed a similar range of involvement in areas of personality and psychopathology. Silverman, for example, found that paranoid schizophrenics exhibited significantly more extensive scanning behavior and utilized significantly narrower categories than nonparanoid schizophrenics (Silverman, 1964). Gardner and Long (1962) reported that extreme scanning was marginally related to ratings of isolation, projection, and generalized delay on the Rorschach. This latter finding that scanning behavior tends to be associated with two different defense mechanisms suggests the possibility that extensive scanning may serve different purposes under different circumstances or, perhaps, that there may be two distinct types of scanning. The association with isolation, which is a preferred defense mechanism of obsessives, suggests that the scanning may occur in the service of information seeking, as reflected in the obsessive's concern with exactness to offset doubt and uncertainty. The association with projection, which is a preferred defense mechanism of paranoids, suggests that the scanning may occur in the service of signal detection, particularly danger-signal detection, as reflected in the paranoid's concern with accuracy to offset suspicion and distrust. Some current research at Educational Testing Service attempts to differentiate empirically between these two possible types of scanning. This is done by use of perceptual search tasks in which the subject is required to locate stimuli (signals) embedded

in meaningfully organized visual fields, e.g., to locate faces camouflaged in pictorial scenes or four-letter words embedded in sentences. Upon completion of the search task, the stimulus materials are removed, and the subject is then asked specific questions about the content of the scenes or the meaning of the set of sentences. Subjects who incidentally acquire information about the field in the process of scanning can thus be differentiated from those whose concern is apparently limited to detecting the signals.

With this brief characterization of cognitive styles in mind, let us now consider some of their possible implications for educational practice and evaluation. To begin with, cognitive styles, by embracing both perceptual and intellectual domains and by their frequent implication in personality and social functioning, promise to provide a more complete and effective characterization of the student than could be obtained from intellectual tests alone. These stylistic dimensions offer for our appraisal new types of process variables that extend the assessment of mental performance beyond the crystallized notion of achievement levels to a concern with patterns of cognitive functioning. These stylistic characteristics should have relevance, although direct research evidence is admittedly very scanty, not only for the course of individual learning in various subject matter areas, but also for the nature of teacher-pupil interactions and of social behavior in the classroom.

Thus, cognitive styles, by virtue of their widespread operation, appear to be particularly important dimensions to assess in the evaluation of instruction. Yet the very pervasiveness that underscores their importance at the same time interferes with the measurement of other important personal characteristics, such as dimensions of specific aptitude. This is because cognitive styles operate in testing situations as well and frequently interact with test formats and test conditions to influence the examinee's score. Consider, for example, the possibility that the five-alternative multiple-choice form of quantitative aptitude tests may favor subjects who prefer broad categories on category-width measures. Initial, rough approximations to the quantitative items might appropriately be judged by these subjects to be "close enough" to a given alternative, whereas "narrow range" subjects may require more time-consuming exact solutions before answering. Significant correlations between category preferences and quantitative aptitude tests have indeed been found, but the level of the correlation turns out to vary widely as a function of the spacing of alternatives on multiple-choice forms of the quantitative items. Scores for breadth of categorizing were found to be substantially correlated with quantitative aptitude scores derived from a multiple-choice form having widely-spaced alternatives, marginally correlated with scores on a free-response quantitative

test, and negligibly correlated with scores derived from a narrowly-spaced form. This suggests that wide spacing of alternative enhances, and narrow spacing disrupts, the "approximation" strategy that broad categorizers tend to employ on multiple-choice quantitative tests (Messick & Kogan, 1965, pp. 493-497). Such findings suggest that we should consider the "fairness" of our aptitude and achievement tests not only for different cultures and different sexes, but for individuals having different stylistic propensities. Thus, it is quite possible that cognitive styles are already being reflected in standard evaluation devices, however their operation under these circumstances is not being assessed for evaluation purposes but serves to contaminate the interpretation of other measures.

Information about cognitive styles offers several possibilities for instructional practice, but choices among them depend upon the results of much needed empirical research. For example, as soon as we are able to assess the cognitive styles of students, we have the possibility of placing them in classrooms in specified ways, perhaps in homogeneous groupings or in particular mixes or combinations. At this point it is by no means clear which particular placements will foster learning for individuals, just as it is by no means clear that homogeneous ability grouping is uniformly beneficial. Similarly, if we can assess the cognitive styles of students, we can also assess the cognitive styles of teachers and consider the possibility of

assigning teachers to students to obtain particular combinations of styles that would optimally foster learning. We could also consider selecting particular teaching methods that would be especially appropriate for certain cognitive styles and certain subject matters. As yet, of course, there is very little research to guide us on these points. But even in considering the possibility of matching the student to the teacher or the teaching method and remembering that with our present assignment procedures some students are in effect so matched while others are not, we should ponder what the criterion of success in this enterprise should be. Should it be the maximal learning of content skills and information?

Consider a possibility that, in the sciences at least, students with an articulated field approach, and perhaps reflective students as well, might learn better with an inductive or "discovery" method of teaching, since it would probably capitalize upon their propensities for analysis and careful consideration of alternatives. More global and more impulsive students, on the other hand, might learn content information better with a directed method of teaching in which rules and principles are specified rather than induced. Consider the likelihood, however, that in our efforts to optimize the learning of subject matter we may so solidify the global child's cognitive style that he may never learn to discover anything in his entire school career. This possibility suggests that teaching to

produce maximal learning of subject matter is not enough. We should also be concerned with the student's manner of thinking. One possibility here is that we should attempt to foster alternative modes of cognition and multiple stylistic approaches to problem solving.

Such a goal will not be easily attained, however, since there are many cognitive and personality dimensions that could interact with properties of teaching methods to produce negligible or adverse results. It makes a difference, for example, when and to whom and to what subject matter an inferential discovery method of teaching is applied. Kagan warns us, as an instance, that "impulsive children are apt to settle on the wrong conclusion in the inferential method and become vulnerable to developing feelings of inadequacy...Since these impulsively derived hypotheses are apt to be incorrect, the impulsive child encounters a series of humiliating failures and eventually withdraws involvement from school tasks (Kagan, 1967, pp. 153-163; Kagan, Pearson, & Welch, 1966, pp. 583-594).

The success of attempts to develop multiple modes of cognition in the individual will depend to a large extent upon the degree to which cognitive styles are malleable. Cognitive styles, as usually conceived, are habits that are spontaneously applied without conscious choice in a wide variety of situations. The possibility being considered here is that through manipulation of educational experience we might convert cognitive styles into cognitive

strategies, by which I mean to imply a conscious choice among alternative modes of perceiving, remembering, thinking, and problem solving as a function of the conditions of particular situations. If the cognitive styles are relatively mutable, such efforts at change and multiple development might be feasible at all levels of the educational sequence. If the cognitive styles, or at least some of them, are relatively immutable, it may be necessary to focus attention on the early years and attempt to foster multiple modes of cognition before particular styles crystallize and become predominant. This latter possibility of predominant cognitive styles may be inevitable, regardless of our educational efforts, but we might at least be able to increase somewhat the power of alternative cognitive modes in the hierarchy, thereby reducing to some extent the preemptiveness of habitual thought. As always, however, we must also consider and evaluate the potential dangers in such an enterprise: our efforts to foster multiple modes of cognition in a child may prevent him from soaring in the unfettered application of his preferred style in a particular field.

I have not discussed affective variables at length because most educators, at least when pressed, affirm the importance of enhancing curiosity and of implanting in the student massive and enduring positive affects toward learning and subject matter. Most of us would agree, therefore, that even when an instructional program does not

attempt to enhance positive attitudes directly, these variables should still be monitored if possible in the evaluation of the program to guard against unintended decreases in interest or involvement. In the measurement of these affective reactions, however, it seems to me unfortunate that evaluation studies rely so heavily upon the engineering model, which relates inputs and outputs, for there is a marked tendency to assess student achievement and attitudes only at the beginning and the end of the course. As Scriven has emphasized, the medical model is the appropriate paradigm for educational research (Scriven, 1966, pp. 33-49) and one derivative from that model should be an explicit attempt in evaluating a program to take account of the student's attitudes and feelings about the course of the treatment and not just the end result.

I wish to close by underscoring the importance of affect for learning and hence the importance of assessing affect in the evaluation of instruction. This point has been elegantly summarized by John Barth (1964, p. 17) in his novel, The Sot-Weed Factor:

> ...of the three usual motives for learning things--necessity, ambition, and curiosity--simple curiosity was the worthiest of development, it being the "purest" (in that the value of what it drives us to learn is terminal rather than instrumental), the most conducive to exhaustive and continuing rather than cursory or limited study, and the likeliest to render pleasant the labor of learning. ...this sport of teaching and learning should never become associated with certain hours or particular places, lest student and teacher alike...fall into the vulgar habit of turning off their

alertness, as it were, except at those times and in those places, and thus make by implication a pernicious distinction between learning and other sorts of natural human behavior.

REFERENCES

Barth, J. The sot-weed factor. New York: Grosset & Dunlap, 1964.

Bieri, J. Complexity-simplicity as a personality variable and preferential behavior. In D. W. Fiske & S. R. Maddi (Eds.), Functions of varied experience. Homewood, Ill.: Dorsey Press, 1961.

Bieri, J., Atkins, A. L., Scott, B., Leamon, R. L., Miller, H., & Tripodi, T. Clinical and social judgment: The discrimination of behavioral information. New York: Wiley & Sons, 1966.

Bruner, J. S. Toward a theory of instruction. Cambridge: Harvard University Press, 1966.

Bruner, J. S., & Tajfel, H. Cognitive risk and environmental change. Journal of Abnormal Psychology, 1961, 62, 231-241.

Dyer, H. S. The discovery and development of educational goals. Proceedings of the 1966 Invitational Conference on Testing Problems. Princeton, N. J.: Educational Testing Service, 1967, 12-24.

Dyk, R. B., & Witkin, H. A. Family experiences related to the development of differentiation in children. Child Development, 1965, 36, 21-55.

Fitzgibbons, D., Goldberger, L., & Eagle, M. Field dependence and memory for incidental material. Perceptual and Motor Skills, 1965, 21, 743-749.

Gardner, R. W., Holzman, P. S., Klein, G. S., Linton, H. B., & Spence, D. Cognitive control: A study of individual consistencies in cognitive behavior. Psychological Issues, 1959, 1, No. 4.

Gardner, R. W., Jackson, D. N., & Messick, S. Personality organization in cognitive controls and intellectual abilities. Psychological Issues, 1960, 2, No. 8.

Gardner, R. W., & Long, R. I. Control, defense, and centration effect: A study of scanning behavior. British Journal of Psychology, 1962, 53, 129-140.

Gardner, R. W., & Schoen, R. A. Differentiation and abstraction in concept formation. Psychological Monographs, 1962, 76, No. 41.

Goodenough, D. R., & Karp, S. A. Field development and intellectual functioning. Journal of Abnormal and Social Psychology, 1961, 63, 241-246.

Harvey, O. J., Hunt, D. E., & Schroder, H. M. Conceptual systems and personality organization. New York: Wiley & Sons, 1961.

Holzman, P. S. Scanning: A principle of reality contact. Perceptual and Motor Skills, 1966, 23, 835-844.

Holzman, P. S. The relation of assimilation tendencies in visual, auditory, and kinesthetic time-error to cognitive attitudes of leveling and sharpening. Journal of Personality, 1954, 22, 375-394.

Holzman, P. S., & Klein, G. S. Cognitive system principles of leveling and sharpening: Individual differences in assimilation effects in visual time-error. Journal of Psychology, 1954, 37, 105-122.

Kagan, J. Reflection-impulsivity and reading ability in primary grade children. Child Development, 1965, 36, 609-628.

Kagan, J. Personality and the learning process. In J. Kagan (Ed.), Creativity and Learning. Boston: Houghton Mifflin, 1967, Pp. 153-163.

Kagan, J., Moss, H. A., & Sigel, I. E. Conceptual style and the use of affect labels. Merrill-Palmer Quarterly, 1960, 6, 261-278.

Kagan, J., Moss, H. A., & Sigel, I. E. Psychological significance of styles of conceptualization. In J. C. Wright, & J. Kagan (Eds.), Basic cognitive processes in children. Monograph of Society for Research in Child Development, 1963, 28, No. 2, 73-112.

Kagan, J., Pearson, L., & Welch, L. Conceptual impulsivity and inductive reasoning. Child Development, 1966, 37, 583-594.

Kagan, J., Rosman, B.L., Day, D., Albert, J., & Phillips, W. Information processing and the child: Significance of analytic and reflective attitudes. Psychological Monographs, 1964, 78, (Whole No. 58).

Kelly, G.A. The psychology of personal constructs. Vol. 1. New York: Norton, 1955.

Klein, G. S. Need and regulation. In M. R. Jones (Ed.), Nebraska symposium on motivation. Lincoln: University of Nebraska Press, 1954, Pp. 225-274.

Klein, G. S., Gardner R. W., & Schlesinger, H. J. Tolerance for unrealistic experiences: A study of the generality of a cognitive control. British Journal of Psychology, 1962, 53, 41-55.

Linton, H. B., & Graham, E. Personality correlates of persuasibility. In. I. L. Janies (Ed.), Personality and persuasibility. New Haven: Yale University Press, 1959

McKenna, V. Stylistic factors in learning and retention. Princeton, N. J.: Educational Testing Service, Research Bulletin, 1967.

Messick, S. Cognitive interference and flexible control. Princeton, N. J.: Educational Testing SErvice, Research Bulletin (In preparation).

Messick, S., & Damarin, F. Cognitive styles and memory for faces. Journal of Abnormal and Social Psychology, 1964, 69, 313-318.

Messick, S., & French, J. W. Dimensions of closure in cognition and personality. Paper delivered at the American Psychological Association, Washington, D. C.: 1967.

Messick, S., & Fritzky, F. J. Dimensions of analytic attitude in cognition and personality. Journal of Personality, 1963, 31, 346-370.

Messick, S., & Kogan, N. Category width and quantitative aptitude. Perceptual and Motor Skills, 1965, 20, 493-497. (a)

Messick, S., & Kogan, N. Differentiation and compartmentalization in object-sorting measures of categorizing style. Perceptual and Motor Skills, 1963, 16, 47-51. (b)

Pettigrew, T. F. The measurement and correlates of category width as a cognitive variable. Journal of Personality, 1968, 26, 532-544.

Proposal for a research and development center for measurement in education. Princeton, N. J.: Educational Testing Service, 1965.

Shclesinger, H. J. Cognitive attitudes in relation to susceptibility to interference. Journal of Personality, 1954, 22, 354-374.

Scott, W. A. Conceptualizing and measuring structural properties of cognition. In O. J. Harvey (Ed.), Motivation and social interaction. New York: Ronald, 1963.

Scriven, M. The methodology of evaluation. In R. W. Tyler., R. M. Gagné , & M. Scriven (Eds.), Perspectives of curriculum evaluation. Chicago: Rand McNally, Pp. 39-83.

Scriven, M. Student values as educational objectives. Proceedings of the 1965 Invitational Conference on Testing Problems. Princeton, N. J.: Educational Testing Service, 1966, 33-49.

Shapiro, D. Neurotic styles. New York: Basic Books, 1965.

Silverman, J. Scanning-control mechanism and cognitive filtering in paranoid and nonparanoid schizophrenia. Journal of Consulting Psychology, 1964, 28, 385-393.

Thurstone, L. L. A factorial study of perception. Psychometric Monograph No. 4. Chicago: University of Chicago Press, 1944.

Wallach, M. A. Creativity and the expression of possibilities. In J. Kagan (Ed.), Creativity and learning. Boston: Houghton Mifflin, 1967. Pp. 36-57.

Witkin, H. A. Psychological differentiation and forms of pathology. Journal of Abnormal Psychology, 1965, 70, 317-336.

Witkin, H. A., Dyk, R. B., Eaterson, H. F., Goodenough, D. R., & Karp, S. A. Psychological differentiation. New York: Wiley, 1962.

Witkin, H. A., Goodenough, D. R., & Karp, S. A. Stability of cognitive style from childhood to young adulthood. Journal of Personality and Social Psychology, 1967, 7, 291-300.

Witkin, H. A., Lewis, H. B., Hertzman, M., Machover, K., Meissner, P. B., & Wapner, S. Personality through perception. New York: Harper, 1954.