

DOCUMENT RESUME

ED 105 583

EA 007 001

AUTHOR Helwig, Carl
TITLE Teacher Evaluation: The State of the Art and a Solution.
PUB DATE [/4]
NOTE 36p.; Best Copy Available

EDRS PRICE MF-\$0.76 HC-\$1.95 PLUS POSTAGE
DESCRIPTORS Behavioral Science Research; *Effective Teaching; Elementary Secondary Education; *Existentialism; Humanistic Education; Interaction Process Analysis; Measurement Instruments; Performance Based Education; Performance Contracts; *Research Methodology; *Self Actualization; Statistical Analysis; Student Teacher Relationship; Teacher Behavior; *Teacher Evaluation; Teaching Quality

ABSTRACT

The author begins with a critique of the failures of most neo-behavioral approaches to teacher evaluation and then assesses the existentialist reconceptualization of the primary relationship between teacher and pupil. He then reports on his studies using an individual's expressed and wanted behaviors in inclusion, control, and affection as predictor variables for authenticity. This probe asked, What might indicate that the teacher was authentic, free, or open? The research decision was not to change or to modify the behavior of the teacher, but to learn which variables might help determine the authentic, open relationship between the teacher and any other he might interact with, especially within the school organization. By following existentialist theory and avoiding any traits, personality, or characteristic approaches in teacher evaluation, especially through group norms or any other normative data, as well as by rejecting totally the neo-behavioral competency/performance-based approach, it is hoped that further systematic evaluation and painstaking isolation of variables contributing to the improvement of teacher evaluation shall be forthcoming through the approach articulated in this report and that other educational researchers will exploit this direction.

(Author/IRT)

BEST COPY AVAILABLE

Teacher Evaluation: The State of the Art and a Solution

by

Carl Helwig

*Old Dominion University, Norfolk, Va.
23508*

The recent resurgence of judging teacher effective-

ness has emerged, in part, from the now subsided student unrest and the continued demands by students for relevant learning. In particular, the works of N. A. Flanders, A. A. Bellack, D. G. Ryans, N. L. Gage, B. F. Skinner, W. J. Popham, and W. J. McKeachie, among others, in the fifties and sixties had added to a revival of behavioristic attempts to find universal criteria empirically in the identification of the "good teacher" or "good teaching." Ironically, this revival comes when data about the effectiveness of various teaching methods indicated rather strongly that "no particular method of college teaching at least is measurably to be preferred over another (Dubin and Taveggia, 1968).

Within this neo-behavioristic revival, change from the comparison of student outcomes on an achievement variable, usually scores on selected standardized tests, has shifted to some form of student opinion poll about teachers and their courses as the alternate criterion. Despite all that has been said under the guise as well as disguise of their brand of "science" and "empiricism," the neo-behaviorists seemed to have provided very little pedagogical illumination in defining the "good teacher"

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS PUBLICATION HAS BEEN SEPARATED FROM THE REGULAR PUBLICATION OF THE NATIONAL INSTITUTE OF EDUCATION. IT IS AVAILABLE FROM THE NATIONAL CENTER FOR EDUCATIONAL RESEARCH, 1200 KENTAVILLE AVENUE, ANN ARBOR, MICHIGAN 48106. THIS PUBLICATION IS AVAILABLE FROM THE NATIONAL INSTITUTE OF EDUCATION, 1200 KENTAVILLE AVENUE, ANN ARBOR, MICHIGAN 48106.

ED105583

EA 607 001

or "good teaching," both in higher as well as public education. Nothing, in addition, seemed to have emerged from Anderson and Hunka's bold 1963 multivariate approach to focus, first, on the evaluator and not the evaluatee, in this instance these authors themselves, moreover, failing to continue their research along the lines of their own proposal. Why, then, this continued and sad state-of-affairs in pedagogical research?

With the revival of the early 20th century behaviorism of F.B. Watson, E. L. Thorndike, and G. S. Hall since the middle fifties and with an added impetus given to the movement in pedagogical research by the neo-behaviorists mentioned above, the emphasis on statistical method in such research far overshadowed the substantivity of its findings in supposedly defining "scientifically" excellence in pedagogy. Sadly, this search has gone on, according to both Barr and Popham, with no substantive advancement toward understanding the "what" of teaching for at least fifty years!

Neo-Behavioristic Approaches

What explanations can be attributed to this conclusive neo-behavioral exercise in futility? Perhaps the neo-behaviorists needed a new "frame of reference." Offhand the neo-behaviorists seemed to be obsessed with statistical methodology, including its applications to computer capability. To deprive the hard, empirical statistically-oriented neo-behaviorists of their a priori "educational objectives" and with it their seemingly continual delight for statistical quantification for all forms of human

behavior would certainly sentence them to a scholastic wasteland. Several explanations for the neo-behavioristic failures seemed possible.

First, the neo-behaviorists often overlooked the simple distinction between measurement (or grading) and evaluation. Kossoff provided this simple analogy: if one had a basketful of eggs, he could grade them easily depending which criterion for grading he decided to choose--weight, size or color. However, none of these methods told one which egg he wanted for breakfast. Only after one ate the egg could he so determine. In these senses, an evaluation thus became an appraisal; a measurement, a quantification. Any determination of teacher effectiveness based on a priori criteria, including behaviorally-stated "educational objectives," therefore, yielded a quantification and not an appraisal (Kossoff, 1971-72).

Second, with their overconcern for quantification, the neo-behaviorists often overlooked an enduring attribute of teaching: teaching as a process. As such, teaching per se was neither too dependent on any given philosophy nor any given psychology, but was highly ordered as symbolism through language. "Both pragmatic thought and behavioristic psychology," remarked Bellack (1963) "emphasize that learning takes place through activity. This view is popularized in the expression that one learns to do by doing. The import of this view is to deprecate the use of language in teaching. . . . This view is held in spite of the fact that it is

difficult, if not impossible, to teach one anything without the use of language."

Admitted that behavioristic psychology has made major contributions--and will continue to do so--to such acts in pedagogy as the three R's, in special education, in physical education, and in driver education, among others, but can the neo-behaviorists provide me, as a learner, with a program of studies, or even a three-hour credit course, including their own behavioral "educational objectives" properly operationalized, which would teach me how to lead the "good life"--indeed a most pragmatic question for me? Or say, teach me something about my dread toward death? Or about my ability "to have reason"? The obvious answers to these rather life-like, complex questions have to be a resounding "no." However, when confronted with such an attack, the neo-behaviorists either ignored the question or merely remarked that since such "educational objectives" could be quantified, they were thus "pseudo" and therefore not "real educational objectives."

Under these conditions, should a neo-behavioristically-contrived criteria, including so-called "educational or instructional objectives" be employed in evaluating teacher performance? A categorical "no!" seemed to be the best likely answer. "What would one substitute in its place?" some would ask. "No substitution would be necessary," could be one form of reply. "But teachers must be evaluated, especially under the latest concept of 'teacher accountability'

or 'performance contracting'," could be the rejoinder. But, at least according to one Office of Educational Opportunity report, neither the concept of "teacher accountability" nor "performance contracting" as behavioral methods of evaluation seemed to be the answer to the problem of teacher evaluation:

During the 1970-71 school year the Office of Economic Opportunity with an authorization of \$7.2 million for the project decided to conduct a major study on performance contracting. Instead of relying on standard teachers and their standard teaching routines, a community signed a contract with outside specialists. This outside firm would be paid only if it improved the students' learning. Some thirty cities experimented with performance contracting heavily subsidized by the OEO.

Six educational firms, eighteen school districts, both city and rural, and 13,000 children in grades one to three and seven to nine, comprised the OEO year-long "experiment." The final results of this \$7.2 million "experiment" to determine the validity of another "experiment," namely, performance contracting: a failure. According to the OEO survey, both the special classrooms--the experimental groups--as well as in the ordinary classes--the control groups--failed to reach the standards expected in terms of national norms. In other words, two forms of behavioral methodology, the experimental or the "new" as well as the control or the "traditional" failed

to achieve the anticipated results (Time, February 14, 1972).

A third and final piece, this time of non-empirical but rational thinking, for the neo-behaviorists before a recommendation to them for a probable change in their most current "frame of reference"--the specific and the microscopic or "micro-teaching." The neo-behaviorists' own behaviorism did not give them license to choose among several alternative behavioral specifications whenever there is more than one alternative to specify--and even more so, to operationalize--an educational behavioral objective. Of equal, if not greater importance--even if such alternatives did exist--the neo-behaviorists could completely justify their selection of an educational objective stated in quantifiable (operationalized) behavioral terms as contrasted to the selection by others of an educational objective stated in non-behaviorial terms. This decision had to represent a choice other than the application of the neo-behaviorists' own behaviorism, since the alternative not to do so also existed through the simplest in logical reasoning.

Since all behaviorial theory rested on statistical quantification--hence the neo-behaviorists' own assertion of their being "scientific" or "empirical," which was not necessarily the case. Frequently the decision to quantify or to not quantify an educational objective seemed to rest on a somewhat silly notion, namely: through the definition an educational objective into quantifiable behaviorial

terms through the introduction of an activity verb. For example, the non-behavioristically stated "to have the pupil understand modern poetry" as an educational objective altered to the behaviorally-stated "to have the pupil write a critical essay on modern poetry."--thus leading to some kind of teacher grading or measurement--did not in and of itself clinch the ultimate wisdom of the neo-behaviorists' decision. In other words, neo-behaviorism by itself per se did not provide sufficient cause for the selection of behavioral over non-behavioral educational objectives (Steinberg, 1972).

A Critique

Most neo-behavioral tactics interfered with the learner's self-actualization. In this respect, Etzioni (1972) observed that "what is becoming increasingly apparent is that to solve social problems by changing people is more expensive and usually less productive than approaches that accept people as they are and seek not to mend them but the circumstances around them." Yet, even here, the problems of individual man was far more primary to individual qua individual man than the problems of his society to him, despite the utterances of Skinner in Beyond Freedom and Dignity (1972) among others, to the contrary. Individual man had, because of the existentialist absurdity of his existence, to cope with the perennial problems of life itself, especially about the inevitability of his own death,

the joy as well as the agony of his love, the reality of choice through his own free will, the experience of his own freedom in a democracy or a totalitarian state, and the fruitfulness or futility of his own interpersonal relationships with others. In other words, individual man, as a learner, had to come into encounter with his own human predicament (Kneller, 1964).

No neo-behaviorially-contrived "educational objectives" could cope with such rather high order, complex and highly individualized life-long concerns, which, moreover, could ^{not} be perceived as "objectives"--ultimate, educational or otherwise, but had to be conceived as continuous processes. Thus, as processes, these concerns did not represent final objectives, purposes, or missions--rather they were "the stuff of life" and hence more real to individual man than success at a vocation or being able to read and to write in a highly industrialized and technologically-oriented society. Therefore, the need for their quantification or their "operational definitions" became a near-impossibility, if not an unnecessary enterprise.

Despite all this, the neo-behaviorists would not abandon their elusive pursuit. As late as 1972, Gage continued to ask; Can "science" contribute the the "art of teaching"? This question, he held, is being answered more and more positively (despite the preceding fifty or more years of futility), and this educational "science" was now "reaching solid ground" by no longer stressing the

search for the universal and the macroscopic, as the earlier research efforts before the sixties, but the specific and the microscopic (Gage, 1972). Research in the specific and the microscopic in modifying teaching behavior should yield no more than another plethora of fragmented, empirical frivolity--as fifty years or so of previous research on the universal and the macroscopic.

Why? Because current teacher evaluation techniques generally suffered from a most serious statistical flaw-- and has been pointed out above, the neo-behaviorists could hardly be expected to abandon their current quantification tactics. This flaw rested upon the questionable assumption of linearity of the observed teacher behavior data. If the teacher-student relationship were viewed as an interaction paradigm, then all quantitative analytic techniques, including analysis of variance as well as other statistical techniques assuming linear, additive relationships among variables as a mathematic model expressed through the linear equation, had to be abandoned for the differential equation.

A feedback interaction paradigm of the teacher-student relationship assumed teacher behavior to be a mathematical function of student behavior and vice versa and both behaviors changed over time--especially if one acknowledged role change through the neo-behaviorists' own enthusiasm for their own "behavior modification" through "operant conditioning." Differential equations occasionally did lend themselves to linear representation; but despite this, it was not the usual linear representation which

provided the estimation techniques for ordinary multiple regression analysis nor analysis of variance. For to test whether an observed set of teacher-student interaction data were rather "well-fitted" by a differential equation required more complex estimation techniques in order to test for a teacher-student feedback loop interaction model (Editors, Harvard Educational Review, May, 1972).

The Mixed Results with Existing Measures

Three instruments have been extensively used with mixed results being obtained in defining desirable teacher and course characteristics.

Purdue Performance Instructor Indicator (1960) by H. H. Remmers and J. H. Sneddeker.

Purdue Rating Scale for Instruction (1950) by H. H. Remmers and D. N. Elliott.

Student Opinion of Teaching and Course (1969) by W. J. McKeachie.

The Fifth and Sixth Year Mental Measurements Yearbooks (1959 and 1965) contained critical evaluations of the Purdue forms, and as McKeachie admitted in a December, 1971 letter to this writer for his form, all three measures lacked any extensively field developed normative data. With the shift by other neo-behaviorists to "micro-teaching" from "macro-teaching" behavior was usually coupled with a shift from normative data interpretation to "competency or performance-based" criteria data interpretation. Thus, for all these reasons the employment of

these measures would result in the obtaining of questionable data for decisions about teacher performance.

It might be appropriate also to remark here that with the current emphasis by some neo-behaviorists on "micro-teaching," that is, their abandonment of their search for the macroscopic and the universal in teaching behavior, has resulted in a somewhat anomalous position on their part. Briefly, they assert teaching was a complex act; it required the separation of it into simpler component behavioral acts so these acts could be more readily quantified. But then, they hardly seemed to demonstrate, how after the appropriate components were discovered, these could be reassembled (other than through factor analysis) so their research could again comply with one of the simplest of mathematical laws, namely, "the whole is equal to the sum of its parts," or in other words, into the high order abstractions, the "good teacher" and "good teaching."

Anderson and Hunka (1963) suggested the employment of various multivariate statistical analyses to focus upon the perceptual dynamics of the evaluator himself and not the evaluatee. Their research proposal had yet to realize any fruition for the neo-behaviorists position, in general, held that students, because of their frequency of direct contact with their teachers, remained the most reliable source for observed data despite the fact that the reliability and validity any evaluatee criteria had yet to be categorically demonstrated by the neo-behaviorists. From

a most respectable search of the existing literature, Anderson and Hunka developed a near common sense hypothesis: the probable existence between anxiety and evaluation, namely, the more anxious the self, the more susceptible it was to anxiety and thus the more quickly the evaluative process took place. Furthermore, Anderson and Hunka also established a probable relationship between individual dogmatism, the degree of one's openness or closedness of his own belief system and anxiety. In short, the greater the self's dogmatism, the greater the anxiety, and hence the greater the misperception.

In summary, according to Anderson and Hunka, not all persons were qualified to evaluate others and their greatest objection to current procedures, besides this, had to do with the problem of criteria, that is to say, current procedures focused on teacher behavior per se and at the same time ignored situational determinants. Current evaluative procedures therefore ignored the evaluator himself and his own psychological make-up, thus ignoring his perceptual framework, the totality of his organized concepts, which, in turn, resulted into a self-concept. In short, as with any other evaluative procedure, the Self evaluated the self and not another.

The traits research approach to desirable teacher characteristics seemed to have ended with the work of Ryans (1960) with negligible results in tying desirable teacher characteristics to actual pupil growth. Much enthusiasm for Ryan's research stemmed from his use of factor analysis.

Several researchers, Hamacheck, 1969; Turner, 1969; and McKeachie, 1969 had in the late sixties revived the characteristic-traits thrust with their published efforts resulting more in exhortations for further research rather than actual, generalizable, systematic findings. Added to these researchers "findings" could be added the effort of Rosenshine, Costin, and Popham in the late sixties and early seventies, and as would be anticipated, the results of their findings were mixed and confusing.

In 1972, Gage criticized past research for its continued use of what he called "the criterion-of-effectiveness" paradigm, that is, a criterion variable was selected and then an attempt was made to find the predictor variables which would statistically predict the criterion. Usually the criterion, or dependent, variable turned out to be some form of pupil achievement variable, mostly in addition, of the cognitive variety, such as cognitive scores on standardized tests or the punier teacher-awarded letter or test score grade.

Gage's own proposed research thrust resulted into a contradiction of conceptualization and his rejection of the "criterion-of-effectiveness" paradigm. Gage himself would obtain initial student ratings on teacher behavior, assume this initial rating to be the predictor and then have the teacher undergo some "behavior modification." He would then employ a second set of student ratings as the criterion. This, of course, was not the interaction, feedback model Gage thought he had but a modified form of his rejected "criterion-of-

effectiveness" model (Editors, Harvard Educational Review, 1972).

With California's enactment of the Stull Act (1971), the need for a radically new approach to teacher evaluation, this time also to implement the legislative will, became evident with the Stanford University sponsored "Conference on the Stull Act," October 12-14, 1972, N. L. Gage, Chairman. Such neo-behaviorists as Glass, Popham, Abbott, Flanders, Alkin, and Baker participated. The Conference's participants were still searching for a viable conceptualization of what it meant to evaluate teacher performance. According to Broudy, the three recognized teaching styles of didactics, heuristics, and philetics tended only to blur further the quantification problem in teacher evaluation (Broudy, 1972).

So with more than fifty years of "empirical research," including the sustained, but disappointing efforts of the neo-behaviorists, would it be reasonable to expect modifications in teaching along the research routes proposed by the neo-behaviorists will dramatically affect teaching? Probably not-- simply because when coupled with the enormous inputs, both biological and cultural, including the educational, into the lives of our students, more than fifty years of research on teaching should by now have provided the American educational establishment, especially its neo-behavioral adherents, with two categorical answers, namely: (1) modifications of the current known teaching styles will not dramatically affect current learning, and (2) no further research on teaching

(Editors, Harvard Educational Review, May, 1972). Despite all this, the neo-behaviorists' shift from the universal and the macroscopic to the specific and microscopic seemed not justified by fifty to seventy-five years of previous futile research but rather symptomized a continued obsession for the quantification of teaching behavior, no matter how inconsequential the results produced by the neo-behaviorists' questionable selected variables as well as educational settings.

An Existentialist Reconceptualization

Along behavioral-pragmatic lines, what, then became the primary relationship between teacher and pupil? The current model of the American teacher seemed to be that of the boss and the lecturer as well as the intrusive conditioner and reinforcer--the last alluding to the primary concerns of the neo-behaviorists, namely through "operant conditioning" and "behavior modification." In contrast, under existentialist educational thought the teacher would be the receptive, Taoist helper in the teacher-student relationship--not necessarily always the pragmatic teacher, but the teacher capable of spontaneous as well as unselfish actions, constantly trying to provide harmony to the conflict the Self encounters in its human existential predicament. In the words of Maslow:

This doctrine of the Real Self to be uncovered and actualized is also a total rejection of the tabula rosa notions of the behaviorists and associationists, who often talk as if anything can be learned or as if anything can be taught, and as if the human being is a sort of passive clay to be shaped, controlled, reinforced in any way that the teacher arbitrarily decided (Maslow, 1968).

Existential educational theory would seem to be therefore the answer. Such a new "conceptualization" would totally eliminate behavioristically-oriented criteria as the primary means in teacher evaluation.

Why, At least four explanations can be given. First, at least since John Dewey's influence, education has depended heavily both philosophically as well as "realistically" on pragmatism. Pragmatism, in turn, has drawn most of its basic tenets from the behavioral sciences, particularly psychology (Kneller, 1958). Second, problems in education have lacked a rigorous experimental approach in research methodology and thus much educational research has been forced to use ex post facto conditions (Kerlinger, 1964).

Third, whenever the pragmatic concept of individual differences among learners has been applied on an experimental basis, the cataloguing of the findings has resulted in accepted social and psychological norms. In turn, this has produced educational programs which have focused on inquiry, consensus, and process rather than on the individual, authentic freedom of the learner himself. Thus, the experimentalistic-pragmatic approach to education had failed to sensitize the learner to his own sense of freedom (Kneller, 1958). Finally, because the behavioral sciences seem to have been overly concerned with externally-placed guides on behavior, education itself has given little attention to the organismic states of the individual learner, and, as a result, the behavioral has been emphasized in subject material as well as in teaching materials almost totally at the expense of the learner's

physiological needs (Hilgard, 1964). There, furthermore, seemed to be turning away from an orientation in education which lead the learner to feel, as a result of his total educational experience, that he was no more than an object about whom conclusions had already been reached. Whenever teachers focused on too much empiricism and behavioral statistical data derived from group norms, this easily could be the result.

As long as the three orientations of logical positivism, operationalism, and pragmatism continued to influence educational thought--the latter two most heavily--then education had no choice but to look toward the behavioral sciences (Di Renzo, 1966). Until existential thought, however, began to threaten behavioral and pragmatic educational thought during the sixties, the latter had almost near monopoly on American educational practices and research, including pedagogy, from about 1900 to 1960. Thus, despite the continued necessity for a pragmatic orientation, the behavioral sciences had yet to distinguish clearly between meaningful and meaningless pedagogical concepts as well as pedagogical concepts corresponding to real experience, on the one hand, and purely verbal definitions on the other (DiRenzo, 1966). Now, according to existentialist thought, the behavioral sciences did have an opportunity to unify subjective with objective knowledge and thus assist education in developing a more profound understanding of the relation of the individual to the group (Kneller, 1958), not through group norms and consensuality, but how to redefine

individual behavior as it related to another. Therefore, below is a proposal and some limited research concerned with teacher authenticity or freedom and a redefining through multivariate statistical techniques what, in fact, might provide a operational definition of the "good teacher." In short, it is proposed to go existentialist through behavioral tactics.

Authenticity as a High Order Concept and Criterion Variable

Beginning with Sartre's 1949 existentialist explanation of "over-determined" behavior as being a manifestation of Inauthenticity, sociologists, particularly, have begun to invoke the concept of Authenticity or its converse, as a crucial concept in organizational theory (Seeman, 1960, 1966; Etzioni, 1968). In the meantime, various researchers in educational administration have also tried to apply this concept to educational organizations through various empirical approaches hoping to gain some insight into how the concept of Authenticity might be explained in operationalized terms (Halpin and Croft, 1963; Halpin, 1966; Brumbaugh, 1968, 1969, 1971; Appleberry and Hoy, 1969; Hodgkinson, 1970). Observed Halpin particularly: "Authentic individuals, in the language of the French existentialists, were given the chance to invent themselves. . . . Authenticity is a tricky concept; it does not lend itself readily to operational definition."

The research herein was specifically addressed to determining whether certain existing empirically-determined

variables purporting to be indicative one way or another of authenticity could, in fact, be linked to an individual teacher's own fundamental interpersonal relations orientation. Five operationalized variables in the extant literature purported to measure some aspect of human behavior which, in turn, their respective researchers held to be one way or another a manifestation of authenticity: Esprit and Thrust, (Halpin, 1966); Custodialism (Willower et al., 1967; Appleberry and Hoy, 1969, 1970); Ambivalence (Seeman, 1960); Dogmatism (Rokeach, 1960).

Halpin held that two variables within his Organizational Climate Description Questionnaire (OCDQ), Esprit and Thrust, could furnish possibilities for an operational definition of Authenticity. "Probably the best single index to Authenticity which we could adduce from our present data is the combination of high OCDQ scores on both Thrust and Esprit," claimed Halpin. Esprit, as operationally defined by him, represented the teacher's own perception of his degree of morale or satisfaction within the school environment, while Thrust represented the teacher's perception of his principal's degree of positive leadership behavior.

Halpin through factor analysis of his own data on the construct, school organizational climate, discovered three major factors, Social Needs, Esprit, and Social Control, to describe the types of behavior that occurred among members of elementary school faculties. He then compared non-empirically these to Schutz' factors of Affection, Control, and Inclusion. Affection

corresponded to Social Needs and Control to Social Control; however, Halpin's factor of Esprit and Schutz's factor of Inclusion did differ for Esprit reflected the interaction between the group and the leader, whereas Inclusion seemed to describe primarily a measure of within the group.

Utilizing Halpin's OCDQ to identify schools with "open" or "closed" climates, the extremes of Halpin's six prototypic school organizational climate identifications of open, autonomous, familiar, controlled, paternalistic and closed, Appleberry and Hoy found their "relatively open" schools to be low on Custodialism or "humanistic" as contrasted to their "relatively closed" schools which were high on Custodialism. As an operational variable, Custodialism had been defined in the original work of Willower et al. with their Pupil Control Ideology Form (PCI Form) and represented a teacher's degree of Custodialism toward his pupils. About their study, Appleberry and Hoy said:

The rationale for the major hypotheses of this study stressed the authenticity of interactions among professional staff in schools with open climates and the inauthenticity of the interactions among professional staff in closed climates... We also theorized that humanistic pupil control ideology would facilitate and be facilitated by authentic interactions between teachers and pupils (Appleberry and Hoy, 1969).

In addition to Thrust, Esprit, and Custodialism as being measures of authenticity, Seeman's Ambivalence variable in his Leadership Ideology and Ambivalence Scales was suggestive of a form of inauthenticity for follow-up interviews with his sample of superintendents indicated that some were not persons

of clarity and decision as their stereotyping indicated (Seeman, 1960). The Ambivalence Scale, according to Seeman, thus gave clues to inauthentic behavior for reality-centered persons did not deny conflict nor exaggerate it, but rather possessed moderate degrees of ambivalence about their own ideological choices on the scale. In short, Seeman implicitly proposed that, based on his data, there were clues to individual psychological differences to authentic behavior.

Finally, Authenticity was suggestive in the work of Rokeach with the variable Dogmatism as operationalized by his Dogmatism Scale. In developing distinctions between the open and the closed mind, Rokeach studied the characteristics of individual belief-disbelief systems. He called the degree of openness to closedness of an individual's cognitive style as Dogmatism. "The more open one's belief system," he held, "the more evaluating and acting on information proceed independently on its own merits, in accord with the inner structural requirements of the situation," and thus, the individual's ability to be "reality-centered." In short, the lower the degree of Dogmatism, the more open the mind and the greater its tendency toward being reality-centered (Rokeach, 1960).

In sum, all these empirically-determined variables, Thrust, Esprit, Custodialism, Ambivalence, and Dogmatism, therefore, purported in one way or another to be suggestive of a higher order abstraction, Authenticity.

The Predictor Variables

Schutz's FIRO-B Scale was constructed to measure how an individual acted in interpersonal situations with others. Specifically, the scale measured an individual's expressed behavior toward others as well as his wanted behavior from others in three areas of interpersonal relationships:

Inclusion, Control, and Affection. In short, Schutz held that he could empirically and parsimoniously explain how an individual oriented and related toward others in all social situations. Specifically, the FIRO-B Scale measured an individual's expressed behavior toward others as well as his wanted behavior from others in three areas of interpersonal relationships: Inclusion, Control, and Affection (Schutz, 1968).

Inclusion was defined as behavior directed toward the satisfaction of the interpersonal need for inclusion and referred to behavior that reflected "belonging," "communication," and "togetherness" as contrasted to "isolated," "lonely," "ignored," and "excluded." Control was defined as behavior directed toward the satisfaction of the interpersonal need for control as suggested by "dominance," "authority," and "rules" and thus contrariwise to behavior that connoted "rebellion," "resistance," and "submission." Finally, Affection reflected behavior directed toward the interpersonal need for affection as characterized by such terms as "like," "personal," and "friendship" as contrasted to "dislike," "cool," and "emotionally distant" (Malpin, 1966).

The FIRO-B Scale, therefore, yielded six scores, expressed as well as wanted inclusion behavior, expressed as well as wanted control behavior, and expressed as well as wanted affection behavior.

As a result of all the above, the expressed as well as wanted behaviors in Inclusion, Control and Affection of an individual were considered herein as predictor variables. From Schutz, it was determined that these subscales could not be averaged and so each predictor variable was treated as a separate entity. The algebraic notations E^i , E^c , E^a , W^i , W^c and W^a , help in fixing mentally these six predictor variables (Schutz, 1968).

Methodology

There, therefore, was the necessity to assess the status of the art in faculty evaluative procedures by taking those instruments already in existence and stressing an individual's own perception of authenticity or freedom or openness, and subjecting them to data analysis through several multivariate procedures. Several researchers have addressed themselves to the concept, Authenticity, and claimed to have noted some aspects of it in their own research and therefore tried partially to link their own empirically-determined variables to this seemingly higher order concept. Of these researchers, Halpin particularly had given much speculation to its possible operational definition (Halpin, 1966).

As a result, what could it perhaps mean in empirical terms for a teacher to perceive accurately certain purported

forms of authenticity? A partial answer thus seemed reasonable in light of these data: the teacher high on his own expressed and wanted behaviors of Inclusion, Control, and Affection would also tend to be high on his own Esprit, low on his own Custodialism, neither be too low nor too high on his own Ambivalence, low on his own Dogmatism and perceive his principal to be high on Thrust.

In order to enter into multivariate prediction, it was hypothesized that a teacher's expressed behavior on Inclusion, Control, and Affection as well as his wanted behavior on Inclusion, Control and Affection, as measured by the FIRO-B Scale, would significantly predict ($p \leq .05$) his own Esprit, Thrust, Custodialism, Ambivalence and Dogmatism, as measured by the OCDQ Esprit and Thrust subtests, the PCI Form, the Ambivalence Scale, and the Dogmatism Scale, respectively. And if the level of acceptance were attained, then a linkage between purported authentic behavior as suggested by the latter five instruments and a teacher's own interpersonal orientation needs would, as suggested by the FIRO-B Scale, be established. And with all this perhaps a partial answer would be forthcoming as to what it meant to be authentic in empirical terms.

To be avoided in this research effort was the typical approach through null hypothesizing. In other words, the first problem became one of data analysis, and not the rigid posing of narrow statistical hypotheses:

There have to be people in the various sciences who concentrate much of their attention on methods of analyzing data and of interpreting the results of

statistical analysis. These have to be people who are more interested in the sciences than in mathematics, who are temperamentally able to seek for scope and usefulness rather than security, and who are willing to err moderately often in order that inadequate evidence shall more often suggest the right answer (Cooley and Lohnes, 1971).

The Results of Several Samplings

The first sample consisted of forty-seven public school teachers from Tidewater Virginia enrolled in two graduate courses at Old Dominion University in the 1970-71 academic year. Thus, pure randomness and fully true representativeness were not attained; however, with ex post facto data and multiple regression analysis, to which canonical correlation was directly related, true representativeness through random selection was not totally necessary (Kerlinger, 1973). And as a variable, individual teacher self-perceived authenticity seemed not easily amenable to experimental conditions.

The statistic utilized was canonical correlation. This statistic allowed for the conceiving of one set of variables as predictors and another set as criteria. The chi square test, moreover, became the appropriate test for significance (IBM Application Program, 1967).

Table I shows the results.

TABLE I
CANONICAL CORRELATION BETWEEN PREDICTOR AND
CRITERION VARIABLES

<u>Predictor Variables</u>	<u>Criterion Variables</u>
	Esprit, Thrust, Custodialism, Ambivalence, Dogmatism
Expressed Behavior	
Inclusion	
Control	$R_c = .603$
Affection	$N = 47$ $df = 30$
Wanted Behavior	
Inclusion	Chi square = 33.837 $p \leq .20$
Control	
Affection	

While the obtained level of significance, .20, did not reach the level of acceptance, nevertheless, it was suggestive. Therefore, since Esprit, Thrust, and Custodialism seemed from the literature to provide a "better fix" upon Authenticity, another canonical correlation was computed without the Ambivalence and Dogmatism variables. Table II shows these results.

TABLE II
CANONICAL CORRELATION BETWEEN PREDICTOR AND
CRITERION VARIABLES

<u>Predictor Variables</u>	<u>Criterion Variables</u>
	Esprit, Thrust, Custodialism
Expressed Behavior	
Inclusion	
Control	$R_c = .561$
Affection	$N = 47$ $df = 18$
Wanted Behavior	
Inclusion	Chi square = 29.909 $p \leq .05$
Control	
Affection	

With an R_c of .561, an .05 level acceptance was obtained. From these data, it could be inferred that a teacher's expressed behavior on Inclusion, Control, and Affection as well as his wanted behaviors on these three variables would predict his own Esprit, his own Custodialism, and his own perception of his principal's Thrust. Thus, if these three criterion variables pointed toward a teacher's authenticity, then his expressed and wanted behaviors did, in fact, help predict it.

A second sample consisted of 105 Tidewater Virginia teachers enrolled in graduate courses. Data here was gathered during the 1971-72 academic year. Neither the first, nor none of the succeeding derived canonical variates proved significant ($p \leq .05$). Thus with this sample, at least, the individual teacher's degrees of expressed and wanted behaviors on Inclusion, Control, and Affection did not predict his degrees of Ambivalence, Dogmatism, Custodialism, Esprit, and Thrust.

The significant findings worth reporting were bivariate correlations as follows:

TABLE III

	<u>Custodialism</u>	<u>Dogmatism</u>	<u>Esprit</u>	<u>Thrust</u>
Custodialism				
Dogmatism		.320 ¹		.205 ¹
Esprit			.243 ²	.337 ³
				.648 ³

¹ $p \leq .05$ (two-tailed test)
² $p \leq .02$ (two-tailed test)
³ $p \leq .01$ (two-tailed test)

These bivariate data confirmed, in part, Willower et. al.'s (1967) findings with their New York and Pennsylvania samples of significant relationships between teacher Dogmatism and Custodialism and Halpin's (1966) assertions of the strong relationship between a teacher's own Esprit and his own perception of his own principal's leader behavior on Thrust. Finally, the significant relationships between Dogmatism and Esprit as well as Dogmatism and Thrust suggested future research with these three variables and Custodialism for they still might give clues to what it might mean to be authentic in existing empirical terms.

A third sample was gathered during the spring 1973. Rather than using teachers enrolled in graduate courses, a random sample of eighty-four teachers in four Tidewater Virginia schools provided the data. Furthermore, upon examination of previous printout data and a thorough analysis of the Brumbaugh and Appleberry and Hoy reports, it was decided to eliminate the Dogmatism and Custodialism variables. Brumbaugh's wobbly findings with Rokeach's Dogmatism variable and Seeman's Ambivalence variable led to the conclusion that these two variables might be labeled as "contaminating" variables, for if two measures were presumed to measure the same construct, a correlation in the proper direction between them could be predicted. However, if the obtained correlation departed from the expectation, there was no way to know whether the fault would lie with either measure or in the formation of the construct itself. With the Custodialism variable, a rereading

of Appleberry and Hoy suggested an unjustified inferential leap by them in relating Custodialism as a variable to Authenticity.

The results of this third sampling follow.

Table IV
Canonical Correlation Between Predictor
and Criterion Variables

<u>Predictor Variables</u>	<u>Criterion Variables</u>
Expressed Behavior Inclusion Control Affection	Thrust and Esprit $R_c = .479$
Wanted Behavior Inclusion Control Affection	$N = 84$ $df = 12$ $\text{chi square} = 25.106$ $p \leq .02.$

Thus, a multivariate statistical link ($p \leq .02$) was established between the six predictor and two criterion variables. With Thrust as the single criterion variable, the six predictor variables also produced a multiple R of .474, significant at the .01 level, while with Esprit as the criterion variable the six predictor variables produced a multiple R of .345 which not significant at the .05 level. Therefore, the single criterion variable multiple regression analysis with the gathered ex post facto data suggested that the link between the six predictors and Thrust was, in fact, a more firm link than between the six predictors and Esprit.

The last item of data analysis was a multivariate partial correlation, that is, $R_{78.123456}$, where 7 and 8

symbolized the Thrust and Esprit variables and 1, 2, 3, 4, 5, 6, the six predictors were "partialled out" and resulted in an R of .418, which squared yielded a .173 thus explaining about 17 percent of the total variance produced upon the two criterion variables by the six predictor variables. In other words, additional criterion variables must be found to help further explain Authenticity, and perhaps the reinsertion of the Custodialism variable in future samplings would have merit.

Conclusions

While the empirical phase of this research report has produced little in the form of any infinite, universal law about teacher behavior and was thus only highly suggestive for another direction in empirical probing, unlike the neo-behaviorists who seemed to be concerned with modifying the teacher's behavior to a priori criteria, macro- and micro; this probe basically asked: what might indicate that the teacher was authentic, free or open? The research decision herein was not to change or to modify the behavior of the teacher, but to learn which variables might help determine the authentic or free, open relationship between the teacher and any other he might interact with, especially within the school organization. By following existentialist theory and avoiding any traits, personality, or characteristic approaches in ~~the~~ teacher evaluation, especially through group norms or any other normative data, as well as rejecting totally the neo-behavioral competency/preformance-based approach, it is

hoped that further systematic evaluation and painstaking isolation of variables contributing to the improvement of teacher evaluation shall be forthcoming through the approach articulated in this report and that other educational researchers will exploit this direction.

Bibliography

- Anderson, C. C. and S. M. Hunka. "Teacher Evaluation: Some Problems and a Proposal," Harvard Educational Review 33:1 (Winter, 1963), 74-95.
- Appleberry, J. B. and W. K. Hoy. "The Pupil Control Ideology of Professional Personnel in 'Open' and 'Closed' Elementary Schools," Educational Administration Quarterly 5:3 (Autumn, 1969), 74-83.
- Argyris, C. Interpersonal Competence and Organizational Effectiveness. Homewood, Illinois: Dorsey Press, 1962.
- Bellack, A. A., editor. Theory and Research in Teaching. New York: Teacher College Press, 1963.
- Broudy, H. S. "Didactics, Heuristics, and Philetics," Educational Theory. 22:3 (Summer, 1973), 251-261.
- Brumbaugh, R. B. "Authenticity: Neglected Concept in Educational Administration," in Changing Behaviors and Values: The Educational Administrator in American Society. Bowling Green, Ohio: Anderson Center for Personnel Development, August, 1969.
- "Conference on the Stull Act," Proceedings at the Stanford Center for Research and Development in Teaching, Stanford University, October 12-14, 1972.

- Cooley, W. W. and P. R. Lohnes. Multivariate Data Analysis,
New York: John Wiley and Sons, 1971.
- Costin, F., W. T. Greenough, and R. J. Menges. "Student Ratings
of College Teaching: Reliability, Validity, and Usefulness,"
Review of Educational Research 41:5 (December, 1971), 511-
535.
- DiRenzo, G. J. editor. Concepts, Thoery, and Explanation in
the Behavioral Sciences. New York: Random House, 1966.
- Dubin, R. and U. C. Taveggia. The Teaching-Learning Paradox.
Eugene, Oregon: University of Oregon, 1968.
- Editors. "Teacher Effectiveness and Teacher Education: The
Search for a Scientific Basis," Harvard Educational Review
42:2 (May, 1972), 282-287.
- Etzioni, A. "Basic Human Needs, Alienation, and Inauthenticity,"
American Sociological Review 33, 1968.
- Etzioni, A. "Human Beings Are Not Easy to Change After All,"
Saturday Review, June 3, 1972, 45-47.
- Flanders, N. A. Analyzing Teacher Behavior. Reading, Mass.:
Addison-Wesley Co., 1970.
- Gage, N. L. Teacher Effectiveness and Teacher Education: The
Search for a Scientific Basis. Palo Alto, California:
Pacific Books, 1972.
- Halpin, A. W. Theory and Research in Administration. New York:
Macmillan Co., 1966.
- Hilgard, E. R., editor. Theories of Learning and Instruction.
Chicago: 63rd Yearbook of the National Society for the
Study of Education, University of Chicago Press, 1964.

- IBM Application Program. (H20-0252-2), 1130 Scientific Sub-routine Package (1130-CM-02X). West Plains, New York: International Business Machines Corporation, 1967.
- Jibou, R. M. and C. A. Pollis. "Student Evaluation of Courses and Instructors," American Sociologist, November, 1971, 312-321..
- Kerlinger, F. N. Foundations of Behavioral Research, 1st edition. New York: Holt, Rinehart, and Winston, 1964.
- Kerlinger, F. N. Foundations of Behavioral Research, 2nd edition. New York: Holt, Rinehart, and Winston, 1973.
- Kerlinger, F. N. and E. J. Pedhuzar. "Educational Attitudes and Perceptions of Desirable Traits of Teaching," American Educational Research Journal 5:4 (November, 1968), 543-560.
- Kneller, G. F. Existentialism and Education. New York: John Wiley and Sons, 1958.
- Kneller, G. F. Introduction to the Philosophy of Education. New York: John Wiley and Sons, 1964.
- Kossoff, E. "Evaluating College Professors by 'Scientific' Methods," American Scholar, Winter, 1971-72, 79-93.
- Krasno, R. M. "Accountability and Research on Teacher Effectiveness," Administrators' Notebook 21:1 (November, 1972) Chicago: Midwest Administration Center, University of Chicago.
- Maslow, A. H. "Some Educational Implications of the Humanistic Psychologies," Harvard Educational Review 38:4 (Fall, 1968), 685-696.

- McKeachie, W. J., Yi-Guang Lin, and W. Mann. "Student Ratings of Teacher Effectiveness: Validity Studies," American Educational Research Journal 8:3 (May, 1971), 435-445.
- McKeachie, W. J. Teaching Tips: A Guidebook for the Beginning College Teacher. Lexington, Mass.: D. C. Heath, Co., 1965.
- Popham, W. J. Designing Teacher Evaluation Systems. Los Angeles: Instructional Objectives Exchange, 1971.
- "Result of a Test: F," Time, February, 14, 1972.
- Rinder, D. I. and D. T. Campbell, "Varieties of Inauthenticity," Phylon 13 (December, 1952), 270-275.
- Rokeach, M. The Open and the Closed Mind, New York: Basic Books, 1960.
- Ryans, D. G. Characteristics of Teachers. Washington: American Council on Education, 1960.
- Schutz, W. C. FIRO-B: A Three-Dimensional Theory of Interpersonal Behavior. New York: Rinehart and Co., 1968.
- Seeman, M. Social Status and Leadership: The Case of the School Executive. Columbus, Ohio: Bureau of Educational Research, Ohio State University, 1960.
- Steinberg, I. S. "Behavioral Definition of Educational Objectives" in Philosophical Redirection of Educational Research, Chicago: 71st Yearbook of the National Society for the Study of Education, University of Chicago Press, 1972.
- Tatsuoka, M. M. Multivariate Analysis. New York: John Wiley and Sons, 1971.

"Teacher Evaluation: Toward Improving Instruction," ERIC
Junior College Research Review, 4:5 (January, 1970).

Washington: American Association of Junior Colleges.

Trilling L. Sincerity and Authenticity. Cambridge, Massachu-
setts: Harvard University Press, 1972.

Turner, R. L. "Good Teaching and Its Contexts," Phi Delta
Kappan, November, 1970, 155-158.

Veldman, D. J. Fortran Program for the Behavioral Sciences.
New York: Holt, Rinehart, and Winston, 1967.