DOCUMENT RESUME

ED 049 176 SP 004 797

AUTHOR	Gage, N. L.; And Others
TITLE	Equilibrium Theory and Behavior Change: An Experiment in Feedback from Pupils to Teachers.
INSTITUTION	Illinois Univ., Urbana. Coll. of Education.
SPONS AGENCY	National Inst. of Mental Health (DHEW), Bethesda, Md.
REPORT NO	R-6
PUB DATE	Aug 60
NOTE	132p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$6.58
DESCRIPTORS	*Behavior Change, Elementary School Teachers, *Feedback, Grade 6, *Student Attitudes, *Teacher Behavior, *Teacher Evaluation

ABSTRACT

Giving sixth-grade teachers information as to how their pupils described their actual and their ideal teacher on 12 items of teacher behavior significantly changed in the direction of the pupils' initial ratings of their ideal teacher, and also made the teachers more accurate in predicting their pupils' descriptions of the teacher. Subjects of the study were 176 sixth-grade teachers in Illinois who were divided into experimental and control groups. They volunteered and cooperated in response to a series of mailings. (The report contains numerous data tables. In addition, appendixes contained samples of the questionnaires used.) (Author/RT)

BUREAU OF EDUCATIONAL RESEARCH
College of Education
University of Illinois
Urbana, Illinois

EQUILIBRIUM THEORY AND BEHAVIOR CHANGE:

AN EXPERIMENT IN FEEDBACK FROM PUPILS TO TEACHERS

N. L. GAGE, PHILIP J. RUNKEL, and B. B. CHATTERJEE

REPORT NO. 6

This investigation was supported by a research grant (M-650 R) from the Institute of Mental Health of the National Institutes of Health, Public Health Service.

STUDIES IN THE GENERALITY AND BEHAVIORAL CORRELATES OF SOCIAL PERCEPTION

AUGUST, 1960

Previous Reports of Studies in the Generality and Behavioral Correlates of Social Perception

(National Institute of Mental Health, Research Grant M-650)

1. Stone, G. C., Leavitt, G. S., & Gage, N. L. Generality of accuracy in perceiving standard persons. (Mimeo.)

2. Gage, N. L., Leavitt, G. S., & Stone, G. C. The intermediary key in the analysis of interpersonal perception. (Published in Psychological Bulletin, 1956, 53, 258-266.)

3. Stone, G. C., Gage, N. L., & Leavitt, G. S. An analysis of the negative correlation between two kinds of accuracy in predicting another's responses. (Published as "Two kinds of accuracy in predicting another's responses." Journal of Social Psychology, 1957, 45, 245-254.

4. Gage, N. L., & Cronbach, L. J. Conceptual and methodological problems in interpersonal perception. (Published in Psychological Review, 1955, 62, 411-422.)

5. Gage, N. L., Leavitt, G. S., & Stone, G. C. Social perception in the classroom. (Published as "Teachers' understanding of their pupils and pupils' ratings of their teachers," Psychological Monographs, 1955, 69, No. 21 (Whole No. 406).)

Abstract

Giving sixth-grade teachers information as to how their pupils described their actual and their ideal teacher on 12 items of teacher behavior changed the teachers' behaviors, as described a month or two later by their pupils, in the direction of the pupils' initial ratings of their ideal teacher, and also made the teachers more accurate in predicting their pupils' descriptions of the teacher.

Equilibrium Theory and Behavior Change: An Experiment in Feedback from Pupils to Teachers

N. L. Gage, Philip J. Runkel, and B. B. Chatterjee[1]
Bureau of Educational Research
College of Education
University of Illinois

This monograph presents the purposes, methods, and results of an experiment in feedback from pupils to teachers. The experiment was performed in intact classrooms in public schools. The research represents an attempt to test social psychological theory in an educationally significant setting.

We begin with an analogy: A blindfolded person throwing darts at a target will not get closer to the bull's eye. Take off the blindfold, and he improves. We say that the improvement is due to knowledge of results, or "feedback."

It is not too far-fetched to think of the teacher as "throwing" her behaviors, gestures and words, at pupils. How "close" she gets to the pupils--how well they like, understand, and learn from the teacher--may depend, in part at least, on the amount and kind of feedback she gets from her pupils.

In everyday teaching, how does the teacher get feedback? She glances at her class and notices signs of interest or boredom, comprehension or puzzlement, favorability or resentment. These signs appear in the pupils' facial expressions, movements, postures, and verbal behavior. The teacher asks direct questions of her pupils: perhaps about the lesson at hand, perhaps about the procedure and objectives of instruction. The teacher gives tests, either standardized or made by herself. The teacher talks informally with her pupils during or after class. She gets some feedback from other teachers who know things about her pupils and their reaction to her teaching. She hears things from parents, from people in the community, and from her principal.

---

[1]Now at the University of Michigan.

Obviously, for most teachers most of the time, these sources of feedback serve fairly well. If this were not so, our classrooms would be chaotic. Our teachers do meet with much success, and pupils do learn fairly well what we want them to learn. A good number of the teacher's darts hit the target. Even so, it is reasonable to suppose that there is room for improvement.

Feedback as an Experimental Variable

Much research has already been done to discover factors that make a difference in the effectiveness of teaching. Characteristics of teachers (traits, abilities, attitudes, etc.) at Time 1 have served as independent variables in a considerable part of this research, with the achievement of pupils at Time 2 as the dependent variable. Such studies have not yielded many positive findings; the relationships obtained have generally been low and inconsistent from one study to the next.

Why has this research been relatively barren? A full answer to this question would require more wisdom than can be offered here. All the same, two possibilities can be briefly indicated. First, it is possible that the characteristics of teachers that have been measured in the past (such as their intelligence, their attitudes, or their perceptual accuracies), however great the variance obtained with the indices used, still do not vary enough to make much difference in the kind of dependent variable with which we have been concerned.

This possibility may be likened to a point made about vision. Among normal people, visual acuity, after correction with spectacles, does not vary markedly. The important variables affecting visual perception in everyday life are not those residing within persons but rather those in the environment. Accordingly, we develop better lighting systems, typographies, traffic signs, and advertising layouts.

Suppose, for example, that accuracy of social perception is important in teaching. And suppose variance in social "acuity," viewed as a trait, is not

sufficiently great to make an appreciable difference in everyday life. Then it follows that we should turn to conditions of the external environment that affect such accuracy. We would not try to hold constant the stimuli and surrounding conditions in which social perception is measured so as to maximize the variance due to individual differences among persons. On the contrary, we would intentionally change and manipulate environmental conditions so as to change accuracy of interpersonal perception.

A second possibility is that studies concentrating solely on prior characteristics of the teacher fail to take into account a major class of determiners of classroom phenomena: those that can broadly be termed characteristics of pupils--their abilities, interests, needs, values, and perceptions. If teaching is viewed as an interactive process, we cannot account for classroom events solely in terms of the characteristics and behaviors of teachers. Rather, these events must be seen as outcomes in which pupil variables, both in themselves and in interaction with teacher variables, have an important effect.

Perhaps this point calls for another look at our dart-throwing analogy. Pupils are not inanimate, standardized, motionless, and passive targets. Nor are the behaviors of teachers vis à vis pupils the same from one pupil to the next. After each effort at hitting the mark, both the "dart" and the "target" may change. Even if we disregard the fact that the teacher is also changing, this formulation makes it easy to understand why measures of teacher characteristics at Time 1 must usually have low value for predicting pupils' learnings at Time 2.

Accordingly, characteristics of pupils must be taken into account and these cannot be considered invariant over classrooms or over occasions within classrooms. What are the research implications of this orientation? Some provision must be made for the ways in which characteristics and behaviors of [illegible] enter into the process. We should not assume that classes are so similar that they will differ only slightly in their response to the teacher's

behavior. An adequate design for the investigation of classroom teaching and learning phenomena must include some provision for the characteristics, behaviors, and perceptions of the pupils in relation to the teacher.

Type of Feedback Chosen for Experiment

Now suppose that in an experimental group of classrooms a treatment is applied aimed at increasing the amount of interaction between teachers and their pupils. The learning of the pupils may then improve, because the behavior of their teacher becomes more appropriate. But first we should determine whether the teacher's behavior changes as a result of this induced interaction. And that is what we tried to ascertain from the present experiment.

The kind of interaction we supplied in the present experiment was feedback from pupils to teachers. Such feedback was only one of several possibilities. It would have been possible to study feedback from teachers to pupils, or even from pupils to pupils. Why, then, did we make the choice indicated? The answer rests on an assumption, well supported by everyday observation, that teachers have most of the power to determine classroom activities in most schools. The teacher's power may not prevail in the "blackboard jungle" or in extremely "democratic" classrooms. But, in most elementary school classrooms, it is the teacher who makes most of the choices, from moment to moment as well as from month to month, as to what the pupils should study, how it should be explained, where pupils may sit, and so on. If this is so, change in classroom processes can be most effectively brought about by supplying feedback information concerning classroom processes to the teacher rather than to the pupils.

Furthermore, established role definitions already provide for considerable feedback from teachers to pupils. Teachers are expected to tell pupils how they perceive and evaluate them--their behavior, achievement, attitude, and like. Therefore, experimentally provided feedback from teachers to pupils

would not be much of a departure from the normal interaction that may already be presumed to take place in classrooms.

We have come to the position that experimentally introduced feedback from pupils to teachers should materially enhance classroom processes. But before going further, we needed to determine whether such feedback would change the behavior of teachers at all.

The Experiment in Brief

Our experiment was aimed at the question, Can teacher behavior be changed by informing the teachers how their pupils describe the behavior of their actual teacher and their ideal teacher? The pupils in our experiment indicated how well certain behaviors characterized their actual teachers. The pupils also indicated how well the same behaviors would characterize their ideal teacher. Some of the teachers (the experimental group) were given information concerning their pupils' opinions; the remaining teachers (the control group) were not given this information. A month or two later, all teachers were again described by their pupils as to how well the behaviors characterized the teachers. Briefly stated, our major hypothesis was that the experimental group of teachers would change its behaviors (as described by pupils) more than the control group. If the change was in a direction which could be considered desirable, the result could be called an "improvement" on the part of the experimental group.

A Note on Strategy

It is easy to think of reasons why this experiment might fail. The influence of the pupils' opinions was brought to bear on teachers through the mails, through printed words and graphs. It was presented in the same form to all teachers in the experimental group, regardless of differences in their personalities and situations. Much has been written on the difficulties of changing teacher behavior. Elaborate programs of diagnosis and therapy are often advocated as necessary to bring about such changes.

Two different strategies might be employed in a program of research and development designed to change teacher behavior. One strategy would begin with an elaborate approach--using tests, questionnaires, observations, diagnostic and therapeutic interviews (individual and group)--which would almost indubitably bring about desired changes. Since such a program would be too expensive for practical use, various features would then be stripped away from the elaborate program, one by one, so that eventually enough of the desired change remains while the program is reduced to a practical scope.

We began at the other end. Our strategy was to begin with a minimal program of practical scope--a kind of "mail-order" program. If this program did not work, we reasoned, additional kinds of influence could be brought to bear upon teachers until significant changes in behavior were brought about. We could, for example, add a personal interview with each teacher designed to interpret the pupils' ratings in terms that the teacher might understand better than a printed report. Additional elements could be added as necessary to secure the change desired. But if some change were achieved with the primitive and simple mail-order approach, we would have made a gain. This was our strategy in undertaking the present experiment.

The Practical Setting of the Experiment

No institution for teacher education can produce finished teachers. And improvement in teachers does not necessarily follow from experience alone. Hence, there has always been a need to help teachers become more competent while on the job. Supervisors, workshops, conferences, and study toward advanced degrees have in part met this need. As the teacher shortage continues, teachers will need even more advice, information, suggestions, and even fundamental reorientation to their task. Otherwise, teachers will fail to overcome the inevitable limitations of their pre-service preparation and their experience in particular classrooms.

The present research tried out a method of influencing teacher behavior which could become a feasible method of improving the teacher's practices on

the job. The particular behaviors on which we tried to focus were those amenable to change through increased awareness by the teacher of her pupils' perceptions of her actions. We certainly do not urge that teachers should always act so as to gain the approval of their pupils; but it does seem plausible that additional information about her pupils' reactions can help a teacher behave more appropriately to her pupils' needs.

The practice of collecting ratings of teachers from their pupils has had a moderate vogue for about 30 years. Advocates of this practice have urged many values of such ratings, among them being the improvement of teacher behavior. But no adequate test of this implied hypothesis has been made. Ward, Remmers, and Schmalzried (1941), for example, employed no control group. From their review, Morsh and Wilder (1954, p. 39) concluded that:

There appears to be considerable opinion that, properly used, student rating has value in bringing about instructor improvement. For example, Schutte (1926), Clem (1930), Flinn (1932), Riley et al. (1950), and Stuit and Ebel (1952), after having students rate instructors in one form or another, state (generally without adequate research evidence [italics added]) that student rating enables instructors to evaluate their courses and teaching performances and that students' opinions often provide a better basis for self-study and instructor self-improvement than do the opinions of supervisors.

Since that paragraph was written, at least two studies of the effect of student ratings on teachers have been undertaken. In 1957, Marjorie Savage investigated such effects among junior high school teachers of home economics. Her experiment differed in several respects from the present one. Her subjects were student teachers who, in the experimental group, tabulated their own pupils' ratings and then discussed them with the supervising teacher; the control group had regular conferences with the supervising teacher but access to the information from pupils' ratings. The subjects in our own experiment were regularly employed teachers who did not have any

with a supervisor in either the experimental or the control group. Furthermore, in Savage's experiment, the interval between the first ratings and the second ratings was only about 20 days; the first ratings were made only five days after the student teacher had begun to teach. As will emerge in our report, the interval between feedback and second ratings may be an important variable in relation to the discernible effect of the feedback. Finally, Savage did not exploit the advantages of analysis of covariance for controlling initial differences between groups in relevant variables. Her results were not statistically significant, and even their trend was not in the hypothesized direction (Savage, 1957).

A second undertaking relevant to our own is that of Bryan (1959). His research, still presumably in progress at this writing, has among its objectives an answer to the question, "To what extent can improvements in teacher effectiveness as judged by students be brought about through the use of written student reactions?" (Bryan, 1959, p. 5). In elaborating this question, Bryan states:

> Testimony to the effect that student reactions have been helpful to individuals and groups is plentiful. Not so numerous are reports of improvement based on a study of favorable changes in average ratings over a period of time. One of these was made by Wilson, who stated, "On those topics on which instructors had made a thoughtful and systematic effort to improve, the June averages were about 25 percentile points higher than in December..." Starrack found that ratings of teachers by students increased "quite materially" with each successive rating over a two-year period (p. 5).

Bryan's method will be to get student reaction to a number of teachers annually for a period of time. In the spring of one school year he will get student reactions from the classrooms of 75 or more teachers classified as the "experimental" group. He will mail to each of these teachers a summary and interpretation of the ratings of his students. Two-thirds (50) of these would be secondary-school teachers (grades 7-12) and one-third (25) would be

elementary-school teachers (grades 4-6). He will repeat the process in the spring of the next two years. In the third spring, each teacher will answer a questionnaire on the ways in which and the extent to which student reactions were helpful.

The control group of 75 similar teachers will be given no information concerning the written reactions of their students. There will be no further communication with them until time to obtain the reactions of their students in the third spring.

Ratings of teachers in the experimental group will be compared with those in the control group for the purpose of determining (a) whether the experimental teachers show more improvement in teaching performance as judged by students than the control teachers; (b) how many teachers show significant gains in each group; (c) on what questions the greatest gains are recorded; and (d) how these gains are related to years of teaching experience, the grade or subject taught, the school or faculty of which the teacher is a part.

As will appear, the present experiment, designed and executed in 1956, has much in common with Bryan's plans, made independently in 1959.

Equilibrium Theory

Although other theoretical frameworks might be equally useful, we have sought a rationale for our hypotheses in what has often been called "equilibrium theory," and what Zajonc (1960) has called "consistency theory." This theory represents a convergence of recent contributions by Heider (1958), Newcomb (1959), Osgood and Tannenbaum (1955) and Festinger (1957).

Heider

Heider's ideas, first published in 1946 and elaborated in his Psychology of Interpersonal Relations (1958), hinge upon the concepts of unit formation, sentiment, and balanced state. Units are entities perceived as belonging together; pupils in a classroom comprise a unit, and a teacher and her act comprise a unit. Sentiments are the ways people feel about or evaluate things.

A balanced state is one wherein perceived units and experienced sentiments co-exist without stress, without pressure toward change either in the unit formation or in the sentiment. Heider's general hypothesis is that the relation between sentiments and unit formations tends toward a balanced state. He sets up schematic situations in terms of a perceiving person, p, an observed other person, o, and a third entity, x, which may be either a third person or an impersonal object. The relationships of unit formation (e.g., belonging, owning, producing, causing) and sentiment (e.g., liking, respecting, admiring) are portrayed by means of U and L respectively, for the positive versions, and not-U and DL for the negative. In triads consisting of p, o, and x, Heider states the formal conditions of balance as follows: "A triad is balanced when all three of the relations are positive or when two of the relations are negative and one is positive. Imbalance occurs when two of the relations are positive and one is negative" (1958, p. 202). "If two negative relations are given, balance can be obtained either when the third relation is positive or when it is negative, though there appears to be a preference for the positive alternative" (1958, p. 206).

In our context, p stands for the teacher, o for the pupil, and x for the teacher's behavior. We assume that typically pLo, that is, that the teacher likes, respects, or is concerned with the good opinion of her pupils. The sentiment of o toward x is determined from the pupils' descriptions of their actual and their ideal teacher's behavior. The greater the difference between the pupils' descriptions of their actual and ideal teacher's behavior, the less the pupils like the teacher's behavior; the smaller the difference, the more the pupils like the teacher's behavior.

Now, if pLo, oLx, and pLx, there is no influence on the teacher to change her attitude toward her own behavior. But suppose the teacher is given evidence that oDLx, that is, that the pupils are critical of her behavior and would like it to change in certain ways. She infers oDLx from the gap between her pupils' descriptions of their actual teacher's behavior and their ideal teacher's

behavior. Then we can infer from Heider's formulations that there will be an influence on the teacher to change pLx to pDLx, always assuming that pLo will remain true. Thus the teacher begins to "dislike" her own behavior, and presumably will want to change it in the direction of the pupils' descriptions of their ideal teacher's behavior. The "imbalance" resulting from a situation in which pLo and pLx, but oDLx will, we hypothesize, tend to be resolved by a tendency toward pDLx. Subsequently, the teacher changes x to x' and restores the balance because, after she has changed her behavior, pLx'. She can assume, because she was told what the pupils want when she was given their description of the ideal teacher's behavior, that oLx'--and of course we continue to assume pLo.

In short, from Heider's theory of balanced states--including its extension by Cartwright and Harary (1956) to systems involving more than three relationships--we can derive the hypothesis that teachers given information concerning how their pupils describe their actual and their ideal teacher's behavior will change their behaviors toward those of the pupils' ideal teacher.

Newcomb

Newcomb concerns himself with "communicative acts" which, in their simplest form, consist of one person (A) transmitting information to another person (B) about something (X).

In the present study, we can consider A to be the teacher, B to be the pupil, and X to be the teacher's behavior. When we assume that the teacher has a positive orientation towards her pupils, and we tell the teachers what the pupils' orientations toward the teacher's behaviors are, we set up what Newcomb labeled "strain toward symmetry" on the part of the teacher to make the teacher develop the same orientation toward X.

What responses might a teacher make when he finds himself under such a strain?" The teacher can reduce strain by altering his own orientations or

his perceptions of his pupils' orientations. From Newcomb's analysis, we can derive the following alternatives:

(1) Influencing pupils toward his own orientation to the behaviors, i.e., making pupils have the same attitudes that he has, whether positive or negative, toward the behaviors.

(2) Changing his own orientation toward the behaviors, i.e., adopting the same attitude toward the behaviors as he perceives the pupils to have.

(3) Cognitively distorting the pupils' orientation, i.e., reinterpreting his perception of his pupils' orientation so that it becomes more like his own.

(4) Modifying his attraction toward the pupils, i.e., liking them less, feeling less "drawn" to them.

(5) Modifying his judgment of his own attractiveness to the pupils, i.e., feeling that the pupils like him less.

(6) Modifying his own evaluation of himself, i.e., liking himself less.

(7) Modifying his judgment of the pupils' evaluation of themselves, i.e., perceiving the pupils to like themselves less.

(8) Tolerating the asymmetry, without change.

How likely is each of these alternatives under the conditions of classroom life? It should be kept in mind that teachers and pupils are constrained to be associated; neither is free to discontinue the association, at least physically. Second, the teacher has subjected himself to a long period of training. At the time of our experiment, he is still in a teaching situation. These facts are evidence that co-orientation toward his pupils and his behaviors is at least in some respects still strong and positive. Third, the teacher's continuation in the role of teacher is evidence, because of the prescriptions of the role, that at least in some respects the teacher is strongly and positively attracted to the pupils. If these assumptions are warranted, Alternatives 4 and 5 above are unlikely.

Similarly, we assume it to be unlikely that, until other alternatives have been exhausted, the teacher will select Alternative 6 (modifying his evaluation of himself) or 7 (modifying his judgment of his pupils' evaluations of themselves). These assumptions stem from the proposition that the self-concept is relatively stable and is perceived as such in others, as compared with concepts of others.

Recall that the objects toward which he perceives asymmetry of relations between himself and his pupils are the teacher's own behaviors, subject to his own control, at least to some extent. Hence, the teacher can employ Alternative 1 above, i.e., attempt to achieve symmetry with his pupils by influencing them towards his own orientation. If he thinks that a certain behavior is very much like himself, while he is informed that his pupils do not consider it so, but they would like it to be so (i.e., they say it would characterize their ideal teacher), a strain toward symmetry will lead the teacher to communicative acts intended to make the pupils also consider the behavior very much like himself. In the classroom, these communicative acts will probably take the form of increased frequency or conspicuousness of the behaviors in question.

After an interval of time, these changes in behaviors will influence the pupils to consider the behaviors more like the teacher. In short, giving the teacher information as to the pupils' orientations toward the behaviors should influence the teacher's behavior.

Alternative 2 (changing his own orientation) will not be as likely to occur because, insofar as both the teacher and pupils consider the behaviors desirable, the teacher will not be likely to change his own orientation to the behaviors toward greater similarity with his pupils' orientations.

Alternative 3, cognitively distorting the pupils' orientation, can be made less probable, by giving the teacher presumably accurate information concerning the pupils' orientations to the behaviors. The teacher's use of native 3 will be revealed by the difference between the teacher's accuracy

in predicting the pupils' responses on a pretest and a posttest. Post-test accuracy should be greater than pretest.

Festinger

Festinger's theory of cognitive dissonance also bears on the kinds of phenomena with which we are concerned. Dissonance is a relationship between two elements in a person's cognition such that, considering these two alone, the obverse of one element would follow from the other. "To state it a bit more formally, x and y are dissonant if not-x follows from y" (Festinger, 1957, p. 13). Dissonance is considered psychologically uncomfortable; it motivates the person to reduce the dissonance and achieve consonance and also to avoid situations and inclinations which would increase the dissonance. To reduce dissonance, a person can (a) change the action or feeling which one of the cognitive elements represents; (b) change an environmental cognitive element by changing the situation to which that element corresponds, if he has sufficient control over the environment; (c) change a cognitive element without changing the corresponding reality, usually by finding others who will agree with and support his new opinion; (d) add new cognitive elements that will increase consonance or reduce dissonance, as through his choice of reading matter; or (e) add a new cognitive element that, in a sense, "reconciles" two elements that are dissonant (Festinger, 1957, pp. 18-24).

How does Festinger's theory bear on our experiment? It seems to us that dissonance can be induced by introducing into the teacher's cognitive field what Festinger would call new cognitive elements: (a) how pupils think their teacher does behave, and (b) how pupils think their ideal teacher should behave. To the extent that there is a discrepancy between the pupils' ratings of their actual and their ideal teacher, the teacher is furnished with a cognitive element that is dissonant with what we must assume to be another cognitive element, namely, the teacher's favorable opinion concerning his own behavior. Assuming the teacher has some respect for the pupil's opinion, we have a

situation in which "the obverse of one element would follow from the other." That is, the obverse of the teacher's favorable self-regard would follow from evidence that he does not conform, as his pupils see him, to his pupils' ideal. Or, what is the same thing, given the discrepancy between his own and ideal behavior, as his pupils see it, the obverse of favorable self-regard would follow from this discrepancy. The consequent dissonance should motivate teachers to change their behavior in the direction of their pupils' ideal.

Osgood-Tannenbaum

The "principle of congruity" stated by Osgood and Tannenbaum (1955) deals with three variables considered significant with respect to the direction of attitude change to be expected in any given situation: (a) existing attitude toward the source of a message, (b) existing attitude toward the concept evaluated by the source, and (c) the nature of the evaluating assertion which relates source and concept in the message.

The principle of congruity implies that "Whenever one object of judgment is associated with another by an assertion, its congruent position along the evaluative dimension is always equal in degree of polarization to the other object of judgment and in either the same (positive assertion) or opposite (negative assertion) evaluative direction." Applying these ideas to our phenomena, we can again consider one object of judgment to be an item of teacher behavior and the other object of judgment to be the teacher's pupils. When we give the teacher information concerning how pupils have rated her behavior and that of their ideal teacher, we are presenting the teacher with an assertion by pupils concerning their evaluation of the teacher's behavior. Let us assume that the teacher evaluates her pupils favorably and that pupils show some dissatisfaction (difference between their ratings of their actual and ideal teachers) with the teacher's behavior. We will then find an increase in "incongruity" on the part of the teacher which she will be constrained to reduce. The teacher's own attitude toward her behavior will tend to become

similarly unfavorable or affected with dissatisfaction. With this change in attitude toward her own behavior on the part of the teacher, the teacher will attempt to change her behavior. And we should then find a greater change in the teacher's behavior, as rated subsequently by her pupils, on the part of the experimental group of teachers, who were given information on pupils' ratings, as compared with the control group.

A Comparison

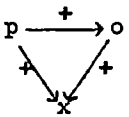
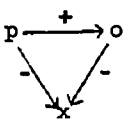
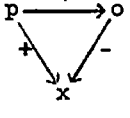
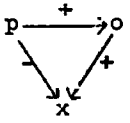
The four versions of equilibrium theory just summarized are brought together in Table 1. For each of eight possible situations, we give in the left-hand column an example of how a teacher might view her class ($p \rightarrow o$), some action or assignment ($p \rightarrow x$), and her class's relation to that action or assignment ($o \rightarrow x$). The situation is schematized, in the manner of Heider, in the second column. In the remaining four columns, we present brief characterizations of the situation in the terminologies of Heider, Newcomb, Osgood-Tannenbaum, and Festinger, respectively.

It should be noted that the four versions of equilibrium theory agree for the most part as to whether equilibrium or disequilibrium exists in each of the eight situations. Only Newcomb disturbs what would otherwise be unanimity, and then only in Examples 7 and 8. Here Newcomb assumes that if p and o are not constrained to continue their association, then situations 7 and 8 are not strained--equilibrium is achieved through dissociation and cessation of communication. If, however, as in the case of teachers who stay in their classroom jobs, p and o are constrained to continue their association, Newcomb argues that a dissociative strain toward balance will exist.

It was not one of the purposes of our experiment to test this difference between Newcomb and the other three. We had hoped to display some results bearing on this issue by using "assumed dissimilarity" as a measure of p_{Lo}. However, the analysis gave results so inconclusive that they will not be reported here.

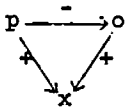
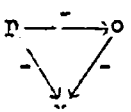
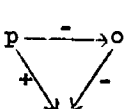
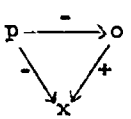
Tat . 1

A Comparison of Four Versions of Equilibrium Theory

Example	Case	Heider	Newcomb	Osgood-Tannenbaum	Festinger
1. A teacher perceives that her class is interested in its current assignment		Balanced: three positive relations	Non-strained: positive attraction to o and symmetry with o toward x	Congruent: o and x same-valued, with positive assertion	Consonant: liking pupils follows from their being interested in teacher's assignment
2. A teacher hears that her class is opposed to a school prank that would impair discipline		Balanced: two negatives and one positive	Non-strained: positive attraction to o and symmetry with o toward x	Congruent: o and x opposite-valued, with negative assertion	Consonant: liking pupils follows from their being against prank
3. A teacher perceives that her class is uninterested in its current assignment		Unbalanced: two positives and one negative	Strained: positive attraction to o and asymmetry with o toward x	Incongruent: o and x same-valued, with negative assertion	Dissonant: obverse of being responsible for assignment follows from pupils disliking it
4. A teacher hears that her class is planning a prank that will impair discipline		Unbalanced: two positives and one negative	Strained: positive attraction to o and asymmetry with o toward x	Incongruent: o and x opposite-valued, with positive assertion: "It is the pupils' plan."	Dissonant: obverse of liking pupils follows from disliking their plan

(continued on next page)

(T[illegible] continued)

Example	Case	Heider	Newcomb	Osgood-Tannenbaum	Festinger
5. A teacher who dislikes her class perceives that it is interested in its current assignment		Unbalanced: two positives and one negative	Strained: dissociation from o but symmetry with o toward x	Incongruent: o and x opposite-valued, with positive assertion	Dissonant: obverse of disliking pupils follows from their being interested in the teacher's assignment
6. A teacher who dislikes her class perceives that it is bored by an assignment that also bores the teacher		Weakly unbalanced: three negatives	Strained: dissociation from o but symmetry with o toward x	Incongruent: o and x same-valued, with negative assertion	Dissonant: obverse of disliking pupils follows from their disliking what teacher dislikes
7. A teacher who dislikes her class perceives that it is bored by the teacher's favorite class activity		Balanced: two negatives and one positive	Strained: dissociation from o and asymmetry with o toward x	Congruent: o and x opposite-valued, with negative assertion	Consonant: disliking pupils follows from their being bored by the teacher's favorite class activity
8. A teacher who dislikes her class perceives that it is planning a prank that will impair discipline		Balanced: two negatives and one positive	Strained: dissociation from o and asymmetry with o toward x	Congruent: o and x same-valued, with positive assertion	Consonant: disliking pupils follows from their planning a bad thing

Our experiment clearly reflects the equilibrium model shared by these four theorists. Although other schemes might have led us to the same plans and expectations, the models outlined above seemed parsimonious and suggestive to us. They called attention to assumptions (e.g., that pLo) and implications (e.g., "resistance to change" of teacher behavior might affect results) that might otherwise have remained unexplicated. Further research should move forward by taking into account some of the assumptions and implications spelled out by these authors. By casting our research into the consistency mold, we hoped to show connections between classroom phenomena and the larger realm of person perception and interpersonal behavior.

Method

In this section, we describe our procedures in selecting subjects and collecting data and in developing the items of behavior on which teachers were described. We also describe the instruments used in collecting data from pupils and teachers, the procedure used in communicating feedback, and finally the formal character of the experimental design.

Procedure in Selecting Subjects and Collecting Data

We wished to work with sixth-grade teachers. Their pupils would be mature enough to handle printed test materials with adequate comprehension, and this is the highest grade in which pupils typically have just one teacher. Having one teacher throughout the day for each class seemed an advantage to our experiment because pupils and teachers would then be subjected to more hours of influence from each other during the week. In higher grades, with "departmentalized" programs and several teachers for each pupil, the phenomena under study might be attenuated by the pupil's interaction with other teachers.

The subjects of the experiment were 176 sixth-grade teachers in Illinois and their approximately 3900 pupils. The distributions of class size among the control and experimental groups were about the same, as will be seen in Table 2.

Table 2

Distribution of Class-size in Control and Experimental Groups

Range of class size		Number of teachers having class size within this range	
		Control	Experimental
30 - 34		5	5
25 - 29		25	24
20 - 24		35	25
15 - 19		18	23
10 - 14		6	6
5 - 9		1	3
	N	90	86
	Median	22	22
	Q_3	26	26
	Q_1	18.5	18
	Q	3.75	4

The median class size was 22 in both control and experimental groups, and quartile values were also almost identical. There were 25 males and 65 females among the control teachers, 25 males and 61 females among the experimentals.

Where the control and experimental teachers were located in Illinois is shown in Figure 1. It is apparent that the teachers were scattered all over the state, in concentrations roughly similar to those of the population.

The Nature of the Sample

To some extent, the teachers were self-selected from a considerably larger group. They volunteered and cooperated in response to a series of mailings. How this came about is portrayed in the steps described below.

Our first step was to approach every superintendent of schools in the state of Illinois whose jurisdiction included a sixth grade. Each superintendent was invited to send us the name of one of his sixth-grade teachers whom we would in turn invite to participate in the research. After having been given names of teachers by superintendents, we tried to retain every possible teacher and her class for use in the final working sample upon which analysis of data would be performed.

Since, however, the entire data collection and the treatment were conducted by mail, the beginning list of teachers was inevitably subject to attrition. Some returns came in too late to be used if the schedule was to be maintained. Some that came in had to be discarded because a teacher had not followed directions. Some losses in the mails occurred because of changes of address, packages destroyed during the Christmas rush, and the like. And of course at each stage of the mailing losses occurred because of nonresponse.

Of 489 teachers originally receiving our invitation to participate in the research, 208 finally returned usable materials from themselves and their pupils at both pretest and posttest. Before analysis, this number was further reduced because some teachers worked under conditions incomparable with those of most teachers, in ways to be explained later.

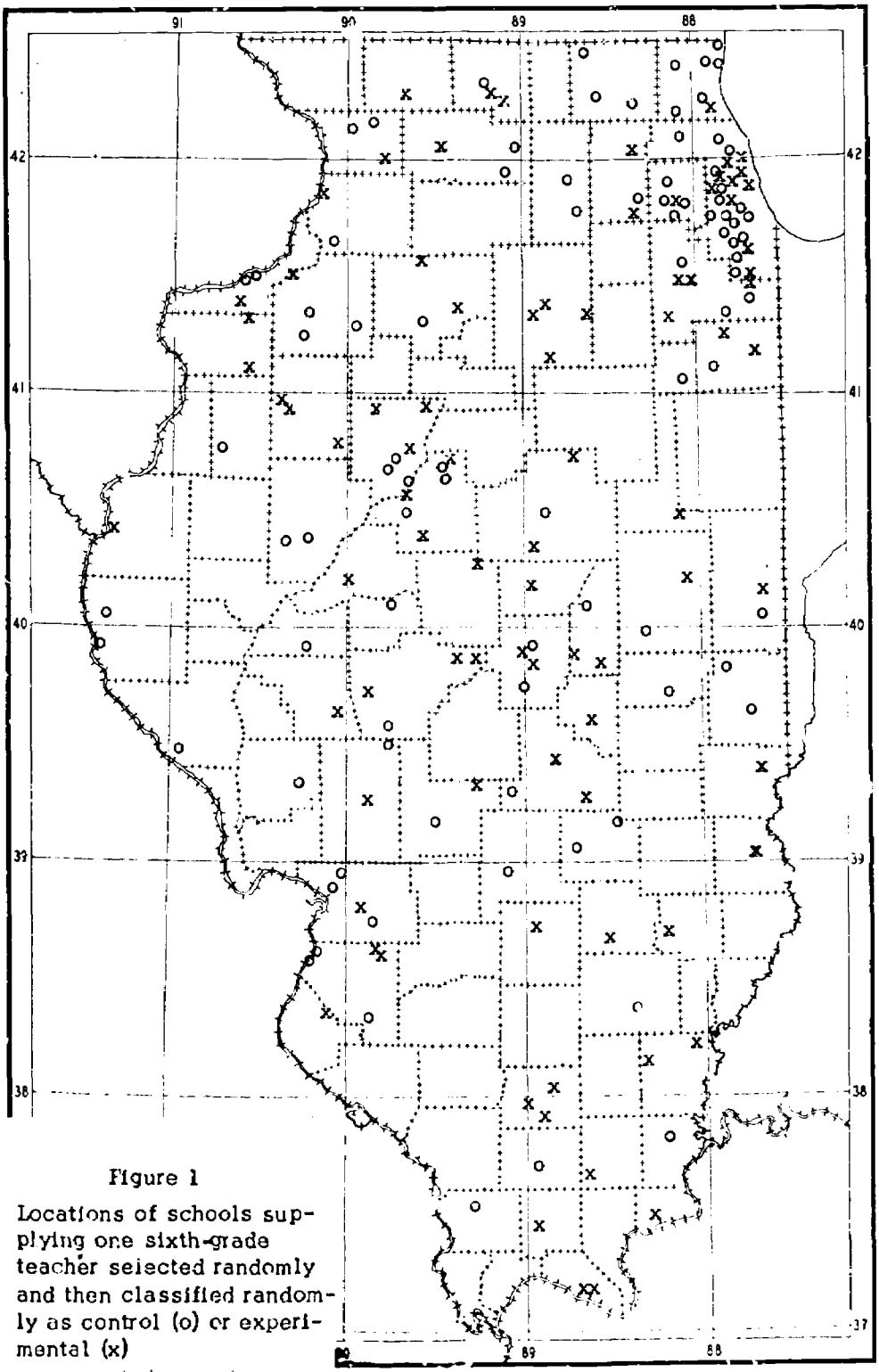


Figure 1

Locations of schools supplying one sixth-grade teacher selected randomly and then classified randomly as control (o) or experimental (x)

The final sample was probably biased in having a more than representative proportion of teachers who were interested in what their pupils think of their actions and who were willing to trust information gathered for them by a university research bureau. Possibly other factors are present, such as the ability to organize one's work so as to make time to administer the questionnaires. Some variables, as suggested by the foregoing evidence concerning class size and geographic location, were probably equally distributed between the two groups. Differences between groups at pretest on relevant variables were controlled by the use of analysis of covariance. Since our primary purpose was to test the effect of feedback to the teacher concerning her pupils' perceptions of her actions, our conclusions should be valid for teachers able and willing to administer the questionnaires and exchange the mailing pieces such as our study required.

Since the final list of subjects was determined primarily by attrition at various mailings, the latter are listed in Appendix A, along with the returns of questionnaires at each stage.

Developing the Items of Teacher Behavior

Our experiment called for four kinds of protocol, each obtained at the beginning (pretest) and end (posttest) of the fall semester, 1956-1957.

(a) SELF: description by the teacher of herSELF

(b) PERC: the teacher's PERCeption of how she would be described by "a pupil who belongs to the majority"

(c) ACT: descriptions by the pupils of their ACTual teacher

(d) IDL: descriptions by the pupils of their IDEAL teacher

All four of these protocols consisted of responses to the following set of 12 "stimuli," or brief verbal descriptions of teacher behavior:

A. Enjoys a funny remark made by a pupil.

B. Praises what a pupil says in class discussion.

C. Tells pupils about some interesting things to read.

D. Explains arithmetic so pupils can understand it.

E. Suggests to pupils new and helpful ways of studying.

F. Talks with a pupil after school about an idea the pupil has had.

G. Asks a small group of pupils to study something together.

H. Shows a pupil how to look up an answer when the pupil can't find it himself.

I. Asks the pupils what they'd like to study in tomorrow's lesson.

J. Acts disappointed when a pupil gets something wrong.

K. Explains something by using examples from games and sports.

L. Asks the class what they think of something a pupil has said.

Since these items determined much of what the experiment could reveal, their selection assumed considerable importance. First, the items were written to be meaningful to sixth-grade pupils; this required, in turn, that they be brief and have few qualifying phrases or clauses that would make pupils uncertain or hesitant. We wanted the items to elicit a quick judgment after a sweep of the pupil's eye across the statement.

Second, each item was intended to describe a reliably recognizable teacher behavior so that pupils would agree with one another as to whether the behavior occurred. This requirement meant that the items should describe teacher behaviors that were reasonably frequent, occurring at least a dozen times per semester, and quickly observed once they occurred.

Third, the items were intended to deal with attributes in which the teacher can change within the time-span of the research, since the major dependent variable of the experiment was to be change in the teacher's behavior.

Fourth, the items were designed to deal with behaviors determined by her-pupil interaction rather than by physical circumstances. Whether the

teacher "Shows movies often" might be determined primarily by her having a movie projector and films, rather than by motives and attitudes that could be influenced by our feedback.

Finally, to maximize the teachers' acceptance of the procedure, we sought to reduce the threat that teachers might experience from their pupils' descriptions. Accordingly, the items should describe desirable, or at worst neutral, kinds of teacher behavior. It would then be impossible for pupils to describe sins of commission on the part of the teacher. The least laudatory descriptions of teachers would then represent merely omissions of desirable acts. Even the least favorably described teacher would not, we hoped, be highly threatened by such a description.

From one point of view, "threat" should be great enough to exert pressure on the teacher to change. This criterion was in a sense opposed to that of maximizing acceptability of the experiment. Meeting both criteria required reducing the range of conditions intended to induce teacher change. Our hope was that the range would still be great enough to produce discernible effects.

The Attribute Interview Study

The 12 items about teacher behavior used in the pretest and the posttest were the end result of much developmental work. The process began with a search of the literature for items of this kind used by other investigators. Items describing specific behaviors easily recognizable on a questionnaire were comparatively rare. Another phase was that of interviewing professional colleagues; this also did not yield a large pool.

One helpful project in this process was the "Attribute Interview Study." Beginning in December 1955, this study was undertaken to explore the attributes used by teachers of grades 5 and 6 in describing or judging their pupils. The dimensions within which the teacher perceived her pupils might constitute a filter, so to speak, through which must pass her impressions of what her pupils thought about her. Dimensions important to teachers in judging their pupils

would provide clues to differences she could perceive among her pupils when they were reacting to her own actions. The pupil could then be asked questions about his teacher's reactions to certain of his behaviors that varied along these dimensions easily perceivable by the teacher. Details of the Attribute Interview Study are given in Appendix B.

The Discriminability Study

The "discriminability study" led to the final 12 items. When the discriminability study began, we had 22 promising items, shown in Appendix C, Table C-1; the 12 finally used are starred.

When used to characterize teachers, the items should not be readily confused with each other. Of any pair of items, a teacher should be able to say consistently "This is more like me than that." The items were to be maximally discriminable in this sense.

The discriminability study went through four stages, designed (a) to select items whose mean discriminability over pairs was a maximum, (b) to test items for spread of responses, and (c) to test which of two methods of labeling a response continuum would produce the greater spread of responses. The four stages are described in Appendix C.

The Questionnaire for Teachers (WDTE)

The questionnaire for the teacher was contained in the booklet, "What Do They Expect?" (An excerpt is shown in Appendix D.) Both the covering letter and the introductory material in the WDTE represented our project as offering a service to the teacher--that of providing her with information about how her pupils perceived her classroom behavior. And eventually we did indeed--in the booklet entitled "Report on Your Pupils' Opinions" (M13)--send every teacher a summary of the responses given by her pupils on our questionnaires. Depicting the project as a "Teachers' Information Service" was intended to heighten receptivity to the influence of the feedback.

The WDTE booklet was attractively prepared, since in soliciting the teacher's participation we depended entirely upon this booklet and a one-page mimeographed follow-up letter. Designed in cooperation with John Massey, then of the art department of the University of Illinois Press,[2] the booklet was printed in three colors, bound in a heavy gloss cover, and enlivened with whimsical drawings. The text was printed in short easy-to-read lines, justified on only one side and surrounded with much white space. The first seven pages of the booklet--in an informal, intimate, light vein--described the project and invited the teacher to participate. The last 17 pages contained the questionnaires and instructions for filling them out.

The care taken with this booklet seems justified. As shown in Appendix A, Table A-1, we got back 360 of the 489 sent out--a return of 74 percent. In view of the time, work, and dislocation of daily routine asked of the teachers, this rate of return seems considerably higher than that expected from rates typically reported.

The 12 items in the questionnaires were divided into two groups of six, hereafter denoted A-F and G-L. Within each group of six, 10 triads were formed in a balanced design. Altogether, the total of 20 triads comprised the "relative" format[3] of each questionnaire.

The WDTE contained four sections: (a) A 20-triad section for collecting relative data in which the teacher was asked to describe herself. (b) A 12-item section for collecting irrelative data in which the teacher was again asked to describe herself. (c) A 20-triad section for relative data in which the teacher was asked for her perception of how her modal pupil would describe her. (d) A 12-item section for irrelative data in which she was again asked how her modal pupil would describe he. The same 12 items appeared in every 20-triad relative section as in each 12-item irrelative section.

---

[2]The booklet was appealing enough to be admitted to an exposition sponsored by the Art Directors Club of Chicago.

[3]The terms "relative" and "irrelative" are used here as in Coombs (1953).

Excerpts from the instructions in the WDTE are given below.

Section (a) "On the next four pages a number of behaviors or actions are listed in groups of three. Read over the three behaviors in each group. Decide which of the three is most like you. You will find the letters 'M' and 'L' following that behavior. Please encircle the 'M.' Then decide which of the three is least like you. Encircle the 'L' following this behavior. For example,

Goes to movies often.	M	L
Likes to travel.	M	(L)
Reads a lot.	(M)	L

Then go on to the next group of three."

Section (b) "On the next two pages you will find again some things you have already met. But this time they come one at a time. After each thing are six different answers. Pick one of these answers and carefully make an 'X' in the box in front of the answer.

Goes to movies often.

- [ ] Very much LIKE me.
- [ ] Somewhat LIKE me.
- [ ] A little bit LIKE me.
- [ ] A little bit UNLIKE me.
- [ ] Somewhat UNLIKE me.
- [ ] Very much UNLIKE me."

Section (c) "In Part 1 you told us what actions were most and least like you and to what degree the actions were like you. Part 2 asks for your estimate of what answers your pupils would give if asked the same questions about your behaviors.

"Now, we know that you know that all your pupils would not answer alike. There are no doubt a few pupils whose answers it would be almost impossible to guess.

"Just the same, it's almost certain that a majority, or at least a large number, of pupils would answer these questions in the same way. Think of a pupil who belongs to this majority. Keep this pupil in mind and answer the questions in this section the way you estimate this pupil would answer them if he were asked to: 'Read over the three things in each group. Decide which of the three things is most like your teacher. Make a circle around the 'M' after it, and then make a circle around the 'L' after the thing which is least like your teacher.'"

Section (d) "Please answer the items on the next two pages also according to your best estimate as to how this pupil who is typical of the majority would answer them if he were asked to: 'Read the sentence which tells what your teacher might do. Then make an 'X' in front of one of the six answers.'"

In summary, the WDTE booklet provided relative and irrelative data for the teacher's perception of herself and the teacher's perception of how her pupils perceived her.

The Pupil Opinion Booklet (POB)

The Pupil Opinion Booklet (M9), printed in the same four sections as the WDTE, contained the following instructions:

> (For the first relative section): "You are to read over the three things in each group. Then decide which one of the three things is most like your teacher and also which one is least like your teacher."
>
> (In the first irrelative part): "First read the thing which tells what your teacher might do Then mark 'X' after one of the six answers."
>
> (In the second relative part): "This time, think of the best teacher you can imagine. (Do not think any more about the teacher you really have.) In the rest of the booklet, think of the best teacher you can imagine and think how that teacher would act."
>
> (In the second irrelative part): "Read the thing which tells what a teacher might do. Then mark 'X' after the answer which tells how much this thing would be like the best teacher you can imagine."

These instructions were supplemented by mimeographed material (M8) for the teacher to read aloud as her pupils prepared to answer the questionnaires. The POBs were provided with separate answer sheets (M9) to save postage and to allow using the booklets again for the posttest. The answer sheets were made to fold in the middle and were gummed around the edges. The outside bore the address of our Research Bureau. The face of the booklet bore this legend:

> "Your answers will be sealed up tight when you are finished. Then they will be sent to the University of Illinois. No one in your town--not your teacher--nor your principal--nor anyone else--will ever know how you answered these questions."

Here are the instructions concerning the answer sheet, read to the pupils by the teacher:

> "This is the answer sheet. (Hold up demonstration copy.) You will be marking your answers on this sheet. Notice that there is a line of glue around the edge (point this out). The glue is there so that you can seal up your answers when you have finished. When you are all through--not now!--you will lick the glue, fold the answer sheet closed (demonstrate) and seal it. The answer sheet, sealed closed, will then be like a letter in a large envelope. You should think of the answer sheet as a letter--a letter to an office in Champaign, Illinois. The people there need to know your ideas about things. They are very interested in the answers you will give to the questions. But they do not wish to tell anyone who gave any particular answer. Therefore the people in Champaign will carefully keep your name secret. Your name, written inside the answer sheet, will help the people in Champaign to keep separate all the different ideas and opinions. But they will never tell anybody what the names of the pupils were. When you have finished answering all the questions in this booklet, you will

seal the answer sheet with your name and your answers inside. Then I will put them all in a larger envelope and mail them to the office in Champaign, Illinois."

The instructions to the teacher for administering the POBs included the following comments:

"(A) Stand far enough away from the first row of pupils so that you cannot look down at their papers. Stay that far away during the entire session.

(B) If a pupil asks a question, do not walk to his desk to help him. Stay at the front of the room and use your copy of the questionnaire for demonstration, if needed.

(C) As the pupils complete their questionnaires and bring the answer sheets to you, be sure that every pupil has sealed his answer sheet closed. If anyone has not done so, ask him in a clear voice to do so, so that everyone nearby can hear you.

(D) When all are finished, let the pupils see you put all the answer sheets in their mailing envelope and seal it closed. All these actions will help the pupils to feel secure."

In our memory, every answer sheet came to us sealed.

The POBs used for the posttest were identical with those for the pretest. The WDTEs for the posttest were identical with those of the pretest except that the cover and introductory material had been removed.

Communicating Feedback: The Report on Your Pupils' Opinions (RYPO)

The Report on Your Pupils' Opinions (M13), printed in blue and red, was made so that individual information could be entered for each teacher. In each booklet, 12 charts appeared, one for each of the 12 items. The chart for each item had two parts: (a) a histogram showing how many of her pupils chose "Very much like my teacher," "Somewhat like my teacher," etc. (b) a histogram showing how many of her pupils chose "Very much like a 'best' teacher," "Somewhat like a 'best' teacher," etc. Also, on each chart the position of the median answer was shown. How these charts looked is shown in Appendix E. In the 12 charts, wide bands of blue and red ink were used in order to make the histograms, and the medians were indicated by means of triangular red and blue markers.

In short, the booklet gave the teacher two histograms for each item, one showing the distribution of pupils' answers characterizing their actual teacher (ACT) and one showing the distribution characterizing their ideal teacher (IDL). Furthermore, the median answer in her class was indicated in each chart.

Aside from these 12 charts, the teacher was also given information on the reliability of her pupils' answers. This information was put in terms of the consistency with which the pupils were able to say that one item was more like her than another item. The booklet explained the concept of consistency in answering the items appearing in triads and then bore the following list.

Percent of class

"High consistency in opinion of actual teacher and high consistency in opinion of ideal teacher _____

High consistency in opinion of actual teacher but low consistency in opinion of ideal teacher _____

Low consistency in opinion of actual teacher but high consistency in opinion of ideal teacher _____

Low consistency in opinion of actual teacher and low consistency in opinion of ideal teacher _____

The largest group is indicated in red. From these figures, you can see the extent to which opinion on these matters has 'jelled' in your class."

Consistency was computed for each of the two groups of six items separately in the method described earlier. Since each pair of items within the six was replicated in different triads, it was possible to tell whether the pupil contradicted himself in different triads in saying which item was more like his teacher than another.

The Experimental Design

Our experiment embodied what Campbell (1957) has termed the pretest-posttest control group design:

$$O_1 \quad X \quad O_2$$

$$O_3 \qquad O_4$$

where X represents the experimental treatment (i.e., feedback from pupils to teachers); O refers to the process of observation or measurement (i.e., pupils' descriptions of their actual teacher); Xs and Os in a given row are applied to the same persons (teachers); the left-to-right dimension indicates temporal order; and parallel rows represent equivalent samples of persons.

Translating Campbell's mode of expression into one of our major analyses yields the following:

Approximately Mid-October	Approximately Early November	Approximately Mid-December
Experiment group: pre-ACT	Feedback (RYPO)	post-ACT
Control group: pre-ACT	No feedback (letter explaining delay)	post-ACT

Extraneous Variables Controlled

In using this design, we controlled several sources of difference between pre-ACT and post-ACT that might have operated other than the feedback whose effect was to be ascertained. The rival explanations thus eliminated were (again in Campbell's terminology):

History. Specific event series other than X. (E.g., suppose an article had appeared in the Illinois Teacher, at about the time of our feedback, advocating one of our items of teacher behavior.)

Maturation. Effects systematic with the passage of time. (E.g., the possibility that all pupils may become less favorable toward teachers as the fall semester wears on.)

Testing. Persons taking a test the second time making scores systematically different from those taking the test the first time. (Our pupil ratings could modify the phenomenon under study, e.g., by sensitizing teachers to these items of behavior--and hence were probably "reactive" measures; cf. Campbell. But any such effects of testing would similarly influence our control and experimental groups. Hence the difference between them would probably be less than that to be found between our experimental group and a

"posttest only" group. Our control group probably got some unintended influence similar to that of the experimental group simply from participating in the pretesting and seeing what the items were. Hence our experiment tests the effect not of pupils rating teachers but of feedback to teachers of the ratings.)

Instrumental decay. Shifts in measurement conditions, as when raters become more experienced. (Our pupils might guess that the purpose of the posttest was to ascertain change or stability in the ratings and might rate accordingly on the posttest.)

Regression. Shifts toward the mean occurring due to unreliability of the measurements or random instability in the things measured. (Since our two groups were bound to differ in their pre-ACT means, they would regress statistically toward the total group mean; analysis of covariance determined whether changes occurred from pre- to posttest beyond those due to regression.)

Selection. Biased recruitment of subjects in the experimental and control groups. (Although our subjects were self-selected, the biases due to this source probably influenced the control and experimental groups in the same way and left them equivalent.)

Mortality. Drop-out of a biased subset of the subjects. (Same comment applies as for "Selection.")

Extraneous Variables Uncontrolled

Our design failed to rule out several sources of experimental-control difference other than those already noted as controlled. Still following Campbell, we discern the following shortcomings of our design as a basis for drawing generalizations to classrooms beyond those involved in our experiment.

The Interaction Effect of Testing. Our design, as already noted, offers no basis for generalizing to unpretested teachers. Our conclusions, in strict ic, can apply only to teachers who not only receive feedback but who were) pretested, i.e., prerated by their pupils and by themselves. Since we

would have had no information to feed back to the teachers without the prerating by pupils, this limitation is a realistic and necessary one. But it should be realized that the effects of the total program of prerating plus feedback are another matter, to be investigated only by an experiment involving unpretested groups of teachers, in what is called the Solomon (1949) four-group design:

$$\begin{array}{ccc} O_1 & X & O_2 \\ O_3 & & O_4 \\ & X & O_5 \\ & & O_6. \end{array}$$

How the third group (—, X, O_5) could be provided feedback without prerating by pupils is of course a major problem in studies of the present kind of experimental variable. If spurious, fictional feedbacks, serving as placebos, can be justified on ethical grounds, such Xs without pretesting might be used. Otherwise, we should have to use a design like the following:

$$\begin{array}{ccc} O_1 & X & O_2 \\ O_3 & & O_4 \\ A & & O_5 \end{array}$$

where the A indicates that at a specific time prior to X the third group was made equivalent to the other two "by a random sampling assignment" (Campbell, 1957, p. 304).

Limitations due to measurement procedures. We relied, in this experiment, on ratings as our measurement devices. These ratings of teachers were made by their pupils and the teachers themselves.

The validity of such measurements depends, like all validity, on the purpose or definition of the variables measured. It is defensible to say that ratings like those we obtained are intrinsically significant, quite apart from

The fact remains, however, that the generalizability of the present experiment is limited simply because no measures other than ratings were used. To overcome this limitation, we should eventually use a variety of ways of describing teacher behavior, "all having in common the theoretically relevant attribute but varying widely in their irrelevant specificities" (Campbell, 1957, p. 310). Observations by exper' visitors to the classroom, films, recordings, and perhaps objective tests of pupil achievement--insofar as they can be used without producing "reactive" effects on the phenomena being studied--are possibilities. But at present all these must be relegated to subsequent experiments, and we must limit our conclusions to what we can learn from ratings.

Results with Irrelative Data

We first present results with the irrelative data--the data obtained with the 12-item rating scales where each item was used independently of the others. These results are organized under three major headings: (a) pupil protocols, (b) teacher protocols, and (c) relations between teacher and pupil protocols. In a subsequent section we present results with the relative data.

Pupil Protocols

The four protocols obtained from pupils are listed below, along with the symbols used in referring to them:

pre-ACT -- the pupil's description of his actual teacher on the pretest

post-ACT -- the pupil's description of his actual teacher on the posttest

pre-IDL -- the pupil's description of his ideal teacher on the pretest

post-IDL -- the pupil's description of his ideal teacher on the posttest

Adjusted post-ACT

The most important single concern of this study was, Would teachers--as described by their pupils--change more if they were given information about how their pupils described them and their ideal teachers than if not given such information? It will be recalled that the experimental group was given such information, while the control group was not. Did this information--"feedback"--have effects manifested in how teachers were described by their pupils on the posttest?

Changes in teachers over time might occur "naturally," without being a result of the treatment manipulated tn this experiment. Such changes could occur as a result of unplanned developments in teacher-pupil relationships in our control group during the school semester. By comparing changes in the experimental group with those in the control group, we sought an indication of whether the experimental treatment produced changes above and beyond these "natural" ones.

Specifically, did the experimental and control groups of teachers differ in the post-ACT descriptions of them by their pupils? A straightforward attack on this question would determine whether the post-ACT means on each item were significantly different. This approach would, however, neglect the possibility that the teachers in the two groups may have differed in their initial status--at the time of the pre-ACT ratings by the pupils. Such differences, even if not statistically significant, would affect the comparisons of post-ACT.

The method for taking account of such initial differences is analysis of covariance, with the pre-ACT ratings serving as control variables, the post-ACT ratings as the dependent variables, and the feedback serving as the independent variable. When such analyses of covariance were carried out with each of our 12 items, the results in Table 3 were obtained.

Table 3

Means of Pupils' Ratings of Actual Teachers

($N_{exp.} = 86$; $N_{cont.} = 90$)

Item	pre-ACT		post-ACT		Adjusted post-ACT		Difference between Adjusted post-ACT Means	t^a
	Exp.	Cont.	Exp.	Cont.	Exp.	Cont.		
A	2.73	2.55	2.65	2.49	2.57	2.57	.00	< 1
B	3.32	3.18	3.49	3.58	3.46	3.61	.15	1.65*
C	2.43	2.32	2.37	2.38	2.33	2.42	.09	1.27
D	1.16	1.16	1.15	1.18	1.15	1.18	.03	1.76*
E	2.16	2.23	2.20	2.32	2.22	2.30	.08	1.17
F	4.19	4.07	4.11	4.16	4.07	4.20	.13	1.23
G	3.53	3.49	3.36	3.45	3.35	3.46	.11	1.19
H	2.03	2.07	2.18	2.18	2.19	2.16	-.03	< 1
I	5.34	5.25	5.33	5.27	5.29	5.30	.01	< 1
J	4.33	4.11	4.34	4.25	4.27	4.33	.06	< 1
K	3.29	3.10	3.31	3.36	3.24	3.44	.20	1.91*
L	2.78	2.78	2.84	2.96	2.84	2.96	.12	1.84*

[a] t was computed as $\sqrt{F}$.

*Significant at the .05 level, on a one-tail basis with df = 173.

Note.--In this and all other tables referring to irrelative data, means refer to a scale in which a score of 1 was assigned to the "Very much LIKE" rating scale alternative, 2 to "Somewhat LIKE," and so on, to 6 for "Very much UNLIKE."

Table 3 presents the pre-ACT, post-ACT, and adjusted post-ACT means of the 12 items for the experimental and control groups. Each pre-ACT and post-ACT mean is the mean over teachers of the median of ratings of the teacher by her pupils on the item. Also shown are the differences between the adjusted post-ACT means for the two groups and the t-statistic for estimating the statistical significance of these differences.

For four of the items, the differences are statistically significant at the .05 level; these are the differences for Items B, D, K, and L. It should also be noted that the direction of the difference is the same for 10 of the 12 items. This direction is that in which the post-ACT mean for the experimental group has a smaller numerical value than that of the control group.

Adjusted post-ACT minus pre-IDL

How should we interpret the direction of this difference between adjusted post-ACT means? Is the difference in the direction of the influence exerted by the feedback? To answer this question, we refer to the means of the pupils' median preratings of their ideal teacher (pre-IDL). That is, the feedback given the teachers in the experimental group concerning how their pupils rated their ideal teacher would presumably exert some influence on the teachers to change in that direction. Our hypothesis was that the difference between adjusted post-ACT and pre-IDL would be smaller for the experimental group. In Table 4, we have shown the means of the median pre-IDL ratings in the experimental groups. When the differences between adjusted post-ACT mean ratings and pre-IDL mean ratings were compared for the experimental and control groups, it turned out that the differences, for 10 of the 12 items, were smaller for the experimental group. The only two items not showing a difference in the hypothesized direction were Item A, in which there was no difference in either the adjusted post-ACT means or pre-IDL means, and Item J. the other items, the experimental group showed the smaller difference

Table 4

Means of Adjusted post-ACT and pre-IDL Ratings

($N_{exp.} = 86$; $N_{cont.} = 90$)

Item	Adjusted post-ACT Mean		pre-IDL Mean		Adjusted post-ACT minus pre-IDL		Is Difference between Columns 6 and 7 in Hypothesized Direction?
	Exp.	Cont.	Exp.	Cont.	Exp.	Cont.	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	
A	2.57	2.57	2.20	2.20	.37	.37	No
B	3.46	3.61	3.21	3.02	.25	.59	Yes
C	2.33	2.42	2.09	1.96	.24	.46	Yes
D	1.15	1.18	1.13	1.12	.02	.06	Yes
E	2.22	2.30	1.60	1.51	.62	.79	Yes
F	4.07	4.20	3.31	3.14	.76	1.06	Yes
G	3.35	3.46	2.81	2.83	.54	.63	Yes
H	2.19	2.16	1.66	1.57	.53	.59	Yes
I	5.29	5.30	4.06	3.83	1.23	1.47	Yes
J	4.27	4.33	4.54	4.44	-.27	-.11	No
K	3.24	3.44	2.64	2.66	.60	.78	Yes
L	2.84	2.96	2.77	2.80	.07	.16	Yes

between the adjusted post-ACT mean and the pre-IDL mean. In short, although the differences between adjusted post-ACT means are significant at the .05 level for only four of the items, the direction of the difference is in the hypothesized direction for 10 of the items.

It is impossible to make tests of the statistical significance of the combined results here because each item does not constitute an independent experiment or replication; i.e., the same subjects (teachers and pupils) were involved in all items. The consequent possibility of correlation among the results from item to item makes inappropriate the use of the binomial or chi-square models for the testing significance of combined results over all 12 items. It may be possible, through subsequent computations, to apply Hotelling's generalized student test here (Jones & Fiske, 1953). The consistency in the direction of the results does suggest that the hypothesized effect of the feedback did occur.

Adjusted post-ACT by interval. The kind of effect that might result from giving teachers "feedback" would, it is readily appreciated, take some time--days or weeks. After a teacher received information about how pupils described her and their ideal teacher, she might take thought as to how she might change her behaviors in order to come closer to her pupils' ideals. How fast this process might operate--how quickly teachers might change their behaviors--was a question on which we had no data to begin with. It might be that in just a few days the teacher could "internalize" the feedback, do something about it, and make these changes evident enough to her pupils that their ratings of her would reflect these changes. On the other hand, the process might take weeks or months, if it occurred at all. Indeed, there might be a curvilinear relation between the time interval and the amount of change it produced in the teacher's behavior as reflected in pupils' ratings of the teacher; in this event, after a certain interval the teacher might "regress" to her pre-feedback ways of behaving.

It was thus desirable to investigate the relation between change due to feedback and the interval from feedback to postratings of the teacher. We recorded the number of calendar days intervening between the mailing of the feedback information to the teachers and the date on which the teacher collected her pupils' postratings of herself and the ideal teacher. This variable, denoted "INTERVAL," was then used in analyses of the data. The frequency distribution of the number of days of the interval for each of the teachers in the experimental group is shown in Table 5. In this distribution, three interval sub-groups seemed to be apparent. In Table 5, these experimental sub-groups are denoted E_1, E_2, and E_3, for the shortest, medium, and longest intervals, respectively. The range of intervals was from 29 to 59 days--from one to two months. The median interval for subgroup E_1 was about 34 calendar days; for E_2, about 42 days; and for E_3, about 53 days. If we were to estimate the psychological significance of these intervals, we should say that this range is quite small. Ideally, the range should have been much larger, extending to eight months or even a year or two. But exigencies of data collection, school calendars, and other administrative considerations militated against a more adequate range in the present experiment.

The obvious hypothesis is that the approach of the experimental group to the pre-IDL becomes closer as interval became greater. To test this possibility, we repeated the analyses of covariance with the experimental group divided into three subgroups: E_1, E_2, and E_3, for the short, medium, and long experimental groups, respectively. The pre-ACT and post-ACT means resulting from this analysis of covariance are shown in Table 6. Table 7 shows the adjusted post-ACT and pre-IDL means for each of the four groups (Control, E_1, E_2, E_3) for each of the 12 items. In Table 8 are shown the differences in the three experimental-interval subgroups between adjusted post-ACT and pre-IDL means. Also shown in Table 8 are rank orders of the differences for the three experimental-interval subgroups, with a rank of 1 assigned to the largest difference for each item.

Table 5

Frequency Distribution of Intervals (Days)
between Sending "Report on Your Pupils' Opinions" (RYPO)
and Receiving "Pupil Opinion Booklets" (POBs) for Post-Ratings
--Experimental Group

Interval (No. of Days)	Number of Teachers in Experimental Group	Experimental-Interval Sub-group Code
59	1	
56 - 58	5	
53 - 55	3	E_3 (N = 15)
50 - 52	0	
47 - 49	6	
44 - 46	12	
41 - 43	17	E_2 (N = 39)
38 - 40	10	
35 - 37	14	
32 - 34	14	E_1 (N = 32)
29 - 31	4	

Table 6

Means of pre-ACT and post-ACT Ratings
by Experimental-Interval Subgroups

	pre-ACT				post-ACT			
	Control Group[a]	Experimental-Interval Subgroups			Control Group	Experimental-Interval Subgroups		
Item	C	E_1[b]	E_2[c]	E_3[d]	C	E_1	E_2	E_3
A	2.55	2.60	2.79	2.82	2.49	2.58	2.69	2.70
B	3.18	3.36	3.24	3.45	3.58	3.58	3.42	3.49
C	2.32	2.50	2.39	2.41	2.38	2.47	2.34	2.25
D	1.12	1.16	1.14	1.21	1.18	1.13	1.17	1.11
E	2.22	2.17	2.21	1.99	2.32	2.23	2.23	2.05
F	4.07	4.17	4.17	4.28	4.16	4.13	3.97	4.45
G	3.49	3.76	3.38	3.43	3.45	3.40	3.21	3.67
H	2.07	2.02	2.12	1.83	2.18	2.17	2.27	1.92
I	5.25	5.13	5.49	5.42	5.27	5.12	5.47	5.41
J	4.11	4.30	4.38	4.28	4.25	4.36	4.35	4.28
K	3.10	3.19	3.52	2.92	3.36	3.24	3.48	2.99
L	2.78	2.73	2.70	3.13	2.96	2.83	2.79	3.00

[a] N = 90

[b] N = 32

[c] N = 39

[d] N = 15

Table 7

Means of Adjusted post-ACT and pre-IDL Ratings
of Experimental-Interval Subgroups

	Adjusted post-ACT					pre-IDL				
	Control Group	Experimental-Interval Subgroups				Control Group	Experimental-Interval Subgroups			
Item	C	E_1	E_2	E_3	F	C	E_1	E_2	E_3	
A	2.56	2.62	2.56	2.42	< 1	2.19	2.08	2.26	2.31	
B	3.61	3.53	3.43	3.39	< 1	3.02	3.23	3.16	3.31	
C	2.42	2.38	2.33	2.23	< 1	1.95	2.16	2.01	2.08	
D	1.15	1.13	1.18	1.07	3.30	1.12	1.15	1.11	1.11	
E	2.30	2.25	2.22	2.16	< 1	1.51	1.63	1.57	1.54	
F	4.12	4.10	3.93	4.32	1.55	3.14	3.22	3.27	3.61	
G	3.47	3.22	3.30	3.73	2.67	2.83	2.87	2.71	2.92	
H	2.17	2.20	2.21	2.09	< 1	1.56	1.64	1.61	1.75	
I	5.30	5.24	5.20	5.33	< 1	3.83	3.89	4.04	4.47	
J	4.33	4.30	4.23	4.24	< 1	4.44	4.78	4.37	4.48	
K	3.43	3.24	3.23	3.21	< 1	2.65	2.53	2.69	2.73	
L	2.97	2.87	2.84	2.77	1.2	2.80	2.79	2.72	2.82	

Table 8

Adjusted post-ACT minus pre-IDL
Experimental-Interval Groups

Item	Adjusted post-ACT minus pre-IDL in Experimental-Interval Subgroups			Correlations of Obtained with Hypothesized Rank Orders of Differences between Adjusted post-ACT and pre-IDL	Is Rho in Hypothesized Direction?
	E_1	E_2	E_3	Rho[a]	
A	.54	.30	.11	1.0	Yes
B	.30	.27	.08	1.0	Yes
C	.22	.32	.15	.5	Yes
D	-.02	.07	-.04	-.5	No
E	.62	.65	.62	.13	Yes
F	.88	.66	.71	.5	Yes
G	.35	.59	.81	-1.0	No
H	.56	.60	.34	.5	Yes
I	1.35	1.16	.86	1.0	Yes
J	-.48	-.14	-.24	.5	Yes
K	.71	.54	.48	1.0	Yes
L	.08	.12	-.05	-.5	No

[a]Rho is the rank-order correlation between the obtained values for E_1, E_2, and E_3, respectively, and the hypothesis that the values would rank $E_1 > E_2 > E_3$.

In the fifth column of Table 8 are shown Spearman coefficients of rank correlation (rho) between the obtained and hypothesized rank orders. The hypothesis, it will be recalled, is that the effect of feedback is monotonically related to the length of time during which it had a chance to operate. Of the 12 rhos, nine are positive, and eight are .5 or higher. It appears that the adjusted post-ACT means of the three experimental subgroups have approached their respective pre-IDL means to a degree that conforms well with the hypothesis. The 15 teachers in group E_3, who had the feedback for the longest interval before the posttest, approached their pupils' pre-IDLs most closely. As is evident in Table 8, E_3's difference between the adjusted post-ACT and the pre-IDL is the smallest (has rank 3) in seven of the 12 items; by chance, this group would have this rank on only four items. Further, group E_1, which had the shortest interval, had the highest difference between the adjusted post-ACT and pre-IDL means on six of the 12 items, as against the four that would occur by chance. Further, we find that the differences (adjusted post-ACT minus pre-IDL) of these groups fell into the exact hypothesized rank order for four of the items (Items A, B, I, and K) as against two by chance; into the correct order with one reversal for four of the items (Items C, F, H, and J), as against two by chance; and into orders less well in conformity with the hypothesis for only four of the items (Items D, E, G, and L), as against eight by chance. Although it is difficult to evaluate the statistical significance of these results, because the same subjects are involved in all items, there seems little question that they tend to conform to the hypothesis: teachers who had the feedback for a longer time interval moved closer to the pre-IDL of their pupils.

Adjusted post-ACT and other measures of IDL. In the foregoing analyses we have compared the pupils' adjusted mean postratings of their actual teachers with their mean preratings of their ideal teachers. The rationale for using latter variable is that it constituted part of the feedback to the teachers

in the experimental group and presumably, therefore, part of the influence exerted upon those teachers.

It is conceivable that other measures of pupils' ideals were better indices of what teachers might perceive as desirable goals for them in the eyes of their pupils. Among these are the "adjusted post-IDL" and the "average IDL." For the sake of completeness, we also made comparisons with "post-IDL" means.

The adjusted post-IDL mean rating is the mean post-IDL rating adjusted through analysis of covariance for differences between groups in mean pre-IDL rating. The adjusted post-IDL mean might conceivably influence teachers through their continuing social interaction with pupils after the feedback, during the interval between feedback and postrating. Perhaps teachers can pick up cues as to their pupils' "ideals" during this interval. If these ideals do change somewhat, then adjusted post-IDL might provide a better basis for evaluating the mean adjusted post-ACT ratings.

When such comparisons were made, however, on the basis of "adjusted post-ACT minus adjusted post-IDL," the results were not as consistently in favor of the experimental group. Instead of 10 of the 12 differences being smaller for the experimental group, and instead of eight of the 12 items showing trends toward smaller differences with increasing "Interval" in the experimental-interval subgroups, comparisons of adjusted post-ACT with adjusted post-IDL yielded only seven smaller differences for the experimental group, and only four trends as hypothesized in relation to increasing interval.

The average-IDL rating is the mean of the pre-IDL and post-IDL mean ratings. It might be considered a meaningful measure of pupils' ideals on the ground that it provided a more representative measure of what pupils wanted--in the "best teacher you can imagine"--during the interval between the pre- and posttests. When average-IDL was subtracted from adjusted post-ACT, the differences did behave as consistently in favor of the experimental group as when pre-IDL was used; 10 of the 12 differences between adjusted post-ACT and

average-IDL were smaller for the experimental group, only those for Items H (no difference) and J (reversed) failing to go in the hypothesized direction. When the experimental group was divided into subgroups according to experimental interval, the rank order of the adjusted post-ACT minus average-IDL differences did not conform quite as well to the hypothesis; five of the rho's (N = 3) were 1.00, and three were .50, but two were -.50 and two were -1.00.

To determine whether pre-IDL was more relevant and influential than post-IDL as influence on the experimental group, we compared the adjusted post-ACT mean ratings with the post-IDL mean ratings on each item. The latter was not communicated to the teachers; it was collected from pupils and analyzed for various control purposes only. Since it was not part of the feedback to the teachers, it should not be expected to serve as a goal toward which teachers in the experimental group would change. On the other hand, none of the differences between pre-IDL and post-IDL is statistically significant. Accordingly, when adjusted post-ACT minus post-IDL differences are computed for the experimental and control groups on each item, we should not expect the differences for the experimental group to be as consistently smaller than those for the control group, as was the case when we compared the two groups on the basis of adjusted post-ACT minus pre-IDL. But we should not, on the other hand, expect any substantial difference.

Only eight of the 12 items show differences between adjusted post-ACT and post-IDL that are smaller for the experimental group, where 10 items did so when the difference was taken between adjusted post-ACT and pre-IDL. Similarly, when the experimental group is divided into interval subgroups, the correlations with the hypothesized rank order of the adjusted post-ACT minus post-IDL differences are not as high or consistently positive as were those obtained by subtracting pre-IDL from adjusted post-ACT; the differences fall

into the exact hypothesized rank order for only three items (as against four with pre-IDL), into orders with one reversal for six items (as against four with pre-IDL), and into negative-rho orders for three items (the same as with pre-IDL). Although the difference in results with pre-IDL and post-IDL is slight, it is in the direction favoring pre-IDL.

The foregoing analyses provided support for the hypothesis that the experimental group of teachers as seen by pupils, compared to the control group, would be closer to the pupils' pre-IDLs and for the hypothesis that the closeness of teachers on the post-ACT to the pupils' pre-IDLs would be positively related to the length of time the teachers had to show influence by the feedback before the postratings were made by their pupils. There was also tenuous evidence that the change toward the pre-IDL was more consistent over items than that toward the adjusted post-IDL, the average IDL, or the (unadjusted) post-IDL.

pre-IDL vs. Departure from Predicted post-ACT

Our main hypothesis may be restated as follows: The post-ACTs of teachers given feedback should depart from the post-ACTs that would be predicted from pre-ACTs if the teachers were not given feedback, and the departures should be in the direction of their pupils' pre-IDLs. In this formulation, we apply two control variables in the analysis of covariance: both pre-ACT and pre-IDL. That is, we adjust the post-ACT means for differences between the experimental and control groups in both pre-IDL and pre-ACT. The difference in post-ACT--in the form of departure of obtained from predicted post-ACT--is then attributable presumably to the feedback.

To determine how each teacher's post-ACT departed from prediction on each of the 12 items, we developed 12 regression equations, using the "within groups" regression coefficients from the analyses of covariance between pre-ACT and post-ACT. With these regression equations we computed for each teacher a "predicted post-ACT" score on each item. The difference between this score and the teacher's "obtained post-ACT" score was the teacher's "departure" from

prediction. Would these departures for the teachers in the experimental group go toward the ideals of the pupils more than would those of the control group teachers?

A proper answer here required that differences in pre-IDL between the groups should be controlled. So analysis of covariance was again used, with the teacher's "departure" score as the dependent variable, the pre-IDL as the control variable, and feedback as the experimental variable.

Table 9 shows the results of the 12 analyses of covariance performed on this basis. The adjusted means of the departures from predicted post-ACT of the experimental group are algebraically lower than those of the control group for 10 of the 12 items, i.e., for all items except A and I. The differences between adjusted departure means are significant at the .05 level (one-tail) for three items (B, K, and L). The rating scale was scored so that a lower numerical value was always assigned the "very much like" end of the continuum (1 = very much LIKE, 2 = somewhat LIKE, ..., 6 = very much UNLIKE). In the case of all items except Item J, the numerical value of the mean IDL rating is lower than that of the ACT rating. Hence, an algebraically lower adjusted mean departure value signifies a departure in the direction of the pupils' IDL ratings for all items except Item J. Our hypothesis is thus supported by the direction of the results from nine of the 12 items; the three items yielding results in disagreement with the hypothesis consist of two (Items A and I) in which the very slight difference is in the direction opposite from that where the pre-IDL mean of the experimental group is numerically lower than the pre-ACT means and one (Item J) in which the pre-IDL mean of the experimental group is numerically higher than the pre-ACT mean. In the case of the latter item, the hypothesis was that the departure-from-predicted-post-ACT mean of the experimental group would be algebraically higher than that of the control group.

pre-IDL vs. Departure from Predicted post-ACT, by Interval. Would interval between feedback and posttest have a relationship to departure-from-predicted-post-ACT? If so, we should expect to find adjusted departure scores tending to

Table 9

Departure from Predicted post-ACT
in the Experimental and Control Groups

Item	pre-IDL		Departure from Predicted post-ACT				t[c]	Is Difference between Columns 6 and 7 in Hypothesized Direction?[d]
			Unadjusted		Adjusted			
	Cont.[a]	Exp.[b]	Cont.	Exp.	Cont.	Exp.		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A	2.20	2.20	-.0078	.0058	-.0078	.0056	< 1	No
B	3.02	3.21	.0067	-.0907	.0157	-.1008	1.89*	Yes
C	1.96	2.09	.0500	-.0419	.0566	-.0510	1.49	Yes
D	1.12	1.13	.0156	-.0174	.0156	-.0181	< 1	Yes
E	1.51	1.60	.0300	-.0430	.0382	-.0512	1.45	Yes
F	3.14	3.31	.0578	-.0628	.0626	-.0670	1.26	Yes
G	2.83	2.81	.0611	-.0733	.0600	-.0722	1.32	Yes
H	1.57	1.66	.0122	-.0035	.0155	-.0088	< 1	Yes
I	3.83	4.07	-.0089	-.0105	-.0098	-.0095	< 1	No
J	4.44	4.54	.0467	-.0300	.0521	-.0344	< 1	No
K	2.66	2.64	.0989	-.0988	.0979	-.0988	1.92*	Yes
L	2.80	2.77	.0789	-.0605	.0779	-.0586	2.02**	Yes

[a] N = 90

[b] N = 86

[c] t was computed as $\sqrt{F}$.

[d] Positive values are in the hypothesized direction, except for Item J; i.e., the adjusted departure of obtained from predicted post-ACT was hypothesized to be algebraically smaller for the experimental group for all items except J.

*Significant at the .05 level, for one-tail.

**Significant at the .025 level, for one-tail.

become larger, in the direction toward the pre-IDL, for the experimental subgroups with the longer intervals. The algebraic value of the adjusted departure scores, for all items except J, should be lowest (most negative) for E_3 and highest (least negative) for E_1.

Analysis of covariance yielded the adjusted departure means shown in Table 10 for the control and three experimental interval subgroups. For Items D and G, the F-ratio of the between-groups to the within-groups variance is significant at the .05 level. But it is perhaps more revealing to look at Table 11, in which are shown the rank correlations with the hypothesized order of the adjusted mean departures of obtained post-ACT from predicted post-ACT. Six of these correlations are perfect as against the two that would be obtained by chance. Four of the rhos are -.5 or -1.00; four would also be obtained by chance. The data in Table 11 suggest that not only does the feedback make teacher behavior, as described by pupils, change in the direction desired by pupils (as is indicated by the data in Table 9) but that the time interval during which the feedback operates also makes a difference in the hypothesized direction.

Correlations between pre-IDL and Departure of Obtained post-ACT from Predicted post-ACT. So far we have evidence that (a) feedback makes a difference and (b) longer intervals make for greater changes toward pupils' ideals. A third step in the progression of influences would be that stemming from the variance in pre-IDL. Is the amount of the teachers' departure in obtained post-ACT mean from their predicted post-ACT mean correlated with differences in the mean pre-IDL of the teachers' pupils?

Some correlation between "departure" and pre-IDL might occur even without feedback, due to tendencies of pupils of a given teacher to respond at about the same scale level to both the ACT and the IDL rating scale situations. That is, the pupils might tend to be satisfied with the way they perceive their teacher to be at any given moment. Hence, we must compare the correlations

Table 10

Departure of Obtained post-ACT from Predicted post-ACT
in the Experimental-Interval Subgroups and the Control Group[a]

	Pre-IDL				Departure of Obtained post-ACT from Predicted post-ACT								
					Unadjusted				Adjusted				
Item	C	E_1	E_2	E_3	C	E_1	E_2	E_3	C	E_1	E_2	E_3	F
A	2.19	2.08	2.26	2.31	-.008	.047	-.015	-.027	-.01	.07	-.03	-.05	< 1
B	3.02	3.23	3.16	3.31	.067	-.031	-.115	-.153	.08	-.04	-.12	-.17	1.27
C	1.95	2.16	2.01	2.08	.050	.006	-.046	-.133	.06	-.01	-.05	-.14	1.40
D	1.12	1.15	1.11	1.11	.016	-.031	.021	-.087	.02	-.03	.02	-.09	3.00*
E	1.51	1.63	1.57	1.54	.030	-.013	-.044	-.107	.04	-.03	-.05	-.11	< 1
F	3.14	3.22	3.27	3.61	.058	-.022	-.200	.207	.06	-.02	-.20	.19	2.56
G	2.83	2.89	2.68	2.92	.061	-.209	-.119	.313	.06	-.22	-.11	.30	2.72*
H	1.56	1.64	1.61	1.75	-.012	.013	.022	-.093	-.01	.01	.02	-.11	< 1
I	3.83	3.89	4.04	4.47	-.009	-.059	-.021	.013	-.01	-.06	-.02	.01	< 1
J	4.44	4.78	4.37	4.48	.047	.016	-.064	-.060	.06	-.01	-.05	-.06	< 1
K	2.65	2.53	2.70	2.73	.099	-.091	-.100	-.113	.10	-.09	-.10	-.12	1.20
L	2.80	2.79	2.72	2.82	.079	-.041	-.051	-.127	.08	-.04	-.05	-.13	1.43

[a]N for C = 90, for E_1 = 32, for E_2 = 39, for E_3 = 15.

*Significant at the .05 level, with df = 3, 169.

Table 11

Correlations of Obtained with Hypothesized Rank Orders of Adjusted Mean Departures of Obtained from Predicted post-ACT

Item	Group[a] E_1	E_2	E_3	Rho	Is Rho in Hypothesized Direction?
A	3	2	1	1.0	Yes
B	3	2	1	1.0	Yes
C	3	2	1	1.0	Yes
D	2	3	1	.5	Yes
E	3	2	1	1.0	Yes
F	2	1	3	-.5	No
G	1	2	3	-1.0	No
H	2	3	1	.5	Yes
I	1	2	3	-1.0	No
J	1	2	3	-1.0	No
K	3	2	1	1.0	Yes
L	3	2	1	1.0	Yes

[a]Hypothesized order was 3-2-1. Rank 1 was assigned to the largest departure in the algebraically negative direction, etc., except for Item J, where Rank 1 was assigned to the largest departure in the algebraically positive direction, because for Item J the mean pre-IDL values are numerically larger than the mean pre-ACT values.

obtained in the experimental group with those in our control group. Only if the former correlations are consistently higher (algebraically) can we infer that the variance of the mean pre-IDLs communicated to teachers in the experimental group accounted for some of the variance in their departures from predicted post-ACT.

The correlations are shown in Table 12. No consistent trends are apparent. The experimental group's rs are algebraically higher than those of the control group for only four items, or about a chance number. Nor does the r consistently rise or fall as we go from E_1 to E_3; the rank correlations of these rs with a hypothesized order giving E_3 a rank of 1, etc., are positive for six items and negative for six. In short, the data provide no evidence that variance in pre-IDL accounts for some of the variance in departure of obtained post-ACT from predicted post-ACT in the experimental group.

Correlational Indices of the Effect of Feedback

It seemed possible that the feedback would lower the correlations between pre-ACT and post-ACT. This effect of the feedback would presumably result from greater changes in post-ACT in the experimental group, unrelated to pre-ACT. But if the changes in post-ACT tended to be fairly constant for all teachers, the mean post-ACT could be higher or lower than the mean pre-ACT without, of course, systematically affecting the correlation between pre- and post-ACT.

At any rate, the correlations between pre- and post-ACT are shown in Table 13. Eleven of the 12 rs in the control group are larger than the corresponding rs in Group E_1. But the differences between the control group and Groups E_2, E_3, and E_T are not nearly as consistent and can readily be ascribed to chance. For example, only six of the 12 rs in the control group are larger than the corresponding rs in the experimental total group. Apparently, the feedback had no consistent effect on the correlation between pre- and post-ACT.

It also seemed possible that the effect of feedback would manifest itself correlations between IDL and ACT means. In previous sections we have

Table 12

Correlations between (a) pre-IDL and
(b) Departure of Obtained post-ACT from Predicted post-ACT

		Group				
		C	E_1	E_2	E_3	E_T
Item	N	90	32	39	15	86
A		.15	-.01	.37	.33	.29
B		.07	.26	.10	.09	.16
C		.15	.28	-.05	.07	.11
D		.35	-.22	-.09	.40	-.07
E		.07	.34	.36	.05	.31
F		.08	-.14	.09	.36	.04
G		.17	.06	.41	-.40	.07
H		.11	-.14	.27	.11	.06
I		.07	-.25	.04	.15	-.10
J		.25	.02	-.04	.33	.06
K		.03	.27	.10	-.04	.13
L		.14	.34	-.05	-.31	.05

Table 13

Correlations between pre-ACT and post-ACT

	Group				
Item	C	E_1	E_2	E_3	E_T
A	.79	.74	.86	.92	.84
B	.50	.37	.46	.45	.43
C	.79	.62	.50	.62	.56
D	.77	.58	.61	.20	.33
E	.43	.70	.77	.33	.72
F	.67	.57	.72	.85	.67
G	.83	.68	.81	.47	.70
H	.73	.42	.85	.80	.75
I	.65	.62	.62	.85	.68
J	.78	.63	.62	.79	.66
K	.68	.63	.78	.14	.68
L	.73	.71	.55	.09	.56

seen that the differences between pre-IDL and adjusted post-ACT means were systematically smaller in the experimental group and that the decrease in this difference was related to interval. Such closer approximation of the ACT mean to the IDL mean might be considered to betoken an increase in "satisfaction with the teacher" on the part of the pupils. Would the correlation between ACT and IDL also show such greater closeness, or what might be called greater satisfaction? Such correlations would of course reflect a different kind of proximity of ACT to IDL, one based on covariance rather than on similarity of average levels. And it is questionable whether such covariance is a relevant measure of satisfaction, since the difference between mean IDL and mean ACT could be large even when the correlation is almost perfect, and the correlation could be very low or zero even when the means are equal.

Table 14 shows the correlations between pre-ACT and pre-IDL, the correlations between post-ACT and post-IDL, and the change in correlation from pre- to post-test. In the experimental total group, the rs for ten of the 12 items change to higher values in the posttest. But this is also true for nine items in the control group. And only seven of the items, about a chance proportion, show a greater increase in r in the experimental group than in the control group. The mean difference between the experimental and control groups in the amount of change in r is .07, with the greater increase in r appearing in the experimental group, but this value is not significantly different from zero. The changes in rs in the experimental-interval subgroups show no consistent trend toward greater increases in r with increasing interval, only four of the items showing rank order correlations of 1.0 with the hypothesis that the changes would increase from E_1 to E_3.

It might be argued that the increase in correlation of post-ACT with pre-IDL is more reasonably to be expected, since it was pre-IDL and not post-IDL that was contained in the feedback. To check on this possibility we computed the rs between pre-IDL and post-ACT. When these rs were compared with those

Table 14

Correlations between pre-ACT and pre-IDL,
post-ACT and post-IDL, and the Changes in Correlation[a]

	r between pre-ACT and pre-IDL					r between post-ACT and post-IDL					Difference between r pre and r post				
Item	C	E_1	E_2	E_3	E_T	C	E_1	E_2	E_3	E_T	C	E_1	E_2	E_3	E_T
A	.60	.59	.38	.73	.50	.63	.79	.66	.90	.74	.03	.20	.28	.17	.24
B	.53	.40	.64	.39	.50	.72	.70	.38	.37	.51	.19	.30	-.26	-.02	.01
C	.69	.50	.59	.62	.55	.79	.80	.65	.84	.74	.10	.30	.06	.22	.19
D	.45	.18	-.07	-.19	.30	.49	.46	.31	.41	.35	.04	.28	.38	.60	.05
E	.34	.45	.25	.30	.33	.60	.52	.67	.21	.54	.26	.07	.42	-.09	.21
F	.50	.64	.46	.77	.60	.59	.72	.42	.77	.61	.09	.08	-.04	.00	.01
G	.59	.60	.78	.65	.67	.67	.44	.69	.70	.62	.08	-.16	-.09	.05	-.05
H	.69	.72	.69	.08	.55	.65	.67	.86	.66	.80	-.04	-.05	.17	.58	.25
I	.32	.60	.16	.55	.44	.38	.50	.24	.57	.42	.06	-.10	.08	.02	-.02
J	.73	.59	.44	.72	.51	.71	.68	.71	.88	.73	-.03	.09	.27	.16	.22
K	.64	.60	.23	.09	.31	.62	.75	.54	.64	.63	-.02	.15	.31	.55	.32
L	.53	.55	.39	.42	.44	.69	.89	.69	.42	.77	.16	.34	.30	.00	.33

[a] N for C = 90, for E_1 = 32, for E_2 = 39, for E_3 = 15, for E_T = 86.

between pre-IDL and pre-ACT, the increases in r were not consistently greater in the experimental group.

Teacher Protocols

Four protocols were obtained from teachers:

pre-PERC -- the teacher's perception (estimate) of how a pupil who is typical of the majority of the class would answer the item on the pretest

post-PERC -- the teacher's perception (estimate) of how a pupil who is typical of the majority of the class would answer the item on the posttest

pre-SELF -- the teacher's indication of how much the item was "like" herself on the pretest

post-SELF -- the teacher's indication of how much the item was "like" herself on the posttest

These protocols made possible a study of the extent to which feedback produced changes in teachers' perceptions of their pupils and themselves with respect to the 12 items of teacher behavior. Changes in teacher behavior as described by pupils, of the kind we have considered in the section on pupil protocols, need not of course be accompanied by changes in how the teachers view themselves and their pupils' perceptions. But evidence on these perceptions on the teacher's side would throw light on whether teachers' self-perceptions did change, and on whether teachers changed their perceptions of their pupils' perceptions of them.

In this section we deal with changes in the teachers' protocols in themselves, independently of any relations (such as "accuracy") to the pupil protocols. The latter relations are examined in the next section.

Adjusted post-PERC

Table 15 shows the results of 12 analyses of covariance in which post-PERC was the dependent variable, feedback was the experimental variable, and adjustments were made to control for between-groups variance in pre-PERC.

For five of the 12 items, the differences in adjusted post-PERC means of the experimental and control groups are statistically significant at the .05 level (two-tailed) or better. The direction of the differences, whether

Table 15

Means of Teachers' Perceptions
of Their Pupils' Majority Responses

Item	pre-PERC		post-PERC		Adjusted post-PERC		Difference betw. Exp. and Cont. in Adjusted post-PERC	t^a
	Exp.	Cont.	Exp.	Cont.	Exp.	Cont.		
A	2.24	2.04	2.28	2.07	2.22	2.13	.09	< 1
B	2.15	2.39	2.39	2.35	2.43	2.31	.12	1.10
C	2.19	2.35	2.19	2.02	2.23	1.98	.25	1.69*
D	1.65	1.57	1.42	1.58	1.40	1.60	-.20	1.78*
E	2.28	2.24	2.07	2.11	2.06	2.12	-.06	< 1
F	3.56	3.31	3.90	3.44	3.84	3.51	.33	1.76*
G	3.28	3.19	3.12	3.21	3.10	3.24	-.14	< 1
H	1.81	1.70	1.79	1.64	1.77	1.65	.12	1.20
I	4.72	4.54	4.75	4.63	4.73	4.65	.08	< 1
J	3.48	3.24	3.66	3.64	3.60	3.70	-.10	< 1
K	2.88	2.76	2.51	2.81	2.46	2.86	-.40	2.05**
L	2.68	3.10	2.88	2.86	3.04	2.71	.33	1.84*

[a] t was computed as $\sqrt{F}$ with df = 173.

*Significant at the .10 level (two-tailed).

**Significant at the .05 level (two-tailed).

significant or non-significant, is not consistent; five go in one direction and seven in the other. Without some basis other than direction for interpreting these differences, we cannot say at this point what they mean, except that five of the 12 are probably not due merely to chance. We shall later compare these "perceptions" of the teachers with other protocols to see what they mean in terms of accuracy, etc. Suffice it now to say that something, presumably the feedback, made the teachers in the experimental and control groups "perceive" their pupils' descriptions differently at the time of the posttest, on five of the items.

Adjusted post-PERC by interval. Were the adjusted post-PERC means different according to the interval between feedback and posttest? Table 16 shows that analysis of covariance yielded significant between-groups variance in post-PERC, adjusted for pre-PERC variance, for Items F and H. But there is no consistency in the direction of the differences between the interval groups in mean adjusted post-PERC scores. Further interpretation of these means is deferred until we examine relationships between teacher and pupil protocols.

Adjusted post-SELF

Did the feedback produce differences between the experimental and control groups in how they described themselves on the 12 items? Table 17 shows the results of the relevant analyses of covariance. Differences significant at the .10 and .05 levels occurred for three items: C, F, and I. But the direction of these differences was not consistent from one item to the next.

Adjusted post-SELF by interval. As is shown in Table 18, the analyses of covariance yielded only one significant between-groups variance in adjusted post-SELF mean, that for Item L. No substantial evidence appeared for systematic changes in mean SELF perception as a function of feedback interval.

Correlational Indices of Effect of Feedback

...ld the correlations between pre- and post-PERC differ consistently the experimental and control groups?--between experimental-interval

Table 16

Means of Teachers' Perceptions of
Their Pupils' Majority Responses, According to Interval

Item	pre-PERC				post-PERC				Adjusted post-PERC				F^a
	C	E_1	E_2	E_3	C	E_1	E_2	E_3	C	E_1	E_2	E_3	
A	2.04	2.25	2.18	2.36	2.04	2.13	2.33	2.50	2.12	2.05	2.31	2.36	< 1
B	2.39	2.13	2.23	2.00	2.35	2.48	2.26	2.57	2.31	2.53	2.27	2.67	1.06
C	2.35	2.06	2.44	1.79	2.02	2.23	2.28	1.86	1.98	2.33	2.20	2.51	1.08
D	1.57	1.48	1.82	1.57	1.58	1.42	1.54	1.07	1.60	1.48	1.45	1.09	2.00
E	2.24	2.41	2.36	1.71	2.11	2.22	2.10	1.64	2.12	2.15	2.05	1.89	< 1
F	3.31	3.52	3.64	3.50	3.44	3.48	4.00	4.57	3.51	3.44	3.89	4.54	3.82*
G	3.19	3.52	3.15	3.43	3.21	3.16	3.05	3.29	3.24	3.02	3.09	3.20	< 1
H	1.70	1.72	2.00	1.50	1.64	1.69	2.00	1.43	1.65	1.70	1.98	1.49	3.83*
I	4.54	4.38	4.87	5.07	4.63	4.47	4.87	5.07	4.65	4.54	4.80	5.19	< 1
J	3.24	3.72	3.31	3.36	3.64	3.97	3.62	3.07	3.70	3.77	3.64	3.07	1.13
K	2.76	2.65	3.26	2.71	2.81	2.71	3.18	2.43	2.86	2.84	2.92	2.51	< 1
L	3.10	2.58	2.67	2.93	2.86	2.77	2.72	3.57	2.71	3.01	2.89	3.55	2.20

[a] $df = 3, 173$; $F_{.05} = 2.66$, $F_{.01} = 3.90$.

*Significant at the .05 level.

Table 17

Means of Teachers' SELF Descriptions

Item	pre-SELF Exp.	pre-SELF Cont.	post-SELF Exp.	post-SELF Cont.	Adjusted post-SELF Exp.	Adjusted post-SELF Cont.	Difference between Exp. and Cont. in Adjusted post-SELF	t[a]
A	1.85	1.82	2.08	1.94	2.07	1.95	.12	< 1
B	1.96	2.00	1.96	1.96	1.95	1.95	.00	< 1
C	2.18	2.22	1.71	2.00	1.72	1.98	-.26	2.05**
D	1.51	1.43	1.35	1.31	1.34	1.32	.02	< 1
E	2.01	1.83	1.94	1.91	1.88	1.97	-.09	< 1
F	3.25	2.98	3.70	3.24	3.63	3.31	.32	1.90*
G	3.06	3.01	3.04	3.06	3.02	3.07	-.05	< 1
H	1.47	1.45	1.59	1.47	1.58	1.48	.10	1.14
I	4.54	4.50	4.94	4.58	4.93	4.59	.34	1.88*
J	3.86	3.97	3.85	3.91	3.88	3.88	.00	< 1
K	2.61	2.48	2.66	2.55	2.61	2.60	.01	< 1
L	2.68	2.99	2.67	3.11	2.76	3.02	-.26	1.58

[a] t was computed as $\sqrt{F}$ with df = 173.

*Significant at .10 level (two-tailed).

**Significant at .05 level (two-tailed).

Table 13

Means of Teachers' SELF Descriptions, According to Interval

Item	pre-SELF				post-SELF				Adjusted post-SELF				F^a
	C	E_1	E_2	E_3	C	E_1	E_2	E_3	C	E_1	E_2	E_3	
A	1.82	1.78	1.79	2.14	1.92	2.09	2.05	2.14	1.93	2.13	2.08	1.91	< 1
B	2.00	1.94	2.00	1.86	1.96	1.94	1.90	2.21	1.95	1.92	1.89	2.27	< 1
C	2.24	2.13	2.38	1.64	2.00	2.06	2.13	1.71	1.98	2.09	2.02	2.03	< 1
D	1.52	1.53	1.54	1.43	1.46	1.31	1.46	1.21	1.46	1.30	1.45	1.28	1.41
E	1.83	2.28	1.95	1.57	1.91	2.13	1.82	1.93	1.97	1.98	1.80	2.18	1.33
F	3.01	3.16	3.31	3.14	3.27	3.41	3.72	3.93	3.32	3.38	3.62	3.91	1.24
G	3.01	3.22	3.03	2.79	3.02	3.03	2.92	3.29	3.03	2.94	2.93	3.41	< 1
H	1.45	1.44	1.56	1.29	1.47	1.44	1.77	1.43	1.48	1.45	1.71	1.53	1.79
I	4.50	4.41	4.56	4.79	4.58	4.84	4.95	5.00	4.59	4.90	4.93	4.87	< 1
J	3.97	4.06	3.62	3.79	3.91	4.06	3.67	3.79	3.86	3.96	3.83	3.85	< 1
K	2.48	2.56	2.67	2.57	2.54	2.78	2.90	2.29	2.58	2.77	2.81	2.27	1.16
L	3.03	2.81	2.62	2.57	3.11	2.59	2.51	3.29	3.01	2.62	2.67	3.47	2.91*

[a] df = 3, 167; $F_{.05} = 2.66$, $F_{.01} = 3.90$.

*Significant at the .05 level.

subgroups? Such differences would indicate that the feedback upset the relationship between pre- and post-PERC that would normally (i.e., without feedback) prevail.

Table 19 shows the rs obtained between pre- and post-PERC. Comparing the respective rs in the control and experimental groups, we find six larger in one group and six in the other--a chance split. But when the experimental group is broken into the three subgroups by interval, we find five rs smaller in E_1 than in the control group, seven in E_2, and nine in E_3. The trend is in the hypothesized direction: the longer the interval, the lower the r between pre- and post-PERC. Perhaps, with longer intervals, we would find the feedback consistently lowering the r below the norm provided by the control group.

Table 20 shows the results of a similar correlational analysis of pre- and post-SELF. Does the feedback reduce the stability of teachers' views of themselves? No consistent differences in the size of the rs from one group to another appear in these results to support an affirmative answer. The differences between the experimental and control groups are about as often in one direction as in the other. The same is true of the rs in the experimental-interval subgroups.

Our final question concerning the teacher protocols in themselves deals with a form of what has been called "assumed similarity:" the degree to which a person assumes another to have the same opinions as himself. One measure of assumed similarity averaged over teachers is the correlation between the teachers' self-descriptions and those they estimated the majority of their pupils to make of them. A high correlation between pre-SELF and pre-PERC, for example, would mean that teachers in the group tended to rate themselves and predict their pupils to rate their teacher at about the same relative level on the item of teacher behavior.

The pertinent correlations for examining the teachers' assumed similarity are given in Table 21. The rs for all 12 items, for both experimental and

Table 19

Correlations between pre-PERC and post-PERC

Item	Group				
	C	E_1	E_2	E_3	E_T
A	.41	.72	.85	.80	.79
B	.24	.55	.62	.15	.53
C	.61	.60	.64	.97	.68
D	.47	.69	.28	.19	.41
E	.54	.59	.50	.41	.55
F	.60	.54	.54	.55	.52
G	.62	.41	.51	.53	.48
H	.41	.79	.65	.69	.65
I	.62	-.06	.61	.50	.32
J	.50	.33	.65	.48	.52
K	.71	.81	.53	.60	.62
L	.68	.72	.49	.31	.53

Table 20

Correlations between pre-SELF and post-SELF

Item	Group				
	C	E_1	E_2	E_3	E_T
A	.60	.60	.69	.81	.67
B	.46	.82	.60	.40	.63
C	.64	.76	.62	.70	.68
D	.77	.69	.72	.48	.67
E	.62	.70	.56	.73	.63
F	.54	.56	.34	.17	.53
G	.55	.73	.31	.55	.52
H	.64	.50	.68	.09	.59
I	.67	.09	.46	.52	.25
J	.47	.57	.64	.84	.66
K	.69	.87	.79	.71	.80
L	.68	.56	.55	.69	.56

Table 21

Correlations between Teachers' SELF Descriptions and PERC: Assumed Similarity

Item	$r_{pre\text{-}SELF\ vs.\ pre\text{-}PERC}$		$r_{post\text{-}SELF\ vs.\ post\text{-}PERC}$		Change	
	C	E_T	C	E_T	C	E_T
A	.56	.65	.67	.84	.11	.19
B	.67	.71	.53	.62	-.14	-.09
C	.65	.75	.75	.81	.10	.06
D	.79	.48	.80	.66	.01	.18
E	.61	.62	.64	.65	.03	.03
F	.65	.76	.76	.71	.11	-.05
G	.66	.63	.77	.81	.11	.18
H	.58	.69	.59	.68	.01	-.01
I	.83	.66	.85	.61	.02	-.05
J	.53	.72	.72	.79	.19	.07
K	.84	.82	.87	.85	.03	.03
L	.79	.71	.84	.79	.05	.08

control groups, tend to be high and positive. Response set, i.e., a form of test-taking habit characterized by reliable individual differences, no doubt produced much of this high correlation; confounded with such sets as a cause of the high degree of assumed similarity is a genuine psychological tendency to ascribe one's own beliefs to positively valued other persons.

For our purposes, however, in evaluating the effect of our feedback, we are more interested in differences between the control and experimental groups in the amount and direction of change from pre- to post-test. In the control group 11 of the 12 rs increased from pre- to post-test: in the experimental group, eight of the 12 rs increased. By and large, the groups do not differ appreciably or consistently in their levels of assumed similarity or in the amounts or directions of change from pre- to post-test.

When the experimental interval subgroups are examined, using the rs shown in Table 22, a slight trend does appear in the direction of change toward less assumed similarity in the longest interval (E_3) group. All 12 of the E_3 changes are less positive or more negative than those in E_1; nine of them bear a similar relation to the changes in the control group. There appears to be a distinct effect of the feedback, combined with a longer interval before the posttest, toward reducing the amount of assumed similarity that teachers would otherwise manifest. Since assumed similarity often has a kind of "autism" inherent in it, the feedback may be considered to be producing less autism. Whether this change away from autism is tantamount to be improved contact with social reality will be considered when we examine the "accuracy" of the teachers in predicting their pupils' responses.

Relations between the Protocols of Pupils and Teachers

So far we have examined differences between our experimental and control groups in the protocols of pupils and teachers separately. Now what of th ions between these protocols? Such relations can take the following fo

Table 22

Correlations between Teachers' SELF and PERC, on the Basis of Interval

Item	$r_{\text{pre-SELF vs. pre-PERC}}$				$r_{\text{post-SELF vs. post-PERC}}$				Changes from $r_{\text{pre-SELF vs. pre-PERC}}$ to $r_{\text{post-SELF vs. post-PERC}}$			
	C	E_1	E_2	E_3	C	E_1	E_2	E_3	C	E_1	E_2	E_3
A	.56	.64	.58	.86	.67	.83	.86	.83	.11	.19	.28	-.03
B	.67	.60	.77	.42	.53	.68	.66	.27	-.14	.08	-.11	-.15
C	.65	.70	.75	.86	.75	.80	.82	.91	.10	.10	.07	.05
D	.79	.62	.42	.47	.80	.78	.59	.53	.01	.16	.17	.06
E	.61	.54	.66	.73	.64	.65	.69	.77	.03	.11	.03	.04
F	.65	.69	.75	.92	.76	.75	.74	.47	.11	.06	-.01	-.45
G	.66	.44	.80	.72	.77	.75	.79	.95	.11	.31	-.01	.23
H	.58	.75	.68	.50	.59	.68	.69	.42	.01	-.07	-.01	-.08
I	.83	.55	.71	.81	.85	.61	.63	.58	.02	.06	-.08	-.23
J	.53	.57	.69	.93	.72	.86	.80	.70	.19	.29	.11	-.23
K	.84	.91	.79	.82	.87	.83	.89	.62	.03	-.08	.10	-.20
L	.79	.75	.69	.75	.84	.72	.79	.84	.05	-.03	.10	-.19

ACC -- the ACCuracy of the teacher's perception or estimation (PERC) of the typical pupil's response, as measured by its closeness to the pupils' median description of their actual teacher (ACT)

SIM -- the SIMilarity between the teacher's description of herself (SELF) and the pupils' median description of their actual teacher (ACT)

It would also be possible to study relations between the pupils' descriptions of their ideal teacher (IDL) and the teacher protocols (PERC and SELF); these relations are, however, being disregarded at this point.

Adjusted post-ACC

One measure of accuracy can be obtained by computing the difference between the mean adjusted post-ACT of the pupils and their teacher's adjusted post-PERC. The two adjusted values entering into this adjusted post-ACC score have been listed for each item in earlier tables (Tables 3, 7, 15, and 16). Taking the difference between them gives us the ACC measures shown in Table 23.

It seems reasonable to expect the experimental group to be more accurate; simply remembering or extrapolating from the information concerning pre-ACT provided by the feedback should have enabled these teachers to estimate post-ACT more accurately, as compared with the control group which received no such information.

Table 23 shows that this expectation is for the most part borne out; for nine of the 12 items the mean adjusted post-ACC of the experimental group is better (i.e., shows a lower absolute difference between ACT and PERC) than that of the control group. When we tested the significance of these differences over all 12 items, the t of 1.89 (df = 11) was significant at the .05 level (one tail). This estimate of significance is of course an overestimate, since the items are not experimentally independent.

Adjusted post-ACC by interval. Are teachers more accurate in predicting their post-ACT ratings by pupils when a shorter or a longer interval has intervened by the feedback and the post-test? Higher accuracy might go with the

Table 23

Accuracy (ACC) Based on the Difference between Means in Adjusted post-ACT and Adjusted post-PERC, With and Without Intervals

Item	Without Interval Adjusted post-ACC		Is Difference between Columns 2 and 3 in Hypothesized Direction?	With Interval Adjusted post-ACC			
	C^a	E_T		C^a	E_1	E_2	E_3
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A	.44	.35	Yes	.44	.57	.25	.06
B	1.30	1.03	Yes	1.30	1.00	.16	.72
C	.44	.10	Yes	.44	.05	.13	-.28
D	-.42	-.25	Yes	-.45	-.35	-.27	-.02
E	.18	.16	Yes	.18	.10	.17	.27
F	.69	.23	Yes	.61	.66	.05	-.22
G	.22	.25	No	.23	.20	.21	.53
H	.51	.42	Yes	.52	.50	.23	.60
I	.65	.56	Yes	.65	.70	.40	.14
J	.63	.67	No	.63	.53	.59	1.17
K	.58	.78	No	.57	.40	.31	.70
L	.25	-.20	Yes	.26	-.14	-.05	-.78

[a]Occasional discrepancies between values for the control group on the same item in the two analyses of covariance result from using different within-groups regression coefficients in computing the adjusted posttest means.

shorter interval, since less forgetting of the feedback might then occur. (This expectation would be the opposite of that for change-ACT, where we reasoned that more change in the teacher's behavior, as described by pupils, would probably occur in the subgroup of teachers with the longer interval.) On the contrary, it could be argued that the longer interval would give teachers more opportunity to internalize the feedback and allow it to guide their perceptions.

As can be seen in Table 23, nine out of 12 of the adjusted post-ACC means in E_1 are better (smaller) than those of the control group; 12 out of 12 in E_2; and only seven out of 12 in E_3. There is thus a slight tendency for the E_2 group to be more accurate on the posttest than either E_1 or E_3.

Accuracy in correlational terms. Accuracy can be measured in one form by the closeness of two averages, and in another form by the size of a correlation coefficient. Accuracy of the latter kind reflects how well the relative position of the teacher's PERC followed that of her pupils' ACT (description of their actual teacher) on a given item. Table 24 shows the rs obtained between PERC and ACT for the various groups on the pretest and posttest. The right-hand section of Table 24 shows the changes in these rs from pre- to post-test. Eight of the items show changes that were more positive or less negative in the experimental group than those of the control group. A Wilcoxon matched-pairs test shows these changes to be significantly different in the two groups, p being less than .025. Measured in correlational terms, accuracy again seems to improve as a result of feedback.

And when the changes in correlational accuracy are compared in the experimental-interval subgroups, it again appears that the E_2 group gained most consistently; it shows greater gain than the control group on 10 items, while E_1 does so on only four items, and E_3 on seven.

Adjusted post-SIM

Similarity of teachers' self-descriptions to their pupils' descriptions of them can also be measured both as a difference between means and as a correlation. The difference on each item between the adjusted post-SELF and the

Table 24

Accuracy (ACC) Based on the Correlation between PERC and ACT

Item	pre-ACC: $r_{pre\text{-}PERC\ vs.\ pre\text{-}ACT}$					post-ACC: $r_{pre\text{-}PERC\ vs.\ pre\text{-}ACT}$					change-ACC: post-ACC minus pre-ACC					Is C - E_T Difference in Hypothesized Direction?
	C	E_1	E_2	E_3	E_T	C	E_1	E_2	E_3	E_T	C	E_1	E_2	E_3	E_T	
A	.17	.08	.50	.21	.21	.44	.25	.56	.25	.42	.27	.17	.06	.04	.21	No
B	.20	.05	-.03	-.09	-.02	-.10	.40	.14	-.17	.20	-.30	.35	.17	-.08	.22	Yes
C	.29	.29	.16	.50	.23	.29	.11	.28	.66	.27	.00	-.18	.12	.16	.04	Yes
D	.13	.14	.03	-.31	-.03	.49	.18	.70	.56	.54	.35	.04	.67	.87	.57	Yes
E	.16	.23	.08	-.11	.17	.06	.22	.27	.56	.29	-.10	-.01	.19	.67	.12	Yes
F	.16	.07	.03	.75	.17	.15	.38	.07	.55	.26	-.01	.31	.04	-.20	.09	Yes
G	.32	.43	.47	.36	.44	.30	.31	.40	.25	.35	-.02	-.12	-.07	-.11	-.09	No
H	.36	-.04	.45	.15	.28	.15	-.14	.40	.10	.27	-.21	-.10	-.05	-.05	-.01	Yes
I	.25	.06	.06	.01	.10	.24	.34	.10	-.21	.22	-.01	.28	.04	-.22	.12	Yes
J	.19	-.11	.06	.15	.01	.27	-.09	.27	.04	.08	.08	.02	.21	-.11	.07	No
K	.15	.58	.30	-.11	.36	.28	.50	.59	.09	.54	.13	-.08	.29	.20	.18	Yes
L	.17	.54	.18	.19	.30	.23	.35	.32	.27	.33	.06	-.19	.14	.08	.03	No

adjusted post-ACT means is called adjusted post-SIM. As shown in Table 25, the experimental group manifests greater similarity in this sense, presumably as a result of the feedback. Eleven of the 12 items show greater adjusted post-SIM for the experimental group. The difference between the means of the two arrays of 12 adjusted post-SIM values is significant at the .025 level on a one-tail basis, using a t-test with 11 df.

Adjusted post-SIM by interval. The experimental-interval subgroups do not seem to differ consistently in adjusted post-SIM. All three subgroups have consistently smaller values than the control group, but they do not differ consistently among themselves. Apparently feedback increased similarity to about the same degree for all intervals between feedback and posttest.

Similarity in correlational terms. To what extent do the experimental and control groups differ when change in similarity is measured by the change in correlation between SELF and ACT from pretest to posttest? Table 26 shows that the experimental group of teachers gained more in similarity to its pupils on only seven of the 12 items. But the gains on these items were so much larger than the losses on the other five items that the difference between the change-SIM values of the experimental and control groups is significant at the .025 level, with $t = 2.67$ and $df = 11$.

Results with Relative Data

Thus far this report has presented results obtained with irrelative data, i.e., with pupils' and teachers' responses to the items of teacher behavior where each item was to be considered independently of the others. The response was made by choosing one of six alternatives: e.g., "Very much LIKE my teacher," "Somewhat LIKE my teacher," "A little bit LIKE my teacher," "A little bit UNLIKE my teacher," etc.

It will be recalled that all protocols were also collected in relative form. In this form, the items were presented in triads. The pupil or teacher checked one item of the triad as "MOST," and another an "LEAST"--"like you"

Table 25

Similarity (SIM) Based on the Difference between Means in Adjusted post-ACT and Adjusted post-SELF, With and Without Intervals

Item	Without Interval Adjusted post-SIM		Is C - E_T Difference in Hypothesized Direction?	With Interval Adjusted post-SIM			
	C	E_T		C	E_1	E_2	E_3
A	.65	.54	Yes	.63	.49	.48	.51
B	1.63	1.48	Yes	1.66	1.61	1.54	1.12
C	.49	.46	Yes	.44	.29	.31	.20
D	-.13	-.19	No	-.31	-.17	-.27	-.21
E	.32	.32	Tie	.33	.27	.42	-.02
F	.88	.44	Yes	.80	.72	.31	.41
G	.43	.33	Yes	.44	.27	.38	.32
H	.64	.64	Tie	.69	.75	.50	.56
I	.73	.25	Yes	.71	.34	.27	.46
J	.44	.37	Yes	.47	.34	.40	.39
K	.89	.66	Yes	.85	.47	.42	.84
L	-.08	.08	Tie	-.04	.23	.17	-.70

Table 26

Similarity (SIM) Based on the Correlation between SELF and ACT

Item	pre-SIM $r_{\text{pre-ACT vs. pre-SELF}}$					post-SIM $r_{\text{post-ACT vs. post-SELF}}$					change-SIM post-SIM minus pre-SIM				
	C	E_1	E_2	E_3	E_T	C	E_1	E_2	E_3	E_T	C	E_1	E_2	E_3	E_T
A	.37	.08	.51	-.01	.23	.42	.12	.50	-.14	.27	.05	.04	-.01	-.13	.04
B	.14	.19	.11	.05	.12	.17	.56	.32	-.71	.25	.03	.37	.21	-.76	.13
C	.36	.23	.22	.62	.24	.27	.20	.28	.59	28	-.09	-.03	.06	-.03	.04
D	.30	.25	-.01	-.30	.00	.36	.05	.32	.14	.23	.06	-.20	.33	.44	.23
E	.14	.05	.17	.01	.13	.08	.13	.17	.54	.18	-.06	.08	.00	.53	.05
F	.18	-.02	.09	.80	.16	.16	.30	-.09	.33	.13	-.02	.32	-.18	-.47	-.03
G	.31	.43	.39	.27	.39	.26	.21	.41	.31	.34	-.05	-.21	[illegible]	.04	-.05
H	.32	.22	.66	.13	.50	.17	.00	.50	.05	.35	-.15	-.22	-.16	-.08	-.15
I	.14	.16	.02	.20	.13	.20	.59	.28	-.11	.39	.06	.43	.26	-.31	.26
J	.26	-.22	.19	.36	.09	.30	-.10	.26	.13	.09	.04	.12	.07	-.23	.00
K	.22	.60	.52	.01	.48	.30	.59	.60	.38	.59	.08	-.01	.08	.37	.11
L	.14	.25	.40	.03	.26	.19	.34	.33	.49	.36	.05	.09	-.07	.46	.10

(for teachers' SELF); "like what your typical pupil would say about you" (for teachers' PERC); "like my teacher" (for pupils' ACT); "like a best teacher" (for pupils' IDL). In constructing the triads, the 12 items of teacher behavior were divided into two groups of six--Items A-F and Items G-L. Each group of six made possible 20 triads (of which a balanced set of 10 was used) whose items were thus interdependent with one another but independent of those in the other group of six. We thus have two alternate and presumably equivalent groups of six items arranged in the form of 10 triads each. The correlations between these "equivalent" forms can yield estimates of the reliability--in the sense of equivalence--of any score based on a group of six items in the relative--or triadic--form. And the two groups of items permit us to obtain twice any difference between experimental and control groups involving a given score based on the six items in the relative format--as against the 12 differences possible when the items were used singly, in irrelative form.

How does the relative format differ from the irrelative? First, it yields rank orderings of six items; hence all subjects' responses have the same mean and dispersion of the ranks. Individual differences in the mean and dispersion of responses cannot occur. All pupils, for example, are forced to discriminate among a given group of six items, but the whole set cannot differ from one pupil to another in the <u>average</u> degree to which pupils say the items are "like" their teacher. Similarly, the pupils must be equal in the differences they attach to the items in how much they are "like" the teacher. Simply stated, the pupils' and teachers' responses to 10 of the 20 triads--made up of six items in balanced combinations of three--were always reduced to a rank order: 1, 2, 3, 4, 5, and 6.

A second difference between the relative and irrelative data is that the subject's ordering of a set of six items could readily be considered as a whole rather than as a set of six individual responses. In effect, the subject produced a "profile," or pattern of scores on the six items; the level and scatter of the profile were the same for all subjects, and only differences in

shape were possible. Response sets--such as "acquiescence"--conducive to individual differences in the level or dispersion of responses cannot operate on the relative protocols.

We shall present our scoring methods and the results of analyses of the "reliability" of the relative data. Then we shall examine differences between the experimental and control groups in such variables as pupil satisfaction, change in actual teacher, change in SELF, and change in PERC.

Scoring the Relative Data

To score the relative protocols, each subject's responses to the appropriate 10 triads was first converted into a rank ordering of the six component items. For example, a pupil's responses, in describing his actual teacher, to the 10 triads composed of Items A-F might be converted into the following ranking of Items A-F

Item:	A	B	C	D	E	F
Rank:	1	4	3	2	6	5

This conversion of triad responses into ranks was done by counting the number of times each item was considered by the pupil to be "more like my teacher" than each other item in the set of six. The item thus "preferred" most often to all other items was given Rank 1; next most often was given Rank 2; etc.

Now, for example, we obtained one such ranking by each pupil of Items A-F as to how well these items characterized his actual teacher. We also obtained such a ranking by each pupil of Items A-F as to how well these items characterized his ideal teacher. These two rankings of Items A-F can of course be correlated; in this case, we used the term satisfaction (SAT) to refer to the tau (Kendall's coefficient of rank correlation) between the two rank orderings. This tau was converted into a score according to the following table:

Tau Range[4]	Midpoint	Score
1.00	1.00	0
.733 to .933	.833	1
.467 to .667	.567	2
.200 to .400	.300	3
-.067 to .133	.033	4
-.333 to -.133	-.233	5
-.600 to -.400	-.500	6
-.867 to -.667	-.767	7
-1.000 to -.933	-.967	8

Thus the agreement between the pupil's relative protocols for his actual and ideal teachers was expressed as a score from 0 to 8; the lower the score on the 0-8 scale, the higher the positive rank correlation, and the higher the score, the more negative the correlation, with a score of 4 standing for a correlation near zero.

Several scores were obtained on the basis of correlations between the ranks of six items on different protocols. We shall report results from the analysis of the five scores shown below.

Score	Relative Protocols Correlated in Obtaining Scores
1. pre-SAT	pre-ACT vs. pre-IDL
2. post-SAT	post-ACT vs. post-IDL
3. change-ACT	pre-ACT vs. post-ACT
4. change-SELF	pre-SELF vs. post-SELF
5. change-PERC	pre-PERC vs. post-PERC

Reliability of the Scores on Relative Data

The correlations between relative data scores based on Items A-F and those based on Items G-L are shown in Table 27. If it may be assumed that Items A-F and G-L represent equivalent sets in terms of their discriminability, content, social desirability, etc., the rs between scores based on these sets may be considered coefficients of reliability in the sense of equivalence, or what are often called coefficients of equivalence. These coefficients reflect the degree

[4]Gaps occur between the intervals because only certain discrete values uld occur.

Table 27

Correlations between Scores Based on Triads
for Items A-F and G L, Relative Data

	Correlation (r): A-F vs. G-L				
Group:	C	E_1	E_2	E_3	E_T
Score N:	90	32	39	15	86
pre-SAT	.61	.11	.45	.13	.30
post-SAT	.47	.65	.52	.25	.51
change-ACT	.20	-.07	.05	.64	.15
change-SELF	-.11	.08	-.14	.39	.03
change-PERC	-.10	.09	-.16	-.13	-.06

to which a teacher or class who had one standing on a score based on responses to triads of Items A-F tended to have the same relative standing on a score based on Items G-L. If the r is high, it means each score is measuring something reliably. If the correlation is low, it means that the scores are not measuring anything reliably, since the two sets of triads are composed of items assumed to be equivalent in content and statistical properties. (If the assumption of equivalence is abandoned, then the scores may each still be reliable, but we have no evidence of the reliability, and in any case the two scores measure something different.)

The rs in Table 27 show that the reliabilities of the pre-SAT and post-SAT scores ranged from .30 to .61 for the control and E_T groups. The reliabilities of the change-ACT, change-SELF and change-PERC scores were essentially zero, accordingly, further analyses of these latter scores will not be reported.

It is noteworthy that the change-SELF and change-PERC scores are based on the responses of a single individual--the teacher--to six items, which the teacher essentially ranked. Experience has shown that tests of this "length" are seldom much more reliable than the present ones have proved to be. The pre-SAT and post-SAT scores, however, are means of scores based on responses of about 22 pupils on the average. Experience has also shown that mean ratings or rankings of objects, including teachers, by 20 or more judges usually possess substantial reliability, perhaps even higher than those obtained for the pre-SAT and post-SAT scores. As to why the change-ACT scores are so low in reliability, although based on Ns of 22 pupils on the average, we presume that the time interval between the pre- and post-ACT ratings is responsible; not only different items but different occasions are involved in these correlations, while the pre-SAT and post-SAT reliabilities involve only different sets of items, each score being based on data collected on just one occasion.

The coefficients of equivalence between the A-F and G-L items, even in the case of pre-SAT and post-SAT, are rather low in view of the elaborate process in selecting and grouping the items. Three possible explanations for this

come to mind. In the first place, of course, our techniques of item selection may have contained unnoticed faults. Secondly, as we mentioned earlier, the criterion indexes we used in selecting the items were still well below 1.00 for our final 12 items (see Appendix C, Table C-5). Perhaps further refinement in item selection would have yielded higher coefficients of equivalence. Finally, it may be that items differ radically in their meaning from one school room to another. A pair of items easily discriminable (in application to the teacher) for one class of pupils may not be for another. If this is an important factor in reducing the equivalence between two sets of six ranked items, the researcher would need to build separate sets of items for every classroom. The time and expense involved in doing this, not to speak of the demands on the teacher and her pupils, make it easy to see why we rejected this possibility.

At any rate, the data on reliability indicated that only the pre-SAT and post-SAT scores merited further study.

Adjusted post-SAT from Relative Data

What is the meaning of the score for the tau coefficient between the pupil's rank ordering of six items in describing his actual teacher and the same pupil's rank ordering of the same six items in describing his "best imaginable," or ideal, teacher? We have interpreted this score as an index of the pupil's "satisfaction" with the teacher, i.e., the degree to which the pupil sees his teacher's pattern of activities approximating what the pupil desires in a teacher. These scores, as derived from pupils' pre-test protocols, were then averaged over all the teacher's pupils, and the mean was the teacher's pre-SAT score, on Items A-F. By the same method applied to the pupils' post-test protocols, we obtained the teacher's post-SAT score on Items A-F.

Did the experimental and control groups of teachers differ in their pupils' "satisfaction" with them at the end of the experiment? If so, in what direction? Our hypothesis is, of course, that teachers in the experimental group, having received feedback on what their pupils thought their actual and

ideal teachers would be like, would manifest behaviors prior to the posttest that would elicit higher post-SAT scores from their pupils.

In comparing post-SAT scores, we must adjust for differences between the groups in pre-SAT. Analysis of covariance was again used for this purpose. Table 28 shows the results. It should be noted that, in all three pairs of adjusted post-SAT means, the mean for the experimental group is the smaller. A smaller number here means a higher mean tau coefficient. A higher tau in turn signifies closer agreement between the two rank orders involved in adjusted post-SAT measures, namely, post-ACT and post-IDL. In short, the experimental group's behaviors were judged by pupils, at the time of the posttest, to be more in the rank order characteristic of the pupils' ideal teacher, after adjustment for differences in pre-SAT.

The difference between adjusted post-SAT means for Items A-F is significant at the .05 level (one-tail); the differences between means for Items G-L, and for the two sets of items combined (Items A-L) are not significant. Clearly, although the results are suggestive, the hypothesis of greater post-SAT in the experimental group gets only weak support from these relative data.

Adjusted post-SAT by Interval

Is a longer interval between feedback and posttest associated with greater adjusted post-SAT (smaller numerical values) in the experimental group? It will be recalled that evidence favorable to this expectation emerged from analyses of the irrelative data. As shown in Table 29, the relative data go in the same direction. With only one reversal, the mean tau scores for adjusted post-SAT become smaller--and satisfaction hence greater--as we go from E_1 (shortest interval) to E_3 (longest interval). The F-ratios in these analyses of covariance are not, however, statistically significant, with df = 3 and 173. Apparently, the longer the teacher had to assimilate and act upon the feedback, the more she changed in the direction of her pupils' desires. But little confidence can be placed in this finding as yet.

Table 28

Means of Teachers' Tau Scores for the Satisfaction of their Pupils, Relative Data

Items (1)	pre-SAT Cont. (2)	pre-SAT Exp. (3)	post-SAT Cont. (4)	post-SAT Exp. (5)	Adjusted post-SAT Cont. (6)	Adjusted post-SAT Exp. (7)	t^a (8)	Is Difference between Columns 6 and 7 in Hypothesized Direction? (9)
A-F	3.34	3.30	3.18	3.07	3.17	3.08	1.68*	Yes
G-L	3.17	3.15	2.95	2.91	2.94	2.92	.30	Yes
A-L	6.52	6.45	6.14	6.00	6.12	6.02	1.21	Yes

[a] t was computed as $\sqrt{F}$ and its probability was interpreted on a one-tail basis with $df = 173$.

*Significant at the .05 level.

Table 29

Means of Teachers' Adjusted post-SAT Tau Scores by Experimental Interval

Items	Mean Adjusted post-SAT Tau Score C	E_1	E_2	E_3	F	Rho[a]	Is Rho in Hypothesized Direction?
A-F	3.17	3.14	3.03	3.06	1.50	.5	Yes
G-L	2.94	2.94	2.93	2.84	.40	1.0	Yes
A-L	6.12	6.13	5.96	5.94	1.13	1.0	Yes

[a]Rho is the rank-order correlation between the obtained values for E_1, E_2, and E_3, respectively, and the hypothesis that the values would rank $E_1 > E_2 > E_3$.

Summary and Implications

In everyday teaching, the teacher gets feedback by glancing at her class, by giving tests, and by talking informally with her pupils, other teachers, parents, people in the community, and her principal. Obviously, for most teachers most of the time, these sources of feedback serve fairly well. It is reasonable to suppose, however, that there is room for improvement. One possibility suggests itself: manipulate environmental conditions so as to improve teachers' accuracy in perceiving pupils' perceptions of their teacher. Experimentally introduced feedback from pupils to teachers should materially affect classroom processes.

Our experiment was aimed at the question, Can teacher behavior be changed by informing the teachers how their pupils describe the behavior of their actual teacher and their ideal teacher? The pupils in our experiment indicated how well certain behaviors characterized their actual teachers. The pupils also indicated how well the same behaviors would characterize their ideal teacher. Some of the teachers (the experimental group) were given information concerning their pupils' opinions; the remaining teachers (the control group) were not given this information. A month or two later, all teachers were again described by their pupils as to how well the behaviors characterized the teachers. Briefly stated, our major hypothesis was that the experimental group of teachers would change its behaviors (as described by pupils) more than the control group.

When we assume that the teacher has a positive orientation towards her pupils, and we tell the teachers what the pupils' orientations toward the teacher's behaviors (X) are, we set up what Newcomb labeled "strain toward symmetry" on the part of the teacher to make the teacher develop the same orientation toward X. Hence, the teacher can attempt to achieve symmetry with

his pupils by influencing them towards his own orientation. If he thinks that a certain behavior is very much like himself, while he is informed that his pupils do not consider it so, but they would like it to be so (i.e., they say it would characterize their ideal teacher), a strain toward symmetry will lead the teacher to communicative acts intended to make the pupils also consider the behavior very much like himself. In the classroom, these communicative acts will probably take the form of increased frequency or conspicuousness of the behaviors in question. (Our hypotheses can also be formulated in terms of Heider's theory of balance, Festinger's theory of dissonance, and the Osgood-Tannenbaum principle of congruity.) Our generic term for all these theories is equilibrium theory.

After an interval of time, these changes in behaviors will influence the pupils to consider the behaviors more like the teacher. In short, giving the teacher information as to the pupils' orientations toward the behaviors should influence the teacher's behavior and subsequent descriptions by the pupils of their teacher's behavior.

The subjects of the experiment were 176 sixth-grade teachers in Illinois, one from each of 176 school districts, and their approximately 3900 pupils. They volunteered and cooperated in response to a series of mailings.

Protocols consisted of responses to a set of 12 "stimuli," or brief verbal descriptions of teacher behavior. Illustrative of these are

A. Enjoys a funny remark made by a pupil.
B. Praises what a pupil says in class discussion.

A "Report on Your Pupils' Opinions" was made so that individual information could be sent to each teacher. Twelve charts, one for each item, were bound in a booklet. The chart for each item had two parts: (a) a histogram showing how many of her pupils chose "Very much like my teacher," "Somewhat like my teacher," etc.; (b) a histogram showing how many of her pupils chose "Very much like a 'best' teacher," "Somewhat like a 'best' teacher," etc. Also, on each chart, an arrow pointed to the median answer.

The four protocols obtained from pupils were

pre-ACT -- the pupil's description of his actual teacher on the pretest

post-ACT -- the pupil's description of his actual teacher on the posttest

pre-IDL -- the pupil's description of his ideal teacher on the pretest

post-IDL -- the pupil's description of his ideal teacher on the posttest

Did the experimental and control groups of teachers differ in the post-ACT descriptions of them by their pupils? Using analysis of covariance to take account of initial (pre-ACT) differences, we found the differences in adjusted post-ACT were statistically significant at the .05 level for four of the items. The direction of the difference was the same for 10 of the 12 items.

Did the differences go in the direction of the influence exerted by the feedback? The feedback given the teachers in the experimental group concerning how their pupils rated their ideal teacher would presumably exert some influence on the teachers to change in that direction. Our hypothesis was that the difference between adjusted post-ACT and pre-IDL would be smaller for the experimental group. The differences, for 10 of the 12 items, were indeed smaller for the experimental group.

Was there a relation between change due to feedback and the interval between feedback and the postratings of the teachers? The interval was the number of calendar days intervening between the mailing of the feedback information to the teachers and the date on which the teacher collected her pupils' postratings of herself and the ideal teacher. The range of intervals was from 29 to 59 days--from one to two months. The median interval for Group E_1 was about 34 calendar days; for E_2, about 42 days; and for E_3, about 53 days. The 15 teachers in Group E_3 approached their pupils' pre-IDLs most cl[illegible] Group E_1 had the highest average difference between the adjusted p[illegible] and pre-IDL means.

Four protocols were obtained from teachers:

pre-PERC -- the teacher's perception (estimate) of how a pupil who is typical of the majority of the class would answer the item on the pretest

post-PERC -- the teacher's perception (estimate) of how a pupil who is typical of the majority of the class would answer the item on the posttest

pre-SELF -- the teacher's indication of how much the item was "like" herself on the pretest

post-SELF -- the teacher's indication of how much the item was "like" herself on the posttest

For five of the 12 items, the differences in adjusted post-PERC means of the experimental and control groups were statistically significant. The direction of the differences, whether significant or non-significant, was not consistent; we cannot say at this point what they mean. Suffice it now to say that something, presumably the feedback, made the teachers in the experimental and control groups "perceive" their pupils' descriptions differently at the time of the posttest, on five of the items. In any case, adjusted post-SELF differences significant at the .10 and .05 levels occurred for three items, but the direction of these differences was not consistent from one item to the next.

Did the feedback upset the relationship between pre- and post-PERC that would normally (i.e., without feedback) prevail? When the experimental group was broken into the three subgroups by interval, we found five rs smaller in E_1 than in the control group, seven in E_2, and nine in E_3. The trend was in the hypothesized direction: the longer the interval, the lower the r between pre- and post-PERC.

Did the feedback reduce the stability of teachers' views of themselves, i.e., the rs between pre- and post-SELF? No consistent differences in the size of the rs appeared.

A form of what has been called "assumed similarity," averaged over teachers, is the correlation between the teachers' SELF and PERC on a given

item. All 12 of the E_3 changes in assumed similarity were less positive or more negative than those in E_1; nine of them bore a similar relation to the changes in the control group. Apparently feedback, combined with a longer interval before the posttest, tended to reduce the amount of assumed similarity that teachers manifest.

The relations between pupils' and teachers' protocols can take the following forms:

ACC -- the ACCuracy of the teacher's perception or estimation (PERC) of the typical pupil's response, as measured by its closeness to the pupils' mean description of their actual teacher (ACT)

SIM -- the SIMilarity between the teacher's description of herself (SELF) and the pupils' mean description of their actual teacher (ACT)

One measure of ACC, averaged over teachers, is the difference between the mean adjusted post-ACT of the pupils and their teachers' adjusted post-PERC. It seems reasonable to expect the experimental group to be more accurate. This expectation was borne out for nine of the 12 items. Nine out of 12 of the adjusted post-ACC means in E_1 are better (smaller) than those of the control group; 12 out of 12 in E_2; but only seven out of 12 in E_3. The E_2 group was more accurate on the posttest than either E_1 or E_3.

Accuracy can also be measured by the correlation between the teacher's PERC and her pupils' ACT on a given item. Improvements in such correlational accuracy were slightly better in the experimental group and the E_2 group again gained most consistently.

Similarity of teachers' self-descriptions to their pupils' descriptions of them can be measured by the difference on each item between the adjusted post-SELF and the adjusted post-ACT means. The experimental group manifested greater similarity in this sense, presumably as a result of the feedback, on 11 of the 12 items. The increase in similarity appeared to about the same degree for all the interval subgroups. Change in similarity was also measured by change in correlation between SELF and ACT from pretest to posttest.

The experimental group of teachers gained more in such correlational similarity to its pupils on only seven of the 12 items. But these gains were so much larger than the losses on the other five items that the difference between the change-SIM values of the experimental and control groups is significant.

All protocols were also collected in relative form. In this form, the items were presented in triads. The pupil or teacher chose one item of the triad as "MOST," and another as "LEAST"--"like you" (for teachers' SELF); "like what your typical pupil would say about you" (for teachers' PERC); "like my teacher" (for pupils' ACT); "like a best teacher" (for pupils' IDL). To construct the triads, the 12 items of teacher behavior were divided into two groups of six--Items A-F and Items G-L.

To score the relative protocols, each subject's responses to the appropriate triads were first converted into a rank ordering of the six component items. For example, each pupil's responses to the triads of Items A-F were converted to a rank ordering of those items as to how well they characterized his actual teacher and also his ideal teacher. We used the term satisfaction (SAT) to refer to the tau (Kendall's coefficient of rank correlation) between these two rank orderings. Several such correlational scores were obtained between the ranks of six items on different protocols.

The reliabilities of the pre-SAT and post-SAT scores ranged from .30 to .61 for the control and E_T groups. The reliabilities of the change-ACT, change-SELF and change-PERC scores were essentially zero; accordingly, further analyses of these latter scores have not been reported.

Our hypothesis was that teachers in the experimental group would elicit higher post-SAT scores from their pupils. The experimental group's behaviors were indeed judged by pupils, at the time of the posttest, to be more in the rank order characteristic of the pupils' ideal teacher, after adjustment by analysis of covariance for differences in pre-SAT. But only the difference for Items A-F was significant. With only one reversal, the mean tau scores for adjusted post-SAT were found to become smaller--and satisfaction hence

greater--in going from E_1 (shortest interval) to E_3 (longest interval); the F-ratios in these analyses of covariance were not, however, statistically significant.

All in all, our results have the following theoretical and practical implications: Equilibrium theory is supported, or at least not refuted, by the changes in teachers' behaviors, as described by pupils, resulting from the feedback of pupils' opinions concerning the behavior of their actual and ideal teachers. The feedback not only produced change in behavior; it also produced corresponding changes in the accuracy of teachers' perceptions of their pupils' perceptions of their teacher, and in the similarity of teachers' self-descriptions to their pupils' descriptions of the teacher. Whether rival theories of some kind could explain these results as well or better will remain unknown until the attempt is made. For the present, it appears that equilibrium theory survives the test to which it was subjected in the present experiment.

For practical purposes of improving teacher behavior, the method of feedback to teachers of pupils' ratings also seems to possess new promise in the light of our results. Teachers did change, in the direction of pupils' ideals, as described by pupils, as a result of getting feedback. Whether the changes were great enough to have educational significance, whether they would be found if teachers' behaviors were described and measured by expert outside observers rather than pupils, whether the changes toward pupils' ideals are also toward educators' ideals--all these are questions for subsequent investigation. A host of additional issues would arise in the development of such feedback into a usable scheme for teacher improvement. Such development and refinement of the feedback-of-pupils'-ratings technique can now proceed, with all the advantages of recent inventions for rapid data processing, on the basis of some evidence that, in one test of the possibility, at least, teachers' behaviors did change as hypothesized.

Appendix A
The Chronology of Data Collection

1. On September 5, 1956, letters (M1)[5] were sent to all superintendents in Illinois, except those in Chicago, having sixth grades in their charge. Each of the 587 superintendents was invited to give us the name of the sixth-grade teacher coming first in his alphabetical list of sixth-grade teachers, along with her home mailing address. We wanted only one teacher from each community in order to make it difficult for our subjects to compare notes with each other. Each superintendent received a postal card on which to send us the information requested. This mailing is shown as the first entry in Table A-1, which is to be read as follows: On September 5, 587 letters (M1) enclosing post cards (M2) were sent out. By September 15, 373 replies had been received. On this date, a reminder (M3) was sent to the superintendent. By September 25, 477 replies had been received, and by October 5, 497. This was the total response.

2. By September 25, our records of return showed that the rate of return was leveling off. We therefore prepared the next mailing--the booklets entitled "What Do They Expect?" (M5) (excerpts of which are shown in Appendix D)--and on September 28 mailed them to all 475 teachers whose names we had at that time. Table A-1 shows 477 replies from superintendents by September 25, but only 475 "What Do They Expect" (WDTE) booklets mailed on September 28. The explanation for this is that two superintendents returned post cards only to say that they did not wish to participate in the research or sent us illegible post cards or the like. By October 16, 489 WDTEs had been sent out. The difference between this figure and the figure of 497 replies from superintendents again indicates a few replies unusable for one reason or another. This kind of attrition as well as that resulting from nonresponse took place at each stage of the mailings and returns.

[5] The "M" in designations of instruments stands for "mailing piece."

Table A-1

Calendar of Mailings and Receipts

Instrument	Description	Sent out		Received	
		Date	Cumulative Number	Date	Cumulative Number
M1 & 2	Invitation to superintendent	Sept. 5	587	Sept. 15	373
				Sept. 25	477
				Oct. 5	497
M3	Reminder to superintendent	Sept. 15	218		
M4 & 5	What Do They Expect	Sept. 28	475	Oct. 15	178
		Oct. 16	489	Oct. 17	218
				Oct. 24	285
				Nov. 28	360
M6	Post-WDTE reminder	Oct. 15	124		
		Oct. 16	200		
		Oct. 17	284		
M7, 8, & 9	Pupil Opinion Booklets	Oct. 8	8	Oct. 20	77
		Oct. 15	133	Oct. 24	112
		Oct. 27	265	Oct. 25	151
				Oct. 31	210
				Nov. 19	250
M10	Post-POB reminder	Oct. 15	1		
		Oct. 20	48		
		Oct. 26	86		
M12 & 13	Report on Your Pupils' Opinions (Experimental group)	Oct. 24	3		
		Nov. 13	127		
M11	Notification of delay to control group	Nov. 1	1		
		Nov. 14	124		
M14 & 15	RYPO follow-up questionnaire (Experimental group)	Oct. 31	3	Nov. 8	3
		Nov. 21	127	Nov. 21	53
				Nov. 30	88
				Dec. 19	119
M14 & 15	Follow-up mailing of RYPO follow-up questionnaire (Experimental group)	Nov. 21	12		
		Nov. 30	56		
M5 & 9	Post-test WDTE and POB (both groups)	Dec. 3	217	Dec. 6	2
		Dec. 15	249	Jan. 17	223
M12 & 13	RYPO (Control group)	Dec. 11	4		
		Jan. 9	121		
M14 & 15	RYPO follow-up questionnaire (Control group)	Dec. 19	10	Dec. 21	1
		Jan. 15	120	Jan. 15	67

The WDTE (M5) booklet contained a description of what we were asking of the teacher and what service we offered her, and also pages bearing questionnaires for the teachers to fill out. At the end of the booklet, the teacher was asked to tell us how many questionnaires she needed for her pupils. By October 15, 178 WDTEs had been received. Between that date and October 17, 284 follow-up reminders (M6) were sent out. By October 24, 285 WDTEs had been received. This date was used as a cut-off point to maintain our schedule and allow a reasonable length of time for the treatment--that of feeding information back to the teachers of the experimental group concerning how their pupils had described them. Although 360 WDTEs were received in all, those received after October 24 were put aside, and letters of regret were sent to the teachers saying that we would not be sending Pupil Opinion Booklets and that we were sorry not to be able to process their data. (This letter of regret, M21, is not shown in Table A-1.)

As WDTEs were received, the experimental and control groups were established. As each batch of mail was received and before any inspection of the booklets in it, each booklet was alternately marked "Experimental" or "Control."

3. In the meantime, Pupil Opinion Booklets (POBs) were packaged according to the numbers requested by the teachers. Eight packages of POBs (M7, 8, and 9) were mailed on October 8. By October 27, 265 had been mailed. By this time, attrition was at a lower rate; 250 packages of POBs were returned by November 19. Reminders (M10) sent out during October 15-26 presumably helped.

4. The difficult task was now under way of preparing reports to the teachers of the experimental group telling them what their pupils said about their actions. A large staff of computers was trained to prepare booklets (M13) entitled "Report on Your Pupils' Opinions" (RYPO)--an excerpt of which is shown in Appendix E--as rapidly as the POBs were returned. The first three reports went out October 24 and by November 13, RYPOs had been sent to the 127 members of the experimental group as it was then constituted.

Beginning on November 1, teachers in the control group were sent letters (M11) saying that reports to them on what their pupils had answered would be delayed for some weeks. By November 14, the 124 teachers who had been designated members of the control group had been sent these letters.

A follow-up questionnaire--prepared partly to gather additional data but also to encourage the 127 teachers in the experimental group to carefully read their RYPOs--was sent between October 31 and November 21. This questionnaire (M15) asked whether the teacher had read the RYPO booklet, whether she found it interesting, and whether she felt we had provided information she had not had before. A follow-up mailing of 56 replacement copies of this questionnaire took place between November 21 and November 30. In all, 119 were returned from the experimental group by December 19.

5. The WDTEs and POBs required for the posttest were sent to teachers in both the experimental and control groups between December 3 and December 15. In all, 249 packages were sent. Between December 6 and January 17, 223 were returned. Here we experienced some delay and found it necessary to undertake considerable correspondence and detective work because a serious number of our packages were broken in transit through the Chicago post office during the Christmas rush. Most of these, however, were identified and duplicate packages were sent to the teachers involved.

6. Between December 11 and January 9, as the control group's post-test materials came in, 121 RYPO booklets were sent to teachers in the control group. This was followed by 120 copies of the follow-up questionnaire (M15) to the control group; 67 were returned by January 15.

7. In the spring of 1957 a report was prepared for participating teachers giving summaries of pupil responses and teacher responses throughout the state. This report also explained what we hoped to learn from our research. It was sent to every teacher who had at any time participated in the project.

8. At the end of the data collection, 208 teachers had provided us with

106 in the control group. But not all these teachers were used in the analysis of the data. Some teachers had special situations in their schools which made their answers not comparable to those of teachers in other schools: (a) One item, for example, read "Explains arithmetic so pupils can understand it." Some teachers told us that they did not teach arithmetic. (b) Another item read, "Talks with a pupil after school about an idea the pupil has had." Some teachers explained to us that school buses carried away all their pupils immediately after school so that they had no opportunity to talk with pupils after school. (c) Further, some teachers told us that the sixth grade was departmentalized in their schools so that they taught only certain subjects to their pupils. This condition, too, would affect answers to such items as "Explains arithmetic so pupils can understand it." It would also reduce the amount of time pupils of such teachers would be exposed to their influence. Accordingly, the free remarks written in by teachers on instruments they had sent us were scrutinized for evidences of the foregoing unusual conditions. Any teacher who gave such evidence was removed from the sample along with her pupils.

Teachers removed from the sample to be analyzed were not, however, removed from subsequent stages of the experimental procedure. In our introductory material (WDTE), we had promised a service to the teachers--that of providing them with information about their pupils' perceptions of their actions--and this promise was rigorously kept. RYPOs were sent to all teachers who furnished responses from their pupils with which to obtain the information to be put in the report booklet. In addition, all inquiries from teachers were answered. (Eventually, the correspondence file got to be about five inches thick.) In answering questions from teachers, however, we tried not to bias the outcome of the research. While many questions could be answered fully and directly, we merely suggested that the teacher would find her question answered in a future report and that she write again if she did not succeed in doing so.

9. After the list of teachers had been reduced for the reasons described above, the samples used in the final analysis contained 86 teachers and classes in the experimental group and 90 in the control group.

Appendix B

The Attribute Interview Study

The "dimensions" obtained in this study are represented by the groups of items shown in Tables B-1 and B-2. A few examples will illustrate how these groups suggested items finally selected. One item in one of our pre-final sets of items was "Asks a pupil to explain more fully if she doesn't understand what the pupil means to say." Another was, "Explains arithmetic so pupils can understand it." These can be seen to represent the converse of the actions in the fourth group shown in Table B-1. One item we considered was "Tries to help a pupil feel better if he is sad or worried." This obviously fits with the fourth group shown in Table B-2. The item, "Asks a small group of pupils to study something together," has bearing on the first g[illegible] Table B-2. These examples illustrate after the fact our reason for bei[illegible]ested in the cognitive framework through which the teacher perceive[illegible] pupils.

The Attribute Interview Study was exploratory in n[illegible]re and fairly unstructured methods were used. Eight elementary school teacher[illegible]e interviewed. First, each teacher was asked general questions about [illegible]uld come to her mind in making a rough description of her class, abou[illegible]rences and similarities between this class and others she had ta[illegible]e was also asked what she felt it was important to keep in mind when t[illegible]er class, what kinds of things she watched for among her pupils. She [illegible]n asked to pick every fourth pupil from her class roster and to answe[illegible]llowing questions about each one: "How would you describe this pupil if [illegible]e telling about him to some friend you knew well who would understand y[illegible] "Will you pick out the names of two or three pupils who are alike in [illegible]mportant way. How are these two alike? Which pupils are least like these[illegible]" The teacher was finally asked how she would expect the following kin[illegible]pils to differ: and girls, children from poor or well-to-do homes, [illegible]s scoring high and

Table B-1

Approximate Equivalence Groups among 38 Selected Items
from the "Attribute Interview Study"
(Among which 6 Items were Common to No One Group)

First Group	Second Group	Third Group	Fourth Group
410. traveled	286. obese	208. interested in athletics and games	320. quick to respond, when they disagree with you
94. daughter of a doctor	48. attractive	256. likes athletics	304. would be impolite if necessary before they'd let it pass over their heads
376. from South	52. not beauty queens, but physical neatness		
192. home life is as high financially and scholastically as many	74. come to school clean		
120. don't eat well, from poor homes, haven't slept well	278. neat		
122. comes from poor economic conditions	400. tall		
434. because of his background doesn't do so well in work	372. small for age		
380. can't do English when they don't speak well	308. pretty		
416. rich children make problem in a room because they don't understand pupils who have less	20. active physically		
288. had operation in September	102. physical development		
414. a twin	42. good athlete		
116. dresses so much neater than some of other members of her family	268. manners		
144. have greater range of experience	328. a real boy		
378. speak better, have traveled more, have more to contribute in English and reading, do better on I.Q.			
188. hearing has improved			

Table B-2

Equivalence Groups among 23 Selected Items
from the "Attribute Interview Study"

First Group	Second Group	Third Group	Fourth Group
62. careful not to hurt each other's feelings	104. don't have too much difficulty	4. ability to carry on independent projects that are varied	128. emotional problems
82. concerned about each other	150. try to figure things out for themselves	10. ability to use abstractions	148. fearful
86. considerate of others	210. interested in learning	56. bright and able	222. isolates
88. shows proper consideration for his classmates	212. interested in schoolwork	66. easily challenged	276. moody
174. gets along very well socially	238. not hostile to learning	340. responsible and intelligent	432. is withdrawn
250. well liked by boys and girls because of fairness and considerateness	326. do much outside reading		
252. well liked by children and they do try to help him			

low on an intelligence test, and pupils she found more satisfying or less satisfying to have in class.

The protocols were then typed off without any identification, and three judges independently went through the transcripts marking what they considered to be the smallest units of meaning. If two judges agreed in identifying a unit of meaning, this unit was typed on a small card. One judge then looked through the cards for duplicates, and after these were removed, 450 cards remained. Half of these were randomly selected for further analysis. These 225 cards were given to each of five judges. Each judge was instructed to go through the cards placing cards in the same pile if they were similar in meaning. The judges were specifically instructed that the cards should not be separated on the basis of whether they indicated positive and negative aspects of the same attribute, but should be put in the same pile if they reflected the same criterion or basis regardless of whether they mentioned explicitly the positive or negative aspects of the criterion. Thus, "enthusiastic" and "apathetic" might go in the same pile, but "enthusiastic" and "intelligent" would probably go in different piles. The judges were told that they might have as many or as few piles as they wished. After the judges had done their work the results of their sorting were recorded in a 225 x 225 matrix.

Two analyses were then performed, yielding the results shown in Tables B-1 and B-2. In the first analysis, a selection of items was made to obtain a subset of items having the greatest agreement among judges. This was done by assigning weights to each cell of the 225 x 225 matrix as follows: 0, if either no judge or all five judges had agreed in placing two items in the same pile; 1, if either one judge or four judges agreed; 2, if either two or three judges agreed. By adding the weights in all cells corresponding to a given item, a total weight for each item was obtained and the 38 items with the highest weights were selected for further analysis. A new 38 x 38 matrix was then constructed; a "1" was entered in a cell if the two items had been put in

the same piles by three or more judges, and a "0" if fewer than three judges. This 38 x 38 matrix was then manipulated by Festinger's method (1949). This procedure yielded the four groups of items shown in Table B-1.

In the second analysis, the 225 x 225 matrix was inspected to find pairs of items placed by all five judges in the same pile. There were 23 items for which at least four such cells appeared. These items were selected for a new 23 x 23 matrix. In this matrix a "1" was entered in a cell if all five judges had agreed, and a "0" if fewer than five had agreed. This matrix, after manipulation by the Festinger method, yielded the four groups shown in Table B-2.

Appendix C
The Discriminability Study

In Stage 1, four overlapping replications were made of Plan 11.20 from Cochran and Cox (1950, p. 333), each replication presenting six items to each of the eleven subjects. Forty-four questionnaires were prepared, each with a different set of six items, and presented to sixth-grade teachers in Mattoon and Lincoln, Illinois. Each set of six items was grouped into 20 triads--all possible combinations of six items, three at a time. The subjects were asked to indicate which one of the three items was most characteristic of them, and which one was least characteristic.

Since pairs of items appeared in more than one triad, a teacher could say one time that Item A was more characteristic of her than B, and at another time the reverse. For every pair of items, the number of such inconsistent responses could be recorded for each teacher. Finally, the mean consistency was computed for each pair of items over all teachers.

In the next step, a matrix was made with a "1" in the cell corresponding to a pair of items with 100 percent consistency, and a "0" otherwise. The matrix was then manipulated by the Festinger method to find groups of items mutually discriminable from each other. This analysis provided a beginning indication of which items could best be used together in groups.

The questionnaires sent to the teachers also contained the six items presented to the teacher in a list with each item followed by six alternatives among which the teacher could choose to tell us how much the item characterized her. In half the questionnaires, the six responses offered were "Very much like me," "A little like me," etc. In the other half, the choices were "I try to do this very frequently," "I sometimes do this," etc. For both kinds of choices, the mean deviation from the median over subjects was computed to determine which kind of choice-labeling yielded greater stability.

Table C-1

List of Items

A. Tells pupils when they are doing good work.
B. Lets pupils tell about things they think the class would like to hear about.
* C. Enjoys a funny remark made by a pupil.
* D. Praises what a pupil says in class discussion.
* E. Asks a small group of pupils to study something together.
* F. Tells pupils about some interesting things to read.
G. Gives more help to pupils who have a hard time with their schoolwork.
H. Asks a pupil to explain more fully if she doesn't understand what the pupil meant to say.
* I. Shows a pupil how to look up an answer when the pupil can't find it by himself.
* J. Explains arithmetic so pupils can understand it.
* K. Asks the pupils what they'd like to study in tomorrow's lesson.
L. Explains why she wants us to study a certain lesson.
M. Asks pupils to give their opinions about a movie.
* N. Suggests to pupils new and helpful ways of studying.
* O. Acts disappointed when a pupil gets something wrong.
P. Likes pupils who work hard.
Q. Tries to help a pupil feel better if he is sad or worried.
R. Asks pupils who don't say much if they'd like to say something.
* S. Explains something by using examples from games and sports.
T. Shows pupils some ways to avoid mistakes in spelling.
* U. Asks the class what they think of something a pupil has said.
* V. Talks with a pupil after school about an idea the pupil has had.

---

*Items finally used.

Although a number of other kinds of analyses were performed in Stage 1 of the Discriminability Study, the foregoing illustrates the manner in which it was performed.

Stage 2 was similar to Stage 1 except that only 16 items of the original 22 were used. These 16 items were made into 18 questionnaires, each containing six items. Copies of these questionnaires were presented to 48 teachers in Mattoon, Streator, and Bloomington, Illinois.

In Stage 3, we turned to pupils. The same 16 items were used as in Stage 2 and were distributed over different questionnaires according to Plan 11.28 of Cochran and Cox (1950, p. 335). The superintendent of schools in Mahomet, Illinois kindly granted permission for us to present these questionnaires to sixth-grade pupils in his system. Analysis was similar to that of Stages 1 and 2.

By the time Stage 4 was begun, we had selected 12 items for final use out of the original 22. The purpose of Stage 4 was to see whether the 12 items would perform properly when the subject was asked whether the item characterized the behavior of his ideal (not actual) teacher. In this stage two sets of six items each were given to 52 sixth-grade pupils in Arcola, Illinois, and were analyzed as in previous stages.

The final 12 items were chosen on the basis of a number of criteria reflected in measures developed in the Discriminability Study. Some of these criteria have already been mentioned. Another was the relation between the self-characterizations of teachers when answering the two kinds of question forms we offered them. In one part of the questionnaire we asked them which two items were most and least like themselves among three presented at a time. From all the triads a rank order of items was obtained as to the degree to which the teacher said they characterized herself. Another part asked the teacher to tell us how much each item was "like" her by choosing one of six answers. The answers in this part of the questionnaire also yielded a rank order of items based on the particular answer chosen by the teacher for

each item. When the rank correlations (rho) between the responses in these two parts of the questionnaire were computed for each teacher, the mean of the correlations in Stage 2 was .85. When the mean rank was computed for each item over all teachers and the mean taken over items the mean rank correlation was .82. Some further characteristics of the items are shown in Table C-2. Although the final 12 items did not have the highest possible indices, they were far better as a group than the other items and were considered satisfactory for final use.

Table C-2

Characteristics of Items in the Discriminability Study

Statistic	Stage 1	2	3	4 A-F	4 G-L
Number of subjects	38	37	41	49	49
Number of items	22	16	16	6	6
Range of mean percent consistency within pairs of items	50-100	67-100	58-100	83-97	80-94
Mean percent consistency among subjects	74	76	59	76	77
Percent of subjects having percent consistency of 80 or above	62	58	26	49	61
Correlation (r) between (1) percent consistency and (2) rho between relative and irrelative rankings of 6 items	.35	.33	.46	.83	.54
p of r	.05	.05	.01	.01	.01

Appendix D

Excerpt from "What Do They Expect?" Booklet

WHAT DO THEY EXPECT?

some conversation about pupils,
an invitation, and a request

by N. L. Gage, Ph.D., and P. J. Runkel, Ph.D.
College of Education, University of Illinois

Teachers Information Service
Bureau of Educational Research
University of Illinois
Produced in part under Research Grant M-650,
National Institute of Health, U. S. Public Health Service

By *they* we mean your *pupils.*

What do they expect of you?

Naturally, a pupil who gets zero on a test
doesn't expect you to be exactly happy about it.

And the chances are that a pupil who turned up
with a beautiful model airplane
—with working parts, no less—
would expect something more than a
"That's very nice."

"Quinine is the berk of a tree;
canine is the bark of a dog."

And naturally, there are many ways
in which they try to anticipate
what you want and what you will do.

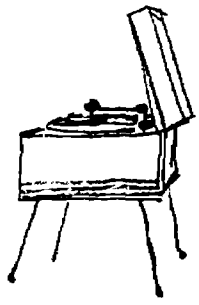
Sometimes they anticipate correctly
and sometimes incorrectly.

All of us sometimes wonder what in the world
the pupils thought we were talking about . . .
and we wonder particularly
when we get answers like these:

"New York is behind Greenwich time
because America was not discovered
until very much later."

"Quinine is the bark of a tree;
canine is the bark of a dog."

"Gravity was discovered by Izaak Walton.
It is chiefly noticeable in the Autumn,
when the apples are falling off the trees."



But there are less funny cases
where a pupil and teacher
have different expectations:

There was the time when a teacher of a sixth grade
was giving her pupils practice in public speaking
by letting them stand up and tell jokes.
At last a very reticent boy,
who lived in a rather depressed neighborhood,
got up his courage
and signalled that he wanted to tell a joke, too.
The only point of the joke
he proudly told
was that it ended in a four-letter word.
Needless to say,
neither the class nor the teacher
burst into delighted laughter.
And the boy, not understanding the shocked silence,
was hurt and bewildered
by the result of his effort to join in the fun.

"Philip, sit up and listen to the music."

And one remembers, too,
a particular music appreciation period
in the sixth grade.
The phonograph was bursting its seams
with the galloping finale
of the *William Tell Overture.*
One boy put his head down upon his arms
and closed his eyes,
so as to be alone with the compelling excitement
of the music.
But the teacher saw what he had done.
She whipped down the aisle to his desk.
She shook the boy sharply by the shoulder.
"Philip!" she said sternly,
"Sit up and listen to the music!"

In each of these examples,
the orientation of the pupil
was not what the teacher thought it was.
The mental direction
in which the pupil was looking, so to speak,
was not the direction
in which the teacher was looking.

What the pupil expected to happen next
was not what came to the mind of the teacher.

Teachers get a lot of information
about how the pupils are following along
as far as subject-matter is concerned.

Like when you ask "How much is 7/8 plus 5/16?"
And he answers "19/16."
You're both talking the same language.

But how the pupil sees *you*
in the midst of all these busy classroom matters,
is information which is not so easy to get.

Suppose you try it.
Suppose you ask a pupil, "How do you think I act?"

Poor child.

In the first place, he probably wouldn't know
what you were talking about.
Even if he had a glimmering,
he'd probably give you some answer
you couldn't use anyway.
Such as, "You're O.K., I guess.
You act O.K."
Which you are already trying to do
with every nerve and muscle you've got.

In the second place,
you every now and then run across a pupil
who doesn't want to talk.
He distrusts teachers.
He might think you were about to hit him,
or about to kiss him,
but he'd rather you'd do either
than find out what he thought about you.
We're glad not many pupils are like that,
but there's no denying
that every school has some of them.

"Every now and then you run across a pupil who doesn't want to talk . . ."

"The teachers make the best guesses they can."

Problems like this
make it hard to find out
whether your pupils are seeing you
as you see yourself—
whether they notice the same kinds of things
about you
that you tend to be concerned about.

And of course, there's the problem of *time,*
which pops up in connection with almost everything
which a teacher tries to do,
these days.

If you could sit down with each one of your pupils
for an hour each,
you could perhaps find out,
at least from most of them,
how they felt about school and about you.

But we all know how difficult it is
to find time for this kind of thing.
Much as we would like to do it.

The result of these difficulties
is that teachers make the best guesses they can.

When she tries to keep in touch
with what is going on
in the minds of 20 or 30 bouncing young creatures,
any teacher has moments
when she feels like the knight
who leaped on his horse
and rode off in all directions.

Most of us would give a lot
to know the many different ways
in which our pupils see
(or understand, or find meaning in)
what *we* are doing
as teachers in the classroom.

We'd like to get this information more reliably
than by the catch-as-catch-can method.

"We've been trying to figure out a way of getting hold of these elusive turns of mind we have just been talking about."

NOW COMES THE PITCH YOU'VE BEEN WAITING FOR.

We've been trying to figure out a way
of getting hold of the elusive turns of mind
we have just been talking about.

You might think, offhand,
that it would take a six-hour interview,
or an electroencephalograph,
to get at this kind of thing reliably.

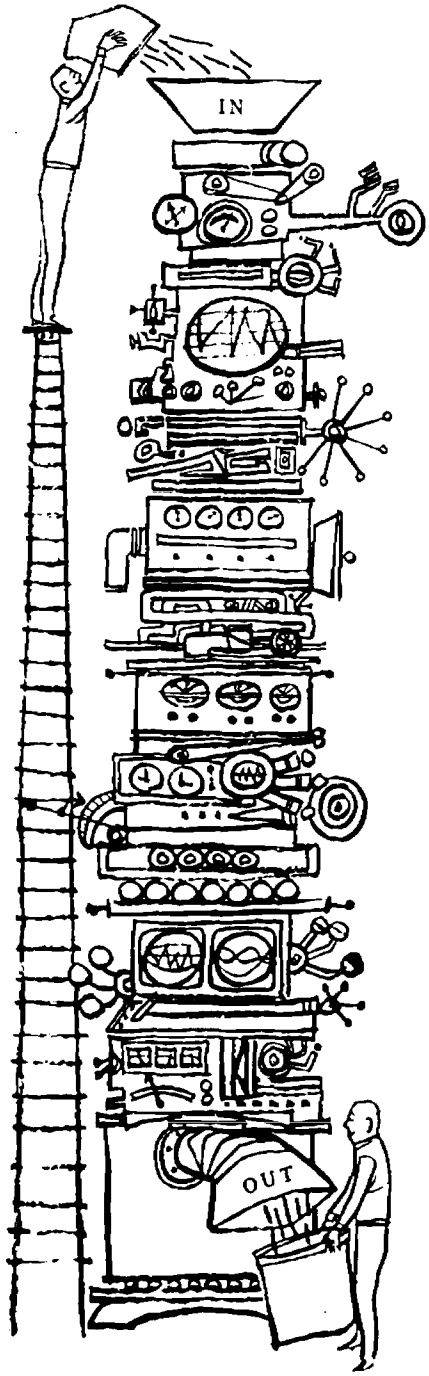
But it won't.
What it requires is
that you answer the questions
on the next few pages,
and then pass out similar questionnaires
to your pupils.

A great amount of potential information
is packed into the few pages
of this questionnaire.
We have been able to achieve this condensation
by using a very special form of questionnaire,
which may at first seem strange to you.
This unusual form enables us, however,
to reduce to a minimum
the time and effort required of you
and of your pupils,
and at the same time
IT WILL ENABLE US TO MAIL BACK TO YOU
A LOT OF INFORMATION ABOUT YOUR PUPILS.

If you are like the average person,
it will take you perhaps thirty or forty minutes
to check off your answers in this booklet.

And once *you* have answered the questions
in *this* booklet,
you will know how to help any of your pupils
who may not understand
the directions for their questionnaires.

You will *not* have to do a lot of counting
or adding
In fact, you won't have to do any.



"You know how fast these electronic computers work."

We will do all the totalling and computing
for you,
and make up an interpretive report
concerning *your particular pupils*
(*not* lumped with those of other teachers).

WE WILL SEND YOU THIS REPORT
AS SOON AS THE ANALYSIS IS COMPLETE.

We will punch all this information on cards,
and run the cards through an electronic computer,
which will handle all these data
in the special way
which this new technique requires.

You know how fast these electronic computers work.

Ours is all set up and ready to go.
We are planning to have the information
about your class
computed and laid out in understandable form
and on its way back to you
at the earliest possible date.

This information will then be yours
to keep and use.

To sum up,
all you need do
is find a few minutes at your convenience
to fill in the remaining pages of this booklet.

Give your pupils a part of an hour
to fill in their questionnaires.

Send the questionnaires to us,
and we will do the rest.

That's all there is to it.

Now we have to admit
there's a small catch to this.

We have been talking about a technique
of analyzing the questionnaires
of a group of people
which comes up with a description
of the viewpoints or orientations
which exist in the group.

"...the information we can send you will be a very worth-while return on the investment of part of an hour's time.."

We're developing this technique
with resources provided
by the University of Illinois
and the U.S. Public Health Service,
both of which care a lot
about teachers and teaching.

This technique has been tested
in a number of situations,
and has turned up useful and reliable information.

But it has not yet been tried out
in many classrooms.

You can see what our concern is in this.

We want to be sure
that this method
of getting and analyzing information
will be useful to teachers in elementary schools
as well as to other types of leaders.

To make sure of this,
we would want you and your pupils
to mark your answers in these booklets
a *second* time,
about six weeks after
you have first filled them out.

This will enable us to check
on the variability shown by pupils
compared to other groups
where this method has been used.

We feel that the information we can send you
will be a very worth-while return
on the investment of part of an hour's time now
and part of an hour's time six weeks from now.

We hope you will want this information
about your pupils
and at the same time
help us make information of this important kind
more widely available to teachers.

Your questionnaire starts on the next page.

PART 1 HOW TO DO IT

This is not a "test."
There are no right or wrong answers.
An answer which tells us what your considered opinion is,
is a "correct" answer.

On the next four pages,
a number of behaviors or actions are listed
in groups of three.

Read over the three behaviors in each group.
Decide which of the three is **most** like you.

You will find the letters "M" and "L" following that behavior.
Please *encircle* the "M".

Then decide which of the three is **least** like you.
Encircle the "L" following this behavior.

For example:

Goes to movies often.	M	L
Likes to travel.	M	(L)
Reads a lot.	(M)	L

Then go on to the next group of three.

(Perhaps all three statements may seem like you,
or perhaps none of them may seem very much like you.
But pick out the one which is **most** like you
and the one which is **least** like you
in each group of three.
Some of these choices may seem hard to make.
But please do the best you can.
Most teachers' answers make good sense to us,
even when the teachers don't think they will.)

Please go along thoughtfully,
but it is not necessary to spend very much time
on any one group.

When you have finished these four pages,
Please check back to make sure
that **you have encircled one "M" and one "L"**
in every group.

Praises what a pupil says in class discussion.	●	L
Explains arithmetic so pupils can understand it.	M	L
Tells pupils about some interesting things to read.	M	●
Acts disappointed when a pupil gets something wrong.	M	L
Asks the class what they think of something a pupil has said.	M	●
Explains something by using examples from games and sports.	●	L
Suggests to pupils new and helpful ways of studying.	M	●
Enjoys a funny remark made by a pupil.	M	L
Praises what a pupil says in class discussion.	●	L
Shows a pupil how to look up an answer when the pupil can't find it by himself.	●	L
Asks the class what they think of something a pupil has said.	M	●
Asks a small group of pupils to study something together.	M	L
Suggests to pupils new and helpful ways of studying.	M	●
Tells pupils about some interesting things to read.	M	L
Praises what a pupil says in class discussion.	●	L

GO ON TO THE NEXT PAGE

Explains something by using examples from games and sports. M L

Asks a small group of pupils to study something together. M L

Shows a pupil how to look up an answer when the pupil can't find it by himself. M L

Tells pupils about some interesting things to read. M L

Enjoys a funny remark made by a pupil. M L

Explains arithmetic so pupils can understand it. M L

Asks a small group of pupils to study something together. M L

Acts disappointed when a pupil gets something wrong. M L

Explains something by using examples from games and sports. M L

Explains arithmetic so pupils can understand it. M L

Suggests to pupils new and helpful ways of studying. M L

Talks with a pupil after school about an idea the pupil has had. M L

Asks the class what they think of something a pupil has said. M L

Asks the pupils what they'd like to study in tomorrow's lesson. M L

Asks a small group of pupils to study something together. M L

GO ON TO THE NEXT PAGE

HOW TO ANSWER THE QUESTIONS IN THE NEXT SECTION:

On the next two pages,
you will find again
some things you have already met.

But this time
they come one at a time.

After each thing
are six different answers.
Pick *one* of these answers
and carefully make an "X"
in the box in front of the answer.

FIRST,

read the sentence
which tells what a teacher might do.

THEN,

pick *one* of the six answers.

Make sure that your "X"
marks just *one*
of the boxes.

LIKE THIS:

Goes to movies ofte:.
- ☐ Very much LIKE me
- ☐ Somewhat LIKE me
- ☐ A little bit LIKE me
- ☐ A little bit UNLIKE me
- ☒ Somewhat UNLIKE me
- ☐ Very much UNLIKE me

Now go ahead.

Suggests to pupils new and helpful ways of studying.

- ☐ Very much LIKE me
- ☒ Somewhat LIKE me
- ☐ A little bit LIKE me
- ☐ A little bit UNLIKE me
- ☐ Somewhat UNLIKE me
- ☐ Very much UNLIKE me

Shows a pupil how to look up an answer when the pupil can't find it by himself.

- ☒ Very much LIKE me
- ☐ Somewhat LIKE me
- ☐ A little bit LIKE me
- ☐ A little bit UNLIKE me
- ☐ Somewhat UNLIKE me
- ☐ Very much UNLIKE me

Enjoys a funny remark made by a pupil.

- ☒ Very much LIKE me
- ☐ Somewhat LIKE me
- ☐ A little bit LIKE me
- ☐ A little bit UNLIKE me
- ☐ Somewhat UNLIKE me
- ☐ Very much UNLIKE me

Acts disappointed when a pupil gets something wrong.

- ☐ Very much LIKE me
- ☐ Somewhat LIKE me
- ☐ A little bit LIKE me
- ☐ A little bit UNLIKE ME
- ☐ Somewhat UNLIKE me
- ☒ Very much UNLIKE me

Asks a small group of pupils to study something together.

- ☐ Very much LIKE me
- ☐ Somewhat LIKE me
- ☐ A little bit LIKE me
- ☐ A little bit UNLIKE me
- ☒ Somewhat UNLIKE me
- ☐ Very much UNLIKE me

Asks the pupils what they'd like to study in tomorrow's lesson.

- ☐ Very much LIKE me
- ☐ Somewhat LIKE me
- ☒ A little bit LIKE me
- ☐ A little bit UNLIKE me
- ☐ Somewhat UNLIKE me
- ☐ Very much UNLIKE me

GO ON TO THE NEXT PAGE

Talks with a pupil after school about an idea the pupil has had.

- ☐ Very much LIKE me
- ☐ Somewhat LIKE me
- ☐ A little bit LIKE me
- ☐ A little bit UNLIKE me
- ☐ Somewhat UNLIKE me
- ☒ Very much UNLIKE me

Praises what a pupil says in class discussion.

- ☒ Very much LIKE me
- ☐ Somewhat LIKE me
- ☐ A little bit LIKE me
- ☐ A little bit UNLIKE me
- ☐ Somewhat UNLIKE me
- ☐ Very much UNLIKE me

Asks the class what they think of something a pupil has said.

- ☐ Very much LIKE me
- ☐ Somewhat LIKE me
- ☐ A little bit LIKE me
- ☐ A little bit UNLIKE me
- ☐ Somewhat UNLIKE me
- ☒ Very much UNLIKE me

Explains something by using examples from games and sports.

- ☐ Very much LIKE me
- ☒ Somewhat LIKE me
- ☐ A little bit LIKE me
- ☐ A little bit UNLIKE me
- ☐ Somewhat UNLIKE me
- ☐ Very much UNLIKE me

Tells pupils about some interesting things to read.

- ☐ Very much LIKE me
- ☒ Somewhat LIKE me
- ☐ A little bit LIKE me
- ☐ A little bit UNLIKE me
- ☐ Somewhat UNLIKE me
- ☐ Very much UNLIKE me

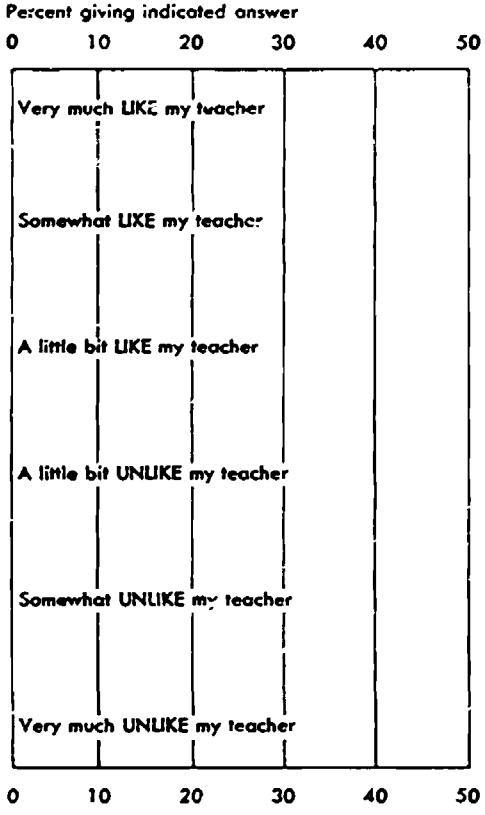
Explains arithmetic so pupils can understand it.

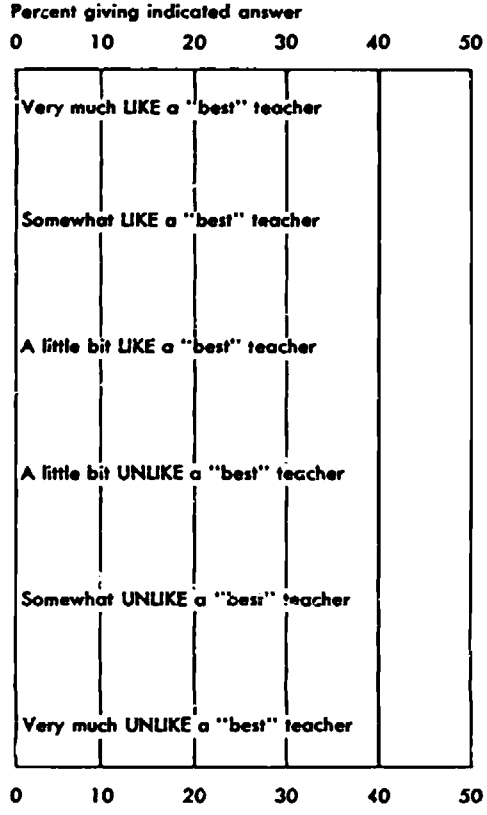
- ☒ Very much LIKE me
- ☐ Somewhat LIKE me
- ☐ A little bit LIKE me
- ☐ A little bit UNLIKE me
- ☐ Somewhat UNLIKE me
- ☐ Very much UNLIKE me

GO ON TO PART 2

Appendix E

Excerpt from "Report on Your Pupils' Opinions" Booklet


Percent giving indicated answer
0 10 20 30 40 50
Very much LIKE my teacher
Somewhat LIKE my teacher
A little bit LIKE my teacher
A little bit UNLIKE my teacher
Somewhat UNLIKE my teacher
Very much UNLIKE my teacher
0 10 20 30 40 50


Percent giving indicated answer
0 10 20 30 40 50
Very much LIKE a "best" teacher
Somewhat LIKE a "best" teacher
A little bit LIKE a "best" teacher
A little bit UNLIKE a "best" teacher
Somewhat UNLIKE a "best" teacher
Very much UNLIKE a "best" teacher
0 10 20 30 40 50

References

Bryan, R. C. The use of student reactions to teachers for the purpose of improving instruction and evaluating teacher merit in public elementary and secondary schools. (Research prospectus) Kalamazoo, Mich.: Western Michigan Univer., July 1, 1959. Mimeographed. 23 pp.

Campbell, D. T. Factors relevant to the validity of experiments in social settings. Psychol. Bull., 1957, 54, 297-312.

Cartwright, D., & Harary, F. Structural balance: a generalization of Heider's theory. Psychol. Rev., 1956, 63, 277-293.

Clem, O. M. What do my students think about my teaching? Sch. & Soc., 1930, 31, 96-100.

Cochran, W. G., & Cox, Gertrude M. Experimental designs. New York: Wiley, 1950.

Coombs, C. H. Theory and methods of social measurement. In L. Festinger & D. Katz (Eds.), Research methods in the behavioral sciences. New York: Dryden, 1953. Pp. 471-535.

Festinger, L. The analysis of sociograms using matrix algebra. Human Relat., 1949, 2, 153-158.

Festinger, L. A theory of cognitive dissonance. Evanston, Ill.: Row, Peterson, 1957.

Flinn, V. Teacher rating by pupils. Educ. Method, 1932, 11, 290-294.

Heider, F. The psychology of interpersonal relations. New York: Wiley, 1958.

Jones, L. V., & Fiske, D. W. Models for testing the significance of combined results. Psychol. Bull., 1953, 50, 375-382.

Morsh, J. E., & Wilder, Eleanor W. Identifying the effective instructor: a review of the quantitative studies, 1900-1952. Lackland AFB, San Antonio, Tex.: Air Force Personnel and Training Research Center, Research Bulletin, TR-54-44.

Newcomb, T. M. Individual systems of orientation. In S. Koch (Ed.), Psychology: a study of a science. (Vol. 3) New York: McGraw-Hill, 1959. Pp. 384-422.

Osgood, C. E., & Tannenbaum, P. H. The principle of congruity in the prediction of attitude change. Psychol. Rev., 1955, 62, 42-55.

Riley, J. W., Jr., Ryan, B. F., & Lifshitz, Marcia. The student looks at his teacher. New Brunswick, N. J.: Rutgers Univer. Press, 1950.

Rotter, J. B. Some implications of a social learning theory for the prediction of goal directed behavior from testing procedures. Psychol. Rev., 1960, 67, 301-316.

Savage, Marjorie L. Changes in student teachers through use of pupil ratings. Unpublished Ed.D. thesis, Univer. of Illinois, 1957.

Schutte, T. H. The teacher, through the students' eyes. Amer. Sch. Bd J., 1926, 73 (3), 72-79.

Solomon, R. L. An extension of control group design. Psychol. Bull., 1949, 46, 137-150.

Stuit, D. B., & Ebel, R. L. Instructor rating at a large state university. Coll. & Univer., 1952, 27, 247-254.

Ward, W. D., Remmers, H. H., & Schmalzried, N. T. The training of teacher-personality by means of student-ratings. Sch. & Soc., 1941, 53, 189-193.

Zajonc, R. The concepts of balance, congruity, and dissonance. Publ. Opin. Quart., 1960, 24, 280-296.