1962

A Study of the Effects of Self-Evaluation Ratings in terms of Effort and Conduct by Homogeneously Grouped Seventh Grade Students

Horace R. Nichols
Central Washington University

A STUDY OF THE EFFECTS OF SELF-EVALUATION RATINGS IN TERMS OF EFFORT AND CONDUCT BY HOMOGENEOUSLY GROUPED SEVENTH GRADE STUDENTS

A Thesis

Presented to

the Graduate Faculty

Central Washington State College

In Partial Fulfillment

of the Requirements for the Degree

Master of Education

---

by

H. R. Nichols

August 1962

APPROVED FOR THE GRADUATE FACULTY

_____

Eldon E. Jacobsen, COMMITTEE CHAIRMAN

_____

Howard B. Robinson

_____

M. L. Pettit

ACKNOWLEDGMENT

I am indebted to Dr. Eldon E. Jacobsen, my graduate committee chairman, for his professional criticism and encouragement in directing the writing of this paper.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

CHAPTER I

INTRODUCTION AND STATEMENT OF THE PROBLEM

I. INTRODUCTION

Authoritative purpose. For years educators have heard the shallow plea for a better marking system in the schools. Marks can and do arouse emotion in students. Symonds states "it is common for the pupil to think that marks indicate how the teacher feels toward him, 'she likes me,' 'she has it in for me'" (9:138). It is not uncommon for students to place the blame for low marks on the teacher. Students claim it was the teacher's evaluation and therefore "his fault." Because evaluations made externally by the teacher have an unfavorable impact on the student, some educators have proposed that the student evaluate himself.

Interest in self-evaluation has brought forth many appealing articles in the journals. Twenty years ago Hamalainen stressed that something needed to be done concerning self-evaluation: "The extent to which the pupil should contribute to his own evaluation is not clear. It is certain that he should enter into the process; yet how and to what extent is only partly understood at present" (4:182). Orata argues that "the teacher should not evaluate the work of the pupils; they should do the

evaluating, themselves. If not, why emphasize the objective of self-direction..?" (5:652). In a more definite vein, Duel enumerates seven reasons why self-evaluation would be beneficial:

1. Students develop more sensitivity to desired outcomes.
2. Students achieve better understanding of how they measure up to established standards.
3. Students develop more awareness of requirements and expectations.
4. Students are motivated through a challenge to "beat themselves."
5. Students are oriented toward a look at "self" and a reliance on "self."
6. Some of the fear component of solely external evaluation is removed.
7. Students develop a clearer frame of reference upon which to base future actions (3:52).

Finally, Rogers upholds the concept of self-evaluation in student-centered courses as empirically sound. He states flatly that "our experience has corroborated the theoretical principle that self-evaluation is the most desirable mode of appraisal in a student-centered course" (7:415). It follows, then, that the external locus of evaluation for the student needs reviewing.

Theoretical background. Because of its special concern with student self-evaluation, this study falls

neatly in the fold of Rogers' client-center theory of personality (7:481:532). Briefly, this theory trusts the individual to determine his own mode of behavior. He contends that the person has a tendency to strive to actualize, maintain, and enhance himself. This forward moving tendency is inherent in all people. It would seem to follow that the tendency of the self to actualize itself is impeded by some fear of external evaluation. The person is not free to make realistic choices congruent with his nature. It is hypothesized that if this threat is removed, he will be able to discriminate between progressive, static, and regressive ways of behavior. The worth of an alternative will be judged by the person and not someone outside himself. He will, therefore, accept his choice in a realistic light. The important concept here is the encouragement of independent decision making. This study is concerned with these theoretical concepts.

II. THE PROBLEM

Statement of the problem. This investigation was undertaken to discover what changes, if any, would occur in seventh grade students' school effort and conduct if they are given the opportunity to evaluate their own effort and conduct at weekly intervals for eighteen weeks. The

problem was divided into the following specific questions to facilitate a thorough investigation: (1) Is there a significant relationship between the ratings given by the student himself (self-rater) and by the teacher (teacher-rater)? (2) Are the student self-raters and the teacher-rater able to discriminate between effort and conduct? (3) Is there improved effort by seventh grade students as rated by the student self-raters and by the teacher-rater? (4) Do academic grades improve following periodic student self-evaluation of effort? (5) Is there improved conduct by seventh grade students as rated by the student self-raters and by the teacher-rater? (6) What feelings and attitudes are held by seventh grade students concerning self-evaluation? So stated, it is necessary to define, somewhat operationally, terms used in this study.

Definition of terms. Self-evaluation is used in the sense of judgments or decisions made personally by the student and manifested by his making a written evaluation on a prepared scale.

School effort entails all demonstrated industry necessary to complete assignments, to study in and out of school, and to perform various tasks in the classroom. These are independently indicated by teacher-ratings and the student self-ratings.

School conduct is an inclusive term used here to mean behavior that conforms to expected school and classroom rules. Again, this is judged independently by the teacher-rating and the student self-rating.

Indeed, marks are a problem in schools when students become emotionally upset about them. That is, they often cannot fully accept the marks given them if they do not understand the evaluation made. This has encouraged many to write on the subject of evaluation, and specifically, self-evaluation. Other than Rogers' statement that self-evaluation "works," we have little experimental evidence to confirm or refute this finding. Using a client-centered framework, this investigator has attempted to throw more light on this relatively novel area of education. Specifically, this study is limited to the problem of determining whether a student's effort and conduct will improve if he is given the well defined opportunity to evaluate himself in these phases of his education.

CHAPTER II

REVIEW OF THE LITERATURE

Although a great deal has been written about self-evaluation on both authoritative and theoretical levels, very little research on self-evaluation has been reported. Self-evaluation has been commonly accepted at the verbal level, but rarely at the action level. Russell, in a 1953 survey of self-evaluation research, reports a lag of scientific study. He states: "This review of some published and unpublished studies indicates a lack of scientific study of the values of day-by-day evaluation in the learning activities of the modern school" (8:570). The research on self-evaluation since 1953 has not increased a great deal. Theoretically, self-evaluation as a technique has many encouraging proponents. Yet few investigators are willing to subject the idea to research methods. The reason this paradox exists is not the concern of this report. However, a hint may be gleaned from examination of the following related literature. The research falls into two categories: (1) relationship between self-estimated ratings and actual rating and (2) use of self-ratings to bring about new learnings.

I. THE RELATIONSHIP BETWEEN SELF-ESTIMATED RATINGS AND ACTUAL RATINGS

Evidence to date suggests that students either over-

rate or under-rate themselves when given the opportunity. Tschechtlin's (11:25-32) study showed this characteristic of elementary school children nearly twenty years ago. Using a sample of thirty-four public and parochial schools throughout the state, with populations ranging from urban to rural and the father's employment socially stratified, she found that girls rated themselves higher on a 22 trait personality scale than other girls, boys, and teachers rated them. Boys tended to under-rate themselves. Webb (14:305-07) duplicated the study with male adult Jews. Using a different personality trait scale, he found a tendency for the individual to rate himself higher than the group rated him. Similarly, Powell (6:225-234) found low correlations between self-insight ratings of adjustment of college girls and the ratings of peers and experts.

We might hastily conclude that self-evaluations of personality are not related to rating of others. However, Ullmann (12:1-36) refined the above studies by separating sexes. He found that external raters are more valid assessors of the adjustment of boys than girls. Yet self-descriptive personality tests such as the California Personality Test appear to have more in common with other indices of adjustment of girls than boys. Self-evaluations apparently are more meaningful to girls than boys.

The tendency for persons to differ in their personality self-ratings from external raters may also hold true in achievement ratings. Asch (1:1-23) investigated the acquisition of knowledge of elementary psychology between student-centered and lecture-oriented groups. He found that only the lecture class gained significantly in learning the subject matter of the course. Because of the self-directed nature of the student-centered class, this group was permitted to assign their own course grade. He believes that the majority of the students were honest in the final grade they gave themselves. However, he is aware that some students gave themselves A's or B's and didn't have the knowledge commensurate as judged by the final test scores. From the various researches it seems that self-raters may often give inaccurate evaluations. However, we must be careful not to conclude self-raters to be completely inaccurate when the ratings are compared to others, such as a teacher-rater.

The well-known Wickman study (15:122) of 1928 even questioned the validity of teacher ratings. When the teachers' ratings were compared with clinicians' ratings of maladjustment in children, negative rank order coefficients of -.22 and -.11 were found. However, studies since 1928 have shown that teachers and clinicians have a more common

judgment of personality problems. For example, in 1951 Ullmann (12:29) found a high relationship (r .86) between teachers and clinicians on traits that teachers considered favorable about pupils. It may be concluded from this that under carefully defined conditions, teachers are relatively accurate in rating personality traits of children. How similar, however, are intellectual understanding of personality traits and personal evaluations of students by the teacher? Does this mean that because teachers objectively believe the way psychologists do that they may still act differently toward the antagonistic child and thus rate him accordingly? Are we left pondering the validity of any rater, even clinicians, when under fire in a classroom? These questions raise a delicate problem of the rater's own self-structure that exceeds the limits of this study. Because rater accuracy is a problem, other investigators have pursued learning outcomes.

II. THE USE OF SELF-RATINGS TO BRING ABOUT NEW LEARNINGS

Until recently, little has been done to measure the outcomes of self-evaluation. The Asch (1:22) study reports that self-raters may gain in other ways in spite of "erroneous" self-evaluation. Using the Minnesota Multiphasic Personality Inventory (MMPI), he found that the

self-raters became better adjusted than the control group. Clearly significant changes in the experimental self-rating group were shown on the MMPI scales.

The two studies that follow are attempts to understand how and to what extent self-ratings are significant. Taylor (10:205-08) gave 120 college students 60 positive and 60 negative items to sort according to their self-ideal. The students were asked to re-rate themselves one week later to determine any difference between scorings. He found positive growth between the subjects' self-ideal and the self but only at the 10 per cent level of confidence. In other words, "self-introspection by self-description without therapy may be accompanied by some of the changes reported in successful counseling, which presumably also involves rather intensive introspection on self. . ." (10:208). Although the finding was not significant at the required critical level of confidence, the trend seemed evident. On the other hand, Duel (2:197-199) did a seemingly excellent study on USAF students in a technical school. His sample consisted of 75 experimentals and 75 controls from School A and 33 experimentals and 33 controls from School B. The experimental groups were given a rating scale at the beginning of technical school, every two or three weeks thereafter, and also after they had completed

the course to assess what they thought their technical knowledge competency was at these intervals. The scales were structured. Their task was to self-rate how well they knew the material learned. Test scores of the controls were compared with the experimentals. Mastery of the course work in favor of the experimentals was significant at the 1 and 2 per cent level of confidence. Duel concludes that, "the results lead to the conclusion that in this particular situation, students given formal and periodic opportunities to evaluate themselves, can achieve to a greater degree than students not having such opportunity" (2:199). Both studies point to the value of self-evaluation. With this positive note as background, the procedures of this investigation will be considered.

CHAPTER III

SAMPLE POPULATIONS AND PROCEDURES

I. SAMPLE POPULATIONS

Two contrasted samples of seventh grade pupils were used in the study. Five homogeneously distinct and separate classes comprised the total seventh grade at Prosser Junior High School during the year (1961-62) of this study. Over 500 students, of which 150 were seventh graders, attended the school. The homogeneous grouping was based on the Stanford Achievement Test scores. The pupils are given the test in the spring, and the results are used to place them in various levels of classes the following fall. One of the seventh grade samples (hereafter referred to as Group I) used in this study received a mean grade placement score of 6.5. The other sample (hereafter referred to as Group II) used in this study received a mean grade placement of 7.4. Tests from the previous spring placed Group I as average achieving among the five sections and Group II as high-average achieving since one class was above them. Table I shows the relationship of Group I and Group II. Boys of Group I and II had a mean grade placement of 6.4 and 7.2 respectively. This difference in measured achievement was significant beyond the 1 per cent level. Girls from Groups I and II, similarly,

TABLE I

COMPARISONS BETWEEN GROUPS I AND II USING

THE CALIFORNIA TEST OF MENTAL MATURITY AND STANFORD ACHIEVEMENT

INSTRUMENT	N	GROUP MEAN	STANDARD DEVIATION	STANDARD ERROR OF THE MEAN	Df	t	SIGNIFICANCE LEVEL
California Test of Mental Maturity							
Boys							
Group I	11	100	13.30	4.20	24	3.18	Beyond 1%
Group II	14	116	10.00	2.77	24		
California Test of Mental Maturity							
Girls							
Group I	13	109	13.19	3.80	24	.66	Not
Group II	12	112	8.09	2.43	24		Significant
Stanford Achievement Test (Composite Spring 6th Grade)							
Boys							
Group I	11	6.4	.47	.15	23	3.33	Beyond 1%
Group II	14	7.2	.66	.18	23		
Stanford Achievement Test (Composite Spring 6th Grade)							
Girls							
Group I	13	6.5	.57	.16	23	3.80	Beyond 1%
Group II	12	7.9	1.14	.34	23		

registered mean grade placements showing that they were two distinct groups in terms of achievement. The mean grade placements of 6.5 and 7.9 respectively is significantly different at the 1 per cent level. Because Groups I and II are not randomly selected from a common population, they cannot be directly compared statistically. On occasion, however, both groups may be compared for the sake of illustrating trends following statistical tests using each group as its own control. Most of the experimental work involves Group I.

II. PROCEDURES

For the first two quarters Group I subjects served as a control on themselves. At the beginning of both the third and fourth quarters, Group I was given the self-evaluation rating sheet to use each Friday. The rating sheet (Appendix A) was developed by the investigator to serve the purposes of this study. It contains four statements with five levels of items from high to low under each statement. Two statements pertain to effort and two relate to conduct. With the help of the teacher, the instructions were made clear in the first rating period. Students were told that they would rate their own effort and conduct with the self-rating sheet on a weekly basis for eighteen weeks. The only restriction asked was that

they be honest about the grade they gave themselves. At the end of the third and fourth quarters, they took home to show their parents the actual grade they had given themselves in effort and conduct. In other words, as far as the Group I students were concerned, they were responsible for their own effort and/or conduct decisions.

In order to study possible progress of Group I using the self-evaluation method the last two quarters, external ratings were obtained from the teacher in the form of (1) Course Grades, and (2) Effort and Conduct Evaluations. Since the same teacher was used in this study for Groups I and II, the teacher difference variable was minimized. The teacher taught both groups geography; Group I was also taught English by the same teacher. Therefore, it was possible to gather quarter grades for these two subjects and compare the results of each quarter for both groups.

Further, the teacher was asked to evaluate both groups in effort and conduct. He did this once at the end of each quarter beginning with the completion of the second quarter. The approach to the rating was kept as similar as possible to the way a teacher normally evaluates students. He used letter ratings ranging from A to X (A superior, B good, C average, D poor, and X failure).

Students in Group I were unaware of the evaluation by the teacher. As far as they were concerned, their self-evaluations were the only ratings in effort and conduct that would be used. In retrospect then, evaluation in the form of grades and teacher evaluations of conduct and effort were obtained to later test the feasibility of self-evaluation.

To throw more light on the feelings that Group I had about self-evaluation, a questionnaire was given to the subjects at the conclusion of the experiment. Group II was given a similar form of the questionnaire. In this way, the feelings Group I expressed about self-evaluation were compared with those of Group II. To add qualitative data, the teacher was given a questionnaire at the conclusion of the study to assess his feelings about this method.

The necessarily contrasted samples posed a problem in research design. Statistically, intra- and inter-comparison were computed for both groups. The same subjects of Group I were used in both capacities of experimental and control conditions. A direct-difference statistical method was used (13:167-171). As previously noted, the external ratings by the teacher were statistically analyzed quarterly. Mean differences were computed to determine a significance level. In other words, the use of the subjects for both

control and experimental conditions handled the problem of matched sampling. Before leaving the statistical procedure for Group I, it is well to note that a correlational study was made between teacher-rater and self-rater. This was done in order to determine if the ratings by the teacher and students were similar.

It seemed that results obtained from Group I would have greater meaning if another group were compared with them. Although no direct statistical comparisons on factors known to be related to achievement can be made, data obtained from Group II can be compared with Group I for illustrative purposes. However, when the proportions from the questionnaire were gathered, a direct comparison was legitimate between both groups. The formula $\delta \text{ prop} = \sqrt{\frac{pq}{N}}$ was used (13:117-20). The assumption was that feelings differ more from person to person than from group to group.

In summary, two contrasted samples of seventh grade pupils were selected. Group I was considered an average achieving group; Group II was high-average. Because both groups are statistically distinct in measured achievement, they cannot be directly compared on this variable. Except for the questionnaire items given to both, other comparative data between the two are useful only for illustrative purposes. Group I served as its own control for one

semester and as an experimental group for the other. After gathering the results of Grades and also Teacher-Ratings for the last two quarters, mean differences between the control and experimental semesters were compared statistically.

CHAPTER IV

RESULTS

The general problem of this study was to determine if seventh grade students' school effort and conduct improve if they are given the opportunity to evaluate their own effort and conduct at weekly intervals. Particularly, the following specific problems were researched:

1. Are the ratings of the self-raters and the teacher-rater similar?
2. Can the self-raters and the teacher-rater discriminate between effort and conduct?
3. Will effort as rated by self-raters and teacher-rater increase following experience in self-evaluation?
4. Will grades of self-raters increase following experience in self-evaluation?
5. Will conduct as rated by self-raters and teacher-rater increase following experience in self-evaluation?
6. What feelings and attitudes are held about self-evaluation of effort and conduct?

Each of the above will be discussed as separate problems to better understand the general problem.

I. THE RELATIONSHIP BETWEEN SELF-ESTIMATED RATINGS AND TEACHER-RATINGS

Related research suggested a careful examination of self-rater accuracy be made. As reported earlier, several studies (1; 6; 10; 11; 12) found low correlation between the self-raters and the teacher-rater. Because many of the studies concluded that self-raters were not in agreement with "other" raters, this investigation sought to determine the relationship between the student self-raters and the teacher-rater.

This investigation found that in general the student did rate himself higher than the teacher rated him. Yet a moderately high positive relationship exists between self-ratings and teacher-ratings in effort and conduct. Examination of Table II reveals a .65 correlation between the teacher-rating of effort and student self-rating of effort for the fourth quarter. Accordingly the mean teacher rating of conduct and the mean student rating of conduct resulting in a moderate r of .49. In both incidences the relationship was significant at the 1 per cent level of confidence.

This study does not completely agree with earlier findings concerning the non-similarity of ratings by self-raters and "other" raters; two reasons may account for this.

TABLE II

CORRELATIONS BETWEEN THE TEACHER AND THE STUDENTS IN TERMS OF EFFORT AND CONDUCT RATINGS FOR THE FOURTH QUARTER

Group I	Teacher Ratings Effort	Teacher Ratings Conduct
Self-ratings Effort	+ .65*	
Self-ratings Conduct		+ .49*

* Significant at 1% level

First, the long rating period of eighteen weeks with defined rating levels on the rating sheet may have aided the self-raters to make more realistic decisions. Or, secondly, the teacher-rater and student self-raters were in closer agreement because both followed the same criteria in judging effort and conduct. Nevertheless, the student- and teacher-raters in this study were more related than other investigators have reported from their findings.

II. THE RELATIONSHIP BETWEEN EFFORT AND CONDUCT RATINGS

Since two purportedly distinct factors of effort and conduct have been rated, it is well to determine if the raters were able to discriminate between the two during rating periods. Table III shows the correlations between self-raters and the teacher-rater for the fourth quarter on these two factors. The self-raters definitely discriminated between effort and conduct as the correlation of r .17 was so low and proved not significantly related. On the other hand, the teacher-rater evaluated effort and conduct quite similarly for Group I. The obtained correlation of r .62 was significant beyond the 1 per cent level. Also, the teacher rating of Group II in terms of effort and conduct was related. This r of .40 was significant at the 5 per cent level. From this, it would seem that the self-raters

TABLE III

CORRELATIONS BETWEEN EFFORT AND CONDUCT RATINGS OF GROUP I (SELF-RATERS AND TEACHER-RATER) AND GROUP II (TEACHER-RATER) FOR THE FOURTH QUARTER

	Student Self-Ratings Conduct Group I	Teacher Ratings Conduct Group I and Group II
Student Self-Ratings Effort Group I	+ .17	
Teacher Ratings Effort Group I		+ .62**
Teacher Ratings Effort Group II		+ .40*

Not significant
* Significant at the 5% level
** Significant at the 1% level

discriminated between effort and conduct more than the teacher-rater. The factors of effort and conduct apparently are somewhat independent for students, but teachers tend to tie effort in with conduct.

III. THE ANALYSIS OF EFFORT RATINGS BY THE TEACHER AND THE STUDENT

With a moderate reliability between the teacher-rater and the self-raters established, we may properly hypothesize that if self-evaluation will result in increased effort for Group I, the raters should be able to demonstrate this growth. A comparison of the ratings given to Group I by the teacher at the conclusion of the second quarter and again at the end of the fourth quarter revealed a non-significant gain in effort. Examination of the results in Table IV demonstrates that by chance the mean difference of .32 could occur 15 times in 100. This great a margin might imply a tendency toward effort improvement to be checked later by larger samples with more teachers involved.

The tendency for effort improvement in Group I is further illustrated when Group I is compared with Group II. Again, the teacher rated the effort of this group at the end of the second, third, and fourth quarters. Group II had arithmetical means of 3.01 second quarter, 2.45 third quarter, and 2.61 fourth quarter. Clearly this is a

TABLE IV

COMPARISONS OF SECOND AND FOURTH QUARTER EFFORT RATINGS FOR GROUP I AND GROUP II

Student and Teacher Effort Ratings for Groups I and II	N	Mean Rating Scores	Mean Difference	Standard Error of the Mean Diff.	df	t	Significance Level
Student Effort Self-Ratings							
Group I			.12	.50	24	.25	Not Significant
Third Quarter	25	2.56					
Fourth Quarter	25	2.68					
Teacher Effort Ratings							
Group I			.32	.22	24	1.45	Not Significant
Second Quarter	25	1.96					
Fourth Quarter	25	2.28					
Teacher Effort Ratings							
Group II			.40	.06	25	6.66	Beyond 1%
Second Quarter	26	3.01					
Fourth Quarter	26	2.61					

decrease in effort by Group II as rated by the teacher, whereas a gain was recorded for Group I between the second and fourth quarters. This gain, to reiterate, was not statistically significant. Although Group II and Group I could not be equated on achievement, it is remembered that Group II was the better achievers. Thus logically, despite inability to compare statistically, the self-raters' improved in effort relative to the group without self-rating experience.

When the self-evaluations of effort by Group I were analyzed for the third and fourth quarters, a slight increase in arithmetical means was found. However, the third quarter mean of 2.56 compared with the fourth quarter mean of 2.68 was not significant. Table IV gives these figures. Although the teacher rating and student self-evaluation shows a slight gain, the hypothesis that self-evaluation will improve effort as judged by the teacher and self-raters must be rejected for this sample. The question that must now be asked is how accurate are the ratings in terms of effort. Did the Group I subjects really register a significant effort gain even though the raters were unable to detect a significant improvement? A look at grades might help answer this question.

IV. COURSE GRADES AS ANOTHER DEFINITION OF EFFORT

A course grade is the result of many factors such as ability, effort, and even conduct. Disregarding all variables except effort, it is possible to hypothesize that grades will improve if effort increases. From the results above it is suggested that grades will not improve significantly. Group I was given the self-evaluation device throughout the second semester. Comparing the first semester grades of geography and English with second semester grades in these two subjects, no improvement was revealed. In fact, examination of Table V for Group I shows a mean difference loss of .08 for both geography and English.

Using Group II as a comparative check, we find no gains for this group in either geography or English. The arithmetical means of Group II for geography first and second semesters respectively were 2.15 and 2.15. For English the means were 2.25 and 2.10, demonstrating a mean difference loss of .15 for this subject.

The comparison of effort and grades by Group I leads to the acceptance of the hypothesis that improvement in effort will lead to improved grades. As was found earlier, Group I did not show a significant increase of effort as

TABLE V

COMPARISONS OF FIRST AND SECOND SEMESTER MEAN GRADE POINT AVERAGES

IN GEOGRAPHY AND ENGLISH FOR GROUP I AND GROUP II

Geography and English Grades of Group I and II	N	Mean G.P.A.	Mean Difference	Standard Error of the Mean Diff.	df	t	Significance Level
Geography Grades							
Group I			.08	.12	24	.66	Not Significant
First Semester	25	1.50					
Second Semester	25	1.42					
English Grades							
Group I			.08	.09	24	.88	Not Significant
First Semester	25	1.78					
Second Semester	25	1.70					
Geography Grades							
Group II			.00	.00	25	.00	Not Significant
First Semester	26	2.15					
Second Semester	26	2.15					
English Grades							
Group II			.15	.14	25	1.07	Not Significant
First Semester	26	2.25					
Second Semester	26	2.10					

rated by the teacher. In conjunction, it was found they did not make better marks. This shows that there may well be the positive relationship long suspected between teacher evaluation of effort and grades, and that the results of ratings on effort might even be used to predict course grades. Consideration of the last three hypotheses discussed suggests that teachers may not always distinguish between conduct, effort, and achievement when rating students. That is, the teacher as an evaluator may make separate ratings for these marks but such may not thoroughly differentiate the traits. Rather, a generalization of characteristics or "halo effect" seems to be present.

V. SELF-EVALUATION OF CONDUCT AS SIGNIFICANT

It was hypothesized that the Group I's conduct would improve as a result of self-evaluation. For the first two quarters Group I was its own control. At the beginning of the third quarter and continuing until the completion of the fourth quarter, Group I was given the opportunity to evaluate their own conduct. The students rated themselves higher than the teacher rated them. Table VI shows a third quarter mean of 3.00 for conduct. This increased to a mean of 3.32 for the fourth quarter. The t difference of 3.55 places their perceived improvement as significant beyond the 1 per cent level of confidence.

TABLE VI

COMPARISONS OF SECOND AND FOURTH QUARTER

CONDUCT RATINGS FOR GROUP I AND GROUP II

Student and Teacher Conduct Ratings for Groups I and II	N	Mean Rating Scores	Mean Difference	Standard Error of the Mean Diff.	df	t	Significance Level
Student Conduct Self-Ratings							
Group I			.32	.09	23	3.55	Beyond 1%
Third Quarter	25	3.00					
Fourth Quarter	25	3.32					
Teacher Conduct Ratings							
Group I			.90	.18	23	5.22	Beyond 1%
Second Quarter	25	2.04					
Fourth Quarter	25	2.94					
Teacher Conduct Ratings							
Group II			.15	.17	24	.88	Not Significant
Second Quarter	26	3.00					
Fourth Quarter	26	3.15					

The teacher ratings also indicated conduct improvement. The results, Table VI demonstrate a mean of 2.04 for the end of the second quarter and a mean of 2.96 at the completion of the fourth quarter. The obtained *t* of 5.22 is clearly significant beyond the 1 per cent level of confidence. Thus, for Group I it may be concluded that the teacher-rater and the student self-raters believe that they did improve their conduct.

Group II, which had no opportunity for self-evaluation, shows no improvement in conduct. The teacher-rated conduct for Group II's second and fourth quarters resulted in means of 3.00 and 3.15 respectively. The *t* of .62 presented in Table VI shows no significant difference in means. Since Group II was originally identified as a higher achieving group than Group I, it is interesting to note how the two groups compared on conduct for the final evaluation quarter. A mean difference of .19 was computed between the Group I mean of 2.96 and Group II mean of 3.15. Applying the *t*-test out of curiosity, since they were not in equivalent achievement groupings, revealed no significant difference (t-.95;df 49). From the results obtained for this problem, it must be concluded that the evaluation of conduct by Group I demonstrated perceived self improvement. To acquire a clearer picture, the feelings and attitudes

that subjects have about self-evaluation are next presented.

VI. FEELINGS AS INDICATORS OF SUCCESS

Since it has been demonstrated that self-evaluation of conduct has been instrumental in improving Group I's rated behavior, this investigator assumes that self-evaluation is emotionally meaningful for the student. That is, self-evaluation without the threat of an external evaluator may very well have a positive effect on the self-concept of the student. They were asked to give their honest feelings about self-evaluation. Subjects were told before they answered the questionnaire that they were part of a research study and that complete honesty for the "sake of science" was necessary.

"This had helped me to do better work. I try to study harder. My grades have been starting to get higher. . ." illustrates unstructured written feedback. Another verbatim comment was: "I'm in favor of the student doing their own evaluation. It shows them what they have to work on to get a better citizenship grade." This particular student's conduct improved from a 2.00 second quarter to a 4.00 fourth quarter. A few more remarks will illustrate that the students fully understood the value of making their own decisions from well defined alternatives. The following

four student statements highlight this concept:

(1) I feel that this is a good idea to let the children evaluate themselves. By doing this the children can truly see their faults and should try to improve.

(2) I liked to evaluate myself because I could see what I was low in and do better the next time.

(3) It is a good idea. To see how you have progressed and not progressed and to try to improve yourself.

(4) I felt that it helped me when I marked down my grade and helped me to improve my grade a little better each week. When the teacher does it I just get the grade and don't know whether to improve my grade or not.

In a more quantifying vein, the results of the questionnaires given to Groups I and II are tabulated in Tables VII and VIII.

A review of a few questions will suffice to demonstrate that a self-evaluation may well affect beneficially student feeling. However, a complete analysis will not be attempted as the results in Tables VII and VIII are self evident. When possible, the questions asked Groups I and II were constructed similarly to facilitate comparison. Further, the results from Group II were arbitrarily interpreted as representative, for sake of comparison, of a seventh grade population. Accepting this assumption, the results from Group I were statistically compared with

TABLE VII

RESULTS OF THE QUESTIONNAIRE GIVEN TO

GROUP I FOLLOWING THE RATING EXPERIENCE

QUESTIONS	PROPORTIONS BY PER CENT GROUP I	GROUP II	TRAIT	df	t	SIG. LEVEL
1. Do you think that you made better grades as a result of evaluating your own effort?						
a. helped very much	14					
b. helped some	73					
c. not helped	12					
d. made poorer grades	1					
2. Do you think that you behaved better in the classroom and school as a result of evaluating your own conduct or citizenship?						
a. helped very much	40					
b. helped some	44					
c. not helped	16					
d. was a worse citizen	0					
*3. What feelings do you have toward evaluating your own grades in effort and conduct?						
a. angry	0	0				
b. somewhat fearful	6	38	FEAR	11	2.36	5%
c. both angry and fearful	10	4				
d. happy	36	8	HAPPY	9	6.51	1%
e. no feelings	48	50				
*4. When evaluating yourself did you give yourself an honest grade each week?						
a. always	55	85	ALWAYS HONEST	34	2.50	1%
b. most often	45	15				
c. half the time	0	0				
d. never	0	0				
5. How seriously did you take the evaluation of yourself in effort and conduct?						
a. very seriously	12					
b. thought about it a little during the week	52					
c. did not worry about my evaluation	36					

* See Table VIII for exact wording for Group II questions

TABLE VII
(continued)

QUESTIONS	PROPORTIONS BY PER CENT GROUP I	GROUP II	TRAIT	df	t	SIG. LEVEL
6. If you gave yourself a low grade in any one of the evaluation questions did you try to do better the next week?						
a. always	28					
b. sometimes	72					
c. never	0					
7. Did you become bored after a few weeks of evaluating yourself?						
a. never	24					
b. sometimes	68					
c. bored every evaluation period	8					
*8. Do you think your evaluation was similar to your teacher's evaluation?						
a. higher than the teacher	25	11	HIGHER	7	2.50	5%
b. the same	63	70	THAN			
c. lower than the teacher	12	19	TEACHER			
*9. Would you rather have your teacher evaluate you in effort and conduct?						
a. definitely have the teacher do it	12	30				
b. have each student do his own as we did it	12	4				
c. both the teacher and the student do it	76	66				
10. Were your parents in favor of your evaluating yourself?						
a. in agreement	16					
b. parent not sure	8					
c. against it - felt this was the teacher's job	4					
d. never told my parents	72					

* See Table VIII for exact wording for Group II questions

TABLE VIII

RESULTS OF THE QUESTIONNAIRE GIVEN TO

GROUP II FOLLOWING THE EXPERIMENT

QUESTIONS	PROPORTIONS BY PER CENT
1. Are you using your study time the best way you can?	
a. yes	30
b. sometimes	70
c. no	0
2. Do you follow the school rules?	
a. always	74
b. sometimes	26
c. never	0
3. Do you think you would study better and follow the school rules better if you had to grade yourself in study habits and school conduct?	
a. definitely	27
b. not sure	61
c. no difference	12
4. Would you grade yourself fairly?	
a. yes	85
b. no	15
5. What feelings would you have about evaluating your effort and conduct?	
a. angry	0
b. somewhat fearful	38
c. both angry and afraid	4
d. happy	8
e. no feelings	50
6. Do you think your evaluation of effort and conduct would be similar to your teacher's?	
a. higher than the teacher's	11
b. the same	70
c. lower than the teacher's	19
7. Would you rather have your teacher evaluate you in effort and conduct?	
a. definitely have the teacher do it	30
b. have each student do his own	4
c. both the teacher and the student do it	66

Group II to determine the difference between proportions. Proportions obtained from Group II, then, are considered hypothetical true proportions. Obtained proportions from Group I are statistically compared with Group II by the formula, $t = \frac{P_1 - P_2}{\sigma \text{ prop}}$ (13:118).

Of interest is the comparison of anger, fear, and happiness between the groups. When Group I was asked what feelings they had about self-evaluation, 8 per cent reported "somewhat fearful" and 36 per cent "happy." The reverse of these percentages of 38 per cent for fear and 8 per cent for pleasure was recorded by Group II. About half in each group did not register any feeling about self-evaluation. However, the reported proportion difference both for fear and happiness between the two groups was significant at the 5 per cent and 1 per cent level of confidence respectively ($\underline{t}$ 2.36 for fear and $\underline{t}$ 6.51 for happy). Thus for Group I self-evaluation was more positively viewed than for the Group II subjects, who had not experienced this method.

Although Group I subjects reported they enjoyed evaluating themselves, they along with Group II subjects expressed the desire to have the teacher help with evaluations. To the question of having both the teacher and the student evaluate effort and conduct, 76 per cent of Group I subjects and 66 per cent of Group II responded

affirmatively. It would seem that students have a real need for structure. Apparently, the structured self-rating scale given Group I did not afford sufficient evaluation security. Yet examination of questions dealing with "honesty" may have a direct bearing on the need for being evaluated by the teacher.

It might be surmised that the majority of students would "always" give themselves ratings that were correct to the best of their judgment. Eighty-five per cent of Group II subjects reported they would "always" be "honest." On the other hand, only 55 per cent of Group I subjects perceived themselves so. The difference between proportions (85 per cent for Group II and 55 per cent for Group I) was significant at the 1 per cent level of confidence. This difference could imply that actual self-evaluation tempts one to be a little dishonest about recording performance.

From the data so far discussed it might be concluded that self-evaluation as practiced by students and conceptualized by others is somewhat distinct in terms of the following: (1) self-raters had less fear and enjoyed the experience more than those anticipating self-evaluation, (2) both the self-raters and the controls expressed the need for teacher's judgment, (3) the self-raters report that they were more dishonest about their ratings than was

expected. From this, two questions might be asked. Perhaps those who have not experienced self-evaluation need the teacher's evaluations because they fear a lack of structure or need a definition of limits. Whereas the need for a teacher rating by those who had the experience was necessitated more from a specific desire to control impulses (dishonesty). Certainly the difference warrants further investigation. Two more questions and results concerning Group I's opinions are next presented to support the quantitative data of this study.

Because the study dealt primarily with effort and conduct, the self-raters were asked to report their opinions. Group I was asked directly if they "made better grades" and "behaved better in the classroom" as a result of self-evaluation. It will be noticed from Table VII that only 14 per cent of the self-raters felt that they were "helped very much" to make better grades. Whereas, 40 per cent replied that they were "helped very much" to be better citizens. It may be recalled that earlier in the report they were rated by their teacher as not making effort improvement, but did demonstrate significant gains in conduct. Both findings would be interpreted to be complementary and should add to the significance of this investigation.

Before digesting the results of this study, an interesting but not surprising feature of the questionnaire was the reply to Question 10. Seventy-two per cent of Group I subjects stated that they never told their parents about their self-evaluation opportunity. Conjectures are left to the reader.

In summary, the problem of this research report was concerned with understanding the effects of self-evaluation in terms of effort and conduct. Two different but homogeneously grouped seventh grade classes provided the subjects. Six problems provided the basis for the results. First, it was concluded that the teacher-rater and the self-raters had a moderate relationship between them when asked to evaluate, separately, effort and conduct. This finding is inconsistent with earlier studies showing little agreement between raters. It was suggested that (1) amount of rating trials and (2) structured rating alternatives may account for the closer agreement between raters in this study. Second, it was found that the self-rater tended to discriminate more between effort and conduct than did the teacher-rater. Third, it was found that self-evaluation in this study did not significantly improve the effort of the self-raters. It was hypothesized that effort, gain or loss, can be quite accurately indicated by observing the barometer

of grades. Fourth, this close relationship between effort and grades was found to exist. Since effort did not improve in the self-raters, it would follow, as the results support, they did not improve their grades as a group. Fifth, self-evaluation did significantly help the self-raters improve their classroom conduct. This was further substantiated when the results from the questionnaire were analyzed. Finally, it was discovered by the questionnaire approach that students under the self-evaluation method differed on some opinions with students not subjected to self-evaluation. They agreed, however, that both wanted the teacher to be part of the evaluation picture.

CHAPTER V

DISCUSSION

It was discovered that students are not accurate self-raters. They tend to either under- or over-rate themselves, and usually over-rated their effort or conduct. But by whose standards? Most students reported they were honest about their self-evaluations. Whether their evaluations were right for them, the fact remains that the teacher didn't see it that way. Perhaps there is a breakdown in communications between teacher (or adult society) and student.

It may be recalled that the teacher and students of this study were in closer agreement than other studies have demonstrated. Was this finding unique because of chance or was it a consequence of the design of this study? The key to the closer relationship between the teacher-rater and the student-rater may lie with the self-rating instrument. That is, both the student and the teacher had a specific external criterion to guide them. True, the teacher was to evaluate his students the traditional way, but he may have been influenced by the rating instrument. Certainly, the students had guidelines to follow for the first time.

Another reason this study found the raters to be in closer agreement may be the duration and choice of rating

alternatives given the self-raters. Eighteen weeks allows the student time to learn how to evaluate himself. Also, the high to low items of the self-rating device structured his responses each rating period. That is, certain limitations were set down in order to honestly rate himself. For example, his assignments for any one week must have been "Always turned in and completed on time" to receive an "A" grade. Whatever the reasons for the greater similarities between raters in this study than in earlier research, this investigation should prove fertile in re-opening the door for further examination of self-evaluation.

A research investigation, ideally, should both answer the original problems and ask new questions. On one point, this study does neither. Paradoxically, the problem of the external rater persists like a "hang-over." First it is revealed that a basic theoretical hypothesis of the study is that the individual will not grow toward independence if judgments and decisions are made for him. Next, it was found that a great deal of experience in self-evaluation is necessary before the student-raters evaluate themselves as teachers would. And finally, the results from the questionnaire seem defeating when a total of 88 per cent of the self-raters would have either the teacher or the student with the teacher do the evaluating. Two suggestions

for improvement in later research will be discussed: (1) more items in self-rating instrument, and (2) combined teacher-student ratings.

There were two effort and two conduct items in the self-rating form. The short self-test was used to minimize repetition at later weekly ratings. A longer instrument might have increased the accuracy of the ratings, but too lengthy an assessment could soon bore the person, and therefore, destroy its basic aim. When asked the specific question only 8 per cent of the self-raters were bored every rating period. It would seem that a few more items could have been used in this study. The length of the self-rating device must certainly be a major consideration for future research and general usage.

Also, it may be necessary to include the teacher-rater in the self-evaluation picture on a limited or temporary basis. Perhaps in the beginning the teacher should take a reflective position for those students that need this security. Great emphasis might well be placed on aiding the student to make his final decision autonomously. When students develop the desired frame of reference and find success in making decisions, growth of a self-directed nature should become evident. Occasionally, the teacher will have to work out an over- or under-

estimation by the student. For example, if a student states that he turned in all assignments completed for that week but didn't acually do this, the teacher could demonstrate the error in the grade book. Chronic over- or under-rating could be one sign, among others, suggesting need for counseling. Over-zealous checking by the teacher should be minimized, however, else he may as well make the decisions himself. The foregoing suggestions, then, are two methods that may delimit the role of the external evaluator and, thus, aid the self-rater to rely on himself to make independent decisions.

Finally, the reported improvement of conduct by Group I may well suggest that self-evaluators can gain more understanding about "values" as a result of self-appraisal. If a student is to gain a set of values wherein he ultimately must be, for the most part, his own judge and if he is to make evaluations of anticipated behavior, it follows that some opportunity to learn self-evaluation is better than consistent practice with external evaluation only. Certainly, this inference is not contrary to the theoretical framework of this study.

CHAPTER VI

SUMMARY

This study was limited to the problem of determining if a student's effort and conduct will improve if he is given the opportunity to self-evaluate these phases of his education. Very few studies have been done on self-evaluation. New techniques for securing student self-ratings were used, such as weekly evaluations and structured rating alternatives. Two distinct and homogeneous seventh grade classes were used. The experimental group had experience using a structured self-rating scale for two quarters. The controls did not rate themselves. The teacher rated the experimental and control groups using the same, traditional, evaluation technique. Both groups completed questionnaires as a final part of the self-evaluation experiment.

The results show: (1) a moderate relationship exists between ratings by the teacher and the student self-raters, (2) self-raters seem better able to discriminate between effort and conduct than the teacher-rater for both groups of subjects, (3) neither self- nor teacher rating of effort improved as a result of self-evaluation, (4) a close relationship between rated effort and grades exists, (5) self-evaluation seems instrumental in significantly improving

the teacher and student ratings of conduct, and (6) by using the questionnaire approach students, under self-evaluation method, differed on some opinions with other students not given the opportunity for self-evaluation. One area of agreement was that both wanted the teacher to be part of the evaluation picture. This led to the conclusion that self-evaluation may not be superficial and that it results in significant change, at least in perceived school conduct.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Asch, Morton J. "Non-directive Teaching in Psychology: an Experimental Study," Psychological Monograph, 65:1-23, 1951.

2. Duel, Henry J. "Effects of Periodic Self-Evaluation on Student Achievement," Journal of Educational Psychology, 49:197-199, August, 1958.

3. _____. "Help Your Students Help Themselves," National Association Secondary School Principals Bulletin, 43:52, October, 1959.

4. Hamalainen, Arthur E. "Existing Practices in the Evaluation of Pupil Growth in the Elementary School," The Elementary School Journal, 42:182, November, 1941.

5. Orata, Pedro T. "Evaluating Evaluation," Journal of Educational Research, 33:652, May, 1940.

6. Powell, Margaret G. "Comparisons of Self-Rating, Peer-Ratings and Experts-Ratings of Personality Adjustment," Educational and Psychological Measurement, 8:225-234, 1948.

7. Rogers, Carl R. Client-Centered Therapy. Boston: Houghton Mifflin Company, 1951. 415-532 pp.

8. Russell, David. "What Does Research Say About Self-Evaluation," Journal of Educational Research, 46:570, April, 1953.

9. Symonds, Percival M. "Pupil Evaluation and Self-Evaluation," Teachers College Record, 54:138, December, 1952.

10. Taylor, Donald M. "Changes in the Self-Concept Without Psychotherapy," Journal of Consulting Psychology, 19:205-208, June, 1955.

11. Tschechtelin, Sister M. Amatora. "Self-Appraisal of Children," Journal of Educational Research, 39:25-32, September, 1945.

12. Ullmann, Charles A. "Identification of Maladjusted School Children," Public Health Monograph, 7:1-36, 1952.

13. Underwood, Benton J. Elementary Statistics. New York: Appleton-Century-Crofts, Inc., 1954. 174 pp.

14. Webb, Wilse B. "Self-Evaluation Compared with Group Evaluation," Journal of Consulting Psychology, 16:305-307, June, 1952.

15. Wickman, E. K. Children's Behavior and Teachers' Attitudes. New York: The Commonwealth Fund, 1928. 122 pp.

APPENDICES

APPENDIX A: SELF-RATING SCALE

Name _____ Grade _____ Teacher _____

Directions:

Study each of the four effort and conduct problems. In each problem choose the one statement that fits you. Place the letter (A, B, C, D, or X) of the statement that fits you in the square for the week you are making this evaluation.

NOTE: Look at the first column on the right side marked sample. Be sure to mark only one letter for each problem. Each of the four must be marked in each time you are asked to evaluate.

	Sample	1st Week	2nd Week	3rd Week	4th Week	5th Week	6th Week	7th Week	8th Week	9th Week
1. Including school and home study, this week I studied:										
A. Three or more hours daily										
B. Approximately two hours daily										
C. Approximately one hour daily										
D. Approximately one-half hour daily										
X. Less than fifteen minutes daily										
2. Daily assignments given me by the teacher this week were:										
A. Always turned in (100%) and completed (100%) on time										
B. Always turned in (100%) but---not always completed										
C. Most always completed. (better than three fourths)										
D. Half completed and turned in										
X. Very seldom turned in										

APPENDIX A: SELF-RATING SCALE
(continued)

	Sample	1st Week	2nd Week	3rd Week	4th Week	5th Week	6th Week	7th Week	8th Week	9th Week
3. Teachers must speak to me about my talking, leaving my desk, and/or disturbing others:										
A. Never this week										
B. Not more than once this week										
C. Not more than three (3) times this week										
D. Not more than five (5) times this week										
X. More than six times (6) this week										
4. Teachers and the principal must speak to me about my behavior in the halls, on the school bus and in the playground:										
A. Never this week										
B. Not more than once this week										
C. Not more than three (3) times this week										
D. Not more than five (5) times this week										
X. More than six (6) times this week										

APPENDIX B: RESULTS OF THE QUALITATIVE QUESTIONNAIRE GIVEN TO THE TEACHER-RATER

1. Do you feel the students understood what they were doing during self-evaluation? That is did they sense the power that was theirs?

   "Yes, very definitely."

2. Did they understand the directions of the self-rating scale at the start of the experiment?

   "Yes, but some needed a second explanation." (How many?) "Four."

3. How long did the evaluation take each week?

   "Ten minutes."

4. Were they all really serious about it or did they joke and look at each other during the rating period?

   "Really very serious."

5. Did you get any pros or cons concerning self-evaluation from any parents?

   "Not one parent contacted me."

6. Was there a marked improvement in effort and/or conduct the second semester? I want your global opinion of the group as a whole.

   "Too many variables."

7. How many students were actually helped by this method? That is, you are sure this method was the reason for their gain?

   "Six people were actually benefited."

8. Would you like to see it used throughout the junior high?

   "Yes, it makes the subjective part of grading more objective."

(continued)

9. Did you find that you had less trouble from this group getting them to work for better grades and citizenship than you did with your other classes?

"The teacher variable is too major, although, you get the feeling that the teacher is more objective and the student understands what to do. Certainly, I would say that they were more responsible."

10. How could this method be improved?

"With closer supervision and teacher consultation with those too far out."