

University of Montana

## ScholarWorks at University of Montana

---

Graduate Student Theses, Dissertations, &  
Professional Papers

Graduate School

---

1978

### Attitudes of group referenced oriented high school students toward task referenced evaluation versus group referenced evaluation and the influence of test grades

Patrick J. Stuber  
*The University of Montana*

Follow this and additional works at: <https://scholarworks.umt.edu/etd>

**Let us know how access to this document benefits you.**

---

#### Recommended Citation

Stuber, Patrick J., "Attitudes of group referenced oriented high school students toward task referenced evaluation versus group referenced evaluation and the influence of test grades" (1978). *Graduate Student Theses, Dissertations, & Professional Papers*. 6015.  
<https://scholarworks.umt.edu/etd/6015>

This Professional Paper is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact [scholarworks@mso.umt.edu](mailto:scholarworks@mso.umt.edu).

ATTITUDES OF GROUP REFERENCED ORIENTED HIGH SCHOOL  
STUDENTS TOWARD TASK REFERENCED EVALUATION VERSUS GROUP  
REFERENCED EVALUATION AND THE INFLUENCE OF TEST GRADES

By

Patrick J. Stuber

B.S., Eastern Montana College, 1976

Professional Paper

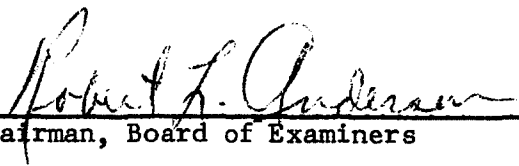
Presented in partial fulfillment of the requirements for the degree of

Master of Education

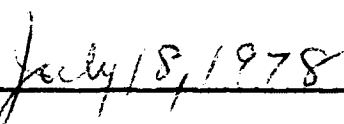
UNIVERSITY OF MONTANA

1978

Approved by:

  
Chairman, Board of Examiners

  
Dean, Graduate School

  
Date

UMI Number: EP36816

All rights reserved

**INFORMATION TO ALL USERS**

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI EP36816

Published by ProQuest LLC (2013). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 - 1346

TABLE OF CONTENTS

	Page
LIST OF TABLES . . . . .	iii
LIST OF FIGURES . . . . .	iv
CHAPTER 1	
INTRODUCTION . . . . .	1
STATEMENT OF THE PROBLEM . . . . .	1
RATIONALE . . . . .	2
DELIMITATIONS OF THE PROBLEM . . . . .	2
DEFINITIONS OF TERMS . . . . .	3
TASKED REFERENCED EVALUATION . . . . .	3
GROUP REFERENCED EVALUATION . . . . .	3
STUDENT PREFERENCE . . . . .	3
HYPOTHESES . . . . .	4
CHAPTER 2	
SURVEY OF RELATED LITERATURE . . . . .	6
GRE NEGATIVISM . . . . .	6
EVALUATION ALTERNATIVES . . . . .	7
TASK REFERENCED EVALUATION . . . . .	8
CHAPTER 3	
SOURCES OF DATA, METHODS AND PROCEDURES . . . . .	10
DEPENDENT VARIABLE . . . . .	10
FACTORIAL DESIGN . . . . .	12
PROCEDURE FOR CHECKING VALIDITY AND RELIABILITY OF DATA . . . . .	12

	Page
INTERNAL VALIDITY . . . . .	12
EXTERNAL VALIDITY . . . . .	13
RELIABILITY . . . . .	14
 CHAPTER 4	
ANALYSIS OF DATA . . . . .	15
STATISTICAL FORMULA . . . . .	15
HYPOTHESIS ONE . . . . .	15
HYPOTHESIS TWO . . . . .	18
HYPOTHESIS THREE . . . . .	19
HYPOTHESIS FOUR . . . . .	20
HYPOTHESIS FIVE . . . . .	21
 CHAPTER 5	
CONCLUSIONS . . . . .	23
SUMMARY . . . . .	23
IMPLICATIONS . . . . .	24
 APPENDIXES	
A. GRE MIDTERM REPORT . . . . .	26
B. TRE MIDTERM REPORT . . . . .	27
C. MIDTERM EVALUATION SURVEY . . . . .	28
BIBLIOGRAPHY . . . . .	29

LIST OF TABLES

Table	Page
1. Arrangement of Fifteen Mean Scores Derived from Midterm Evaluation Survey . . . . .	11
2. Mean Scores Derived from Midterm Evaluation Survey . . . . .	16
3. Mean Scores of TRE by High Letter Grade Groups . . . . .	18
4. Mean Scores of GRE by High Letter Grade Groups . . . . .	19
5. Mean Scores of TRE by Low Letter Grade Groups . . . . .	21
6. Mean Scores of TRE by Low Letter Grade Groups . . . . .	22

LIST OF CHARTS

Chart	Page
1. Results of GRE and TRE Mean Scores by Graded Groups of Subjects . . . . .	17

ATTITUDES OF GROUP REFERENCED ORIENTED HIGH SCHOOL  
STUDENTS TOWARD TASK REFERENCED EVALUATION VERSUS GROUP  
REFERENCED EVALUATION AND THE INFLUENCE OF TEST GRADES

Chapter 1

Introduction

Professional educators are aware of the discrepancies involved in evaluating student progress. There are numerous investigations that have dealt with the subjectivity of grading, and others that deal with the alternative methods of grading. There is an ongoing process to find a method that suits the parents, the pupils, and the system (job market and higher education). Task Referenced Evaluations (TRE) seem to satisfy all three groups, according to Wise and Newman, and other methods do not.<sup>1</sup> There exists however, little if any statistical research in the area of the student perception of TRE as compared to the traditional Group Referenced Evaluation system. The theoretical significance is to determine what methods work best to educate in academic endeavors, especially what motivates student behavior, conducive to optional learning.

Statement of the Problem

The purpose of this study was to determine if students discriminate between Task Referenced Evaluation (TRE, personalized reporting) and Group Referenced Evaluation (GRE, letter grading) as measured by student ratings of both evaluating systems. Specifically, this study sought to



determine if there exists a relationship between student preference for either GRE or TRE, and what grade level shows a preference for either GRE or TRE. The study sought possible answers to the following questions.

(1) Do students really care how they are evaluated? (2) Do students perceive grades as important educational tools? (3) What is the graded status of students who prefer one method of evaluation to another?

### Rationale

It is hoped that this study will answer some of the questions concerning the desirability of GRE or TRE. If feedback is an important aspect of the learning process, then the way educators provide feedback to various student types ought to be researched to provide optimum vehicles for developing maximum potential from each student. It was anticipated that this study would identify which groups of students desire various kinds of feedback from the data that are collected. This study purposely eliminated blanket endorsements for any particular type of evaluation.

### Delimitations of the Problem

The subjects for this research included 49 students of Junior standing enrolled in Forsyth High School, located at Forsyth, Montana. The subjects were divided into two groups, one group was randomly selected from two sections of a U.S. History class. The other group was the remainder of the sections. Neither group contained any handicapped nor gifted students. An examination of the grade book indicated that both groups were representative of the class as represented by grade distributions.

The subjects were given no explanation of the experiment except that a new policy for grading was being examined. All subjects were residing in or around Forsyth, Montana. All subjects had also been in the class from the beginning of the school year until the time the experiment was concluded. A major drawback is that the study does not represent a broad base of the U.S. population.

#### Definitions of Terms

Tasked Referenced Evaluation. TRE is a format that consists of two mastery related items. The student is either evaluated as (a) has completed all tasks as required by class goals, or (b) progressing toward class goals.

Group Referenced Evaluation. GRE is the traditional method of assigning grades ABCDF to students. The grade is based on an ordinal and/or interval rank of ability, competency, or other measurable activity according to teacher standards, towit; A is excellent, B is above average, C is average, D is below average and F is failing.

Student Preference. Student preference in this study contends that students are capable of interpreting graded systems. Student preference refers to a person's ability to perceive a given grading system, and rate it somewhere on a continuum scale between one and ten.

### Hypotheses

It is hypothesized that:

(1) Student preference for TRE is directly related to the usual graded evaluation of the student by GRE methods, and moreover that:

(2) High letter grades and TRE ratings are negatively correlated; the higher the letter grade, the lower the TRE rating.

(3) High letter grades and GRE ratings are positively correlated; the higher the letter grade, the higher the GRE rating.

(4) Low letter grades and TRE ratings are negatively correlated; the lower the letter grade, the higher the TRE rating.

(5) Low letter grades and GRE ratings are positively correlated; the lower the letter grade, the lower the GRE rating.

Previous research by Rinnie indicated that high grades make some people achievers and that "they need standards and feel cheated if they aren't given feedback."<sup>2</sup> Longstreet, however, states that "Grades hinder self-initiating, intellectual, and creative behavior. In the elementary school poor grades set the stage for failure,"<sup>3</sup> and that

"Revision of grading may improve the quality of evaluation feedback, but it cannot alleviate the exaggerated dependency upon grades foisted on pupils when they are too young to defend themselves."<sup>4</sup>

A grade of D or F in American education is a visible sign of failure and those students in such a letter graded category would rather be measured by some other means, as offered by TRE. TRE does not easily relate to norms however, and it might be inferred that so-called A and B students are A and B students because they are motivated to be A and B students, and they might not relate to TRE measurements. The hypothesis does not

state that students in a low achievement status as stated in TRE terms will not eventually come to recognize TRE as still a method of ranking them as inferior, and that only the language has changed. It might be reasoned that no new evaluating system can compete effectively with the present ABCDF grading system. Students dislike being labeled below average (D) or as a failure (F) when they possess a positive feeling about themselves. Therefore, they may regard a report of "is progressing" more desirable than a grade of D or F which may translate into "close to failing or has failed."

A and B grades provide feedback that are both intrinsically and extrinsically reinforcing. A TRE does not rank students within a group, but refers to whether or not a student has completed a task(s). Since TRE feedback is not norm referenced, high grade level students may find TRE less desirable.

---

<sup>1</sup>R. Wise and B. Newman, "The Responsibilities of Grading," Educational Leadership, XVIII (January, 1975), 253-256.

<sup>2</sup>C. Rinnie, "Grading and Growth: Answer to an Editorial," Educational Leadership, XVIII (January, 1975), 247-249.

<sup>3</sup>W. Longstreet, "The Grading Syndrome," Educational Leadership, XVIII (January, 1975), 243-246.

<sup>4</sup>Ibid.

## Chapter 2

### Survey of Related Literature

GRE Desirability. There is adequate information available stressing the desirability of GRE. Rinnie suggests that parents and students need the ABCDF yardstick to measure progress or status within a group.<sup>1</sup> Rinnie also points out that "parents don't want the conference mumbo-jumbo, students feel cheated if they aren't given feedback (A=great, B=good, C=ok, and D=not so hot),"<sup>2</sup> and "if we did not have a grading system with grades we'd have one with numbers."<sup>3</sup> Power indicated that high grades in particular, give students "access to social, academic, and extracurricular activities, and transition to employment, law or medical school, graduate school, etc. will greatly be facilitated by high grades."<sup>4</sup>

GRE Negativism. Researchers have found grading by ABCDF to be victimizing. In a 1975 study, Poole found that "95% of secondary students feel they were victimized by testing and evaluation."<sup>5</sup> The technical inadequacy of the test and subsequent grades did not reflect the true essence of the student. Meyer, in 1976, reported hints of grade inflation whereas "A is ok, nothing else is, and so at sometime you are bound to be evaluated not ok."<sup>6</sup> Meyer is also resentful of GRE because it "is contradictory to the theory that all people or students are not alike,"<sup>7</sup> and that "while objectives are edifying, they are not measureable as defined."<sup>8</sup> Also in 1976, Michaels noted that "when students are sorted on a curve,

no one can tell how well the teacher taught or the student learned,"<sup>9</sup>  
and further, that

"the effort necessary to achieve high performance varies inversely with ability, therefore, those who need to try hardest are given the least incentive to do so. Obviously, using a normal curve grading system, high grades are reinforcing only for those who get them."<sup>10</sup>

Longstreet found grades to be narrow and the GRE "hinders self initiating, intellectual and creative behavior."<sup>11</sup> The heaviest attack on GRE was conducted by Power in 1976 when she referred to GRE as a syndrome. Power's research indicated that

"professors tend to give higher grades in the spring time to females, and to smaller classes. Grades are often incorporated to include subjective material such as dress, promptness, and tardiness. Grades are not used in a comparable manner from school to school, department to department, or professor to professor. All is not well with the curve method either; students are paid to fail to insure higher grades for others, and they produce excessive competition, negativeness, and dishonesty. Grading tends to foster distorted educational values which make the appearance rather than the substance of learning the motivating factor."<sup>12</sup>

Evaluation Alternatives. Finding more meaningful methods of evaluating students is a contemporary issue that is being approached from several angles. Leary, in 1975, gave recognition to the problem of GRE and published a set of "guidelines for switching over to a new grading system."<sup>13</sup> Leary did not speculate what the new system ought to be, however. Rogers advocated "levels grading".<sup>14</sup> The difficulty with Roger's method was that it still used a GRE system, and moreover, those on the lower levels knew that their GRE was not equivalent to higher level students. Bornschuer's "grade contracting"<sup>15</sup> did not get away from GRE entirely. By contracting, the student determined where she/he fit in the group and worked to prove it, thus removing the placement burden from the educator,

provided the contract was completed. Ladas, in 1974, urged "competency grading",<sup>16</sup> but denounced grading in general stating, "What is a grade anyway? Each instructor measures today, using his own foot as a ruler."<sup>17</sup> Simon, Kirschenbaum, and Napier, in their book Whad-ja-get?, described several alternatives to GRE such as written evaluations, performance curriculum, pass-fail, and blanket grading.<sup>18</sup>

Task Referenced Evaluation. Wise and Newman developed a TRE which they described as "meeting the responsibility to both parent and child. How well a student has mastered a particular task is reported."<sup>19</sup> Another method, "Personalized Achievement Reporting"<sup>20</sup> as designed by Hansen is TRE related. Meyer also stressed that grading should be either "has achieved or learning in progress."<sup>21</sup> Research concerning the desirability of task referenced evaluations does not seem to exist in the literature.

---

<sup>1</sup>C. Rinnie, "Grading and Growth; Answer to an Editorial," Educational Leadership, XVIII (January, 1975), 247-249.

<sup>2</sup>Ibid.

<sup>3</sup>Ibid.

<sup>4</sup>Marian Power, "The Grading Syndrome," Journal of Reading, XIX (April, 1976), 568.

<sup>5</sup>A. Poole, "A Teacher-Pupil Dilemma: Student Evaluation and Vicimization," Adolescence, XI (Fall, 1976), 341.

<sup>6</sup>Mary Coeli Meyer, "A New Look at Grading Systems," School and Community, LXII (March, 1975), 26.

<sup>7</sup>Ibid.

<sup>8</sup>Ibid.

<sup>9</sup>James Michaels, "A Simple View of the Grading System," Teaching Sociology, IV (January, 1976), 198-203.

<sup>10</sup>Ibid.

<sup>11</sup>W. Longstreet, "The Grading Syndrome," Educational Leadership,

XVII (January, 1975), 243-246.

<sup>12</sup>Power, 568.

<sup>13</sup>J. Leary, "Assessing Pupil Progress: New Methods are Emerging," Educational Leadership, XVIII (October, 1975), 25-27.

<sup>14</sup>Erleen J. Rogers, "Meeting Student Needs Through the Levels Program and Grade Weighting," The Clearinghouse, LVI (January, 1976), 217-220.

<sup>15</sup>Joan H. Bornschuer, "The Grade Contract Revisited," Foreign Language Annals, IX (April, 1976), 166.

<sup>16</sup>Harold Ladas, "Grades: Standardizing the Unstandardized Standard," Phi Delta Kappan, LVI (February, 1974), 200.

<sup>17</sup>Ibid.

<sup>18</sup>S. B. Simon, H. Kirschenbaum, and R. Napier, Whad-ja-get?, (New York: Hart Publishing Co., 1971).

<sup>19</sup>R. Wise and B. Newman, "The Responsibilities of Grading," Educational Leadership, XVIII (January, 1975), 253-256.

<sup>20</sup>J. Merrell Hansen, "Personalized Achievement Reporting: Grades That are Significant," The High School Journal, LXI (March, 1977), 255.

<sup>21</sup>Meyer, 26.



## Chapter 3

### Sources of Data, Methods and Procedures

Independent variables. There were two independent variables. One variable consisted of a random group of Junior status students from Forsyth, Montana who were subjected to GRE treatment. This group was the control group ( $Ss_1$ ). The second variable, or experimental group ( $Ss_2$ ), consisted of Junior status students also from Forsyth, Montana. There were 25 subjects in the experimental group and 24 subjects in the control group. The experiment was carried out during the fourth quarter of the 1977-78 school year, and all subjects had been exposed to a GRE midterm on three prior occasions. In the fourth quarter of the academic school year, the instructor gave the control group the usual midterm report (see Appendix A, p. 26 and Appendix B, p. 27) with letter grades ABCDF, and the experimental group received TRE statements. The experimental group differed in that all subjects, who would normally have received an A, B, or C grade had "has completed class goals" written on their midterm reports, and the D or F students had "progressing towards class goals" written on their forms.

Dependent variable. Student rating of the TRE and GRE is determined by having the subjects fill out a "Midterm Evaluation Survey" (Appendix C, p. 28). The subjects were to indicate (a) what group they were in, (b) what their usual class grade is, (c) how they rate GRE on a scale of one to ten, and (d) how they rate TRE on a scale of one to ten if they

had been given a TRE Statement. Once the surveys had been returned, the data were broken down into fifteen mean scores for each respective group as illustrated on Table 1.

Table 1  
Arrangement of Fifteen Mean Scores Derived  
from Midterm Evaluation Survey

Student's Letter Grade	Column I GRE Rating by Ss <sub>1</sub> *	Column II TRE Rating by Ss <sub>2</sub> **	Column III GRE Rating by Ss <sub>2</sub> **
A	1	6	11
B	2	7	12
C	3	8	13
D	4	9	14
F	5	10	15

\*Control Group

\*\*Experimental Group

Table 1 illustrates how fifteen mean scores derived from this survey instrument are arranged for comparative purposes according to letter grade. Students surveyed were asked what kind of grade they usually received, and this was used to categorize mean perceptual scores of TRE and GRE. Scores 1-5 in Column I represent how mean values of the control group will be plotted. For example, item 1 is the mean score of "A" students when asked on a scale of 1-10 how they liked GRE. Item 2 is the mean score of "B" students, etc. Column II is the experimental group's mean perceptual rating of TRE. Item 6 is the mean response by "A" stu-

dents, and item 7 is the response of "B" students, etc. Column III is composed of the mean ratings of the experimental group's perception of GRE. Item 11 represents an "A" student's perception of GRE after TRE treatment, item 12 represents a "B" student, etc.

### Factorial Design

In analyzing the data of this study, a factorial design was utilized. This design can be diagrammed as follows:

$Ss_1$ Control	$R_1$		$O_1$
$Ss_2$ Experimental	$R_2$	X	$O_2$

R = a group of subjects

X = a TRE treatment

O = observation

The mean scores were compared in each factor as an indication of student preference. Higher mean values indicate higher student preference for an evaluation as compared to lower mean values.

### Procedure for Checking Validity and Reliability of Data

Internal Validity. Several steps were taken to validate the experiment. Historically both the experimental and control groups were classmates in classes with the same instructional methods, and in the same discipline (U.S. History). The subjects were randomly selected to be either a part of the control group or the experimental group. The process of natural development was the same in both groups, basically because of the short time period of the experiment, and was not a test intervener. No pretest nor discussion about the survey took place. Both groups had

approximately an equal distribution of students in each of the five graded categories (ABCDF).

The survey form and manner of data collection remained consistent. The measuring instrument and data collection remained constant over time and was consistent across groups. All student responses were included in the final data. The experiment suffered no experimental mortality during its course. Spearman rank-order correlations were utilized to test the correlations as hypothesized. Since exact correlations were hypothesized, a visual scan of the data would indicate if the rankings were exact or not. Critical values of the correlations were not sought because of the small number (5) of groups being ranked in each independent variable. An explanation of the Spearman rank-order correlation appears in Chapter 4.

External Validity. Precautions were taken to avoid bias. There was no pretest and although students may have guessed at the intent of the survey, it is likely that there were many different guesses. One major weakness of the experiment is that the sample does not represent the broadest population possible. It supplies data pertinent to students of one particular rural area, but that data may or may not be relevant to subjects of other areas. The data may represent helpful guidelines which indicates what the student population, nation-wide, may perceive. There were no visible reactive effects of the experimental arrangements, no "Hawthorne effect" largely because neither group knew whether they were the control group or the experimental group. The experimental group was aware of something different, but they did not know why or for what purpose they were treated differently. They did not know how long they would be treated differently either, as the experimenter told them that

that they would not get a GRE that midterm. They were not told otherwise, until after the survey form was completed. This was not a laboratory experiment and the subjects did go home after they were given the TRE. Since they did not take the survey until the following day they were exposed to multiple interference. There is no evidence of any outside activity that influenced the subjects' responses to the survey given on the following day.

Reliability. In assessing the reliability of the survey instrument it is known that none of the subjects had seen the form beforehand, it required little memorization, practice, nor experience, in taking the survey. Neither group had specific knowledge that would give it an advantage over the other group. The survey measured affective and not cognitive domains. Other factors such as fatigue, emotional strain, physical conditions of the room, and health of the subjects were not self evident as being an interveing influence on the experiment.

## Chapter 4

### Analysis of Data

Statistical Formula. The following statistic (Spearman rank-order correlation) was used in testing the hypothesis.

$$r_s = 1 - \frac{6 \sum d^2}{N^3 - N}$$

Whereas:

- $r_s$  is the correlation (Spearman method)
- $d$  is the difference between the two ranks (grades and means) squared
- $\sum$  is the Greek letter meaning sum
- $N$  is the number of rank sets (one grade and one mean equals one set).

### Hypothesis One.

Student evaluation for TRE is directly related to the usual graded evaluation of the student by GRE methods.

Subjects who were in the high letter grade groups gave GRE a higher mean rating than they did for TRE. Subjects who were in the low letter grade groups gave TRE a higher mean rating than they did for GRE. The positive correlation expected throughout the GRE group is very high among the mean scores ( $r_s = .986$ ). The negative correlation expected throughout the TRE rating did not develop ( $r_s = .6$ ). (See Table 2, p. 16 and Chart 1, p. 17). Therefore, Hypothesis One is rejected.

The control and experimental groups of A, B, and C students rated letter grading higher than task referenced marks. D students in both groups rated task referenced marks higher than letter grades. F students also ranked task referenced marks higher than letter grades, but did not rank either evaluation method high. D students rated TRE about as high (8.0) as A and B students rated letter grades (8.0, 8.0, 9.0, 7.64, 7.66).

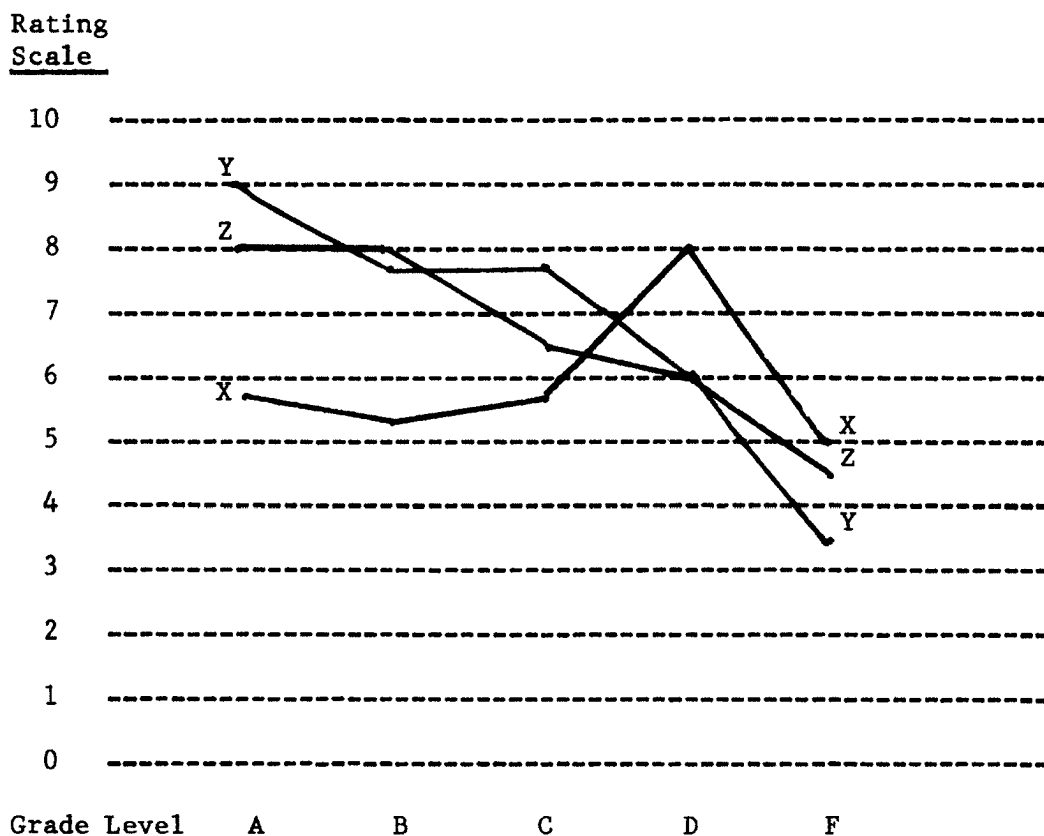
Table 2  
Mean Scores Derived from Midterm Evaluation Survey

Student's Letter Grade	Column I GRE Rating by Ss <sub>1</sub>	Column II TRE Rating by Ss <sub>2</sub>	Column III TRE Rating by Ss <sub>2</sub>
A	8.0	5.75	9.0
B	8.0	5.285	7.64
C	6.46	5.66	7.66
D	6.0	8.0	6.0
F	4.5	5.0	3.5

Table 2 plots the mean scores of student perceptions regarding group referenced evaluation and task referenced evaluation. The control group in Column I correlated very high ( $r_g = .986$ ) by grade group and rating. The experimental group rated task referenced evaluation lower at the higher grades, and higher at the lower grades as seen in Column II as compared to Column I and III. It is noted that Column III contains the extreme scores. The A students' rating was very good (9.0) after being treated to TRE, and F students' rating is the lowest on the table (3.5).

Chart 1

Chart 1 graphically displays the information from Table 2 to show the scope and differences of the three columns of data.



Results of GRE and TRE Mean Scores by Graded Groups of Subjects

X = Experimental Group TRE

Y = Experimental Group GRE

Z = Control Group GRE



Hypothesis Two.

High letter grades and TRE ratings are negatively correlated; the higher the letter grade, the lower the TRE rating.

Hypothesis two rationalizes that students who get high grades would want to retain that system. By comparison, a task referenced evaluation does not give the elite kind of recognition that an A or B does, and that TRE should rate low in A and B groups.

Mean scores for TRE are not negatively correlated to high letter grades ( $r_s = .5$ ). The exact opposite is nearly the case. (See Table 3). There exists a near positive correlation among the high grades with A students giving TRE a 5.75 rating; higher than B's and C's, therefore Hypothesis Two is rejected.

Table 3

Mean Scores of TRE by High  
Letter Grade Groups

Student's Letter Grade	Column I GRE Rating by $Ss_1$	Column II TRE Rating by $Ss_2$	Column III GRE Rating by $Ss_2$
A		5.75	
B		5.285	
C		5.66	
D			
F			

Table 3 illustrates how TRE was rated in terms of mean scores by A, B, and C students. High letter grade students did not give particularly

high scores for the TRE system. The expectation that as the grade got lower, the mean score would increase do not develop.  $r_s = .5$  for this portion of the study.

Hypothesis Three.

High letter grades and GRE ratings are positively correlated; the higher the letter grade, the higher the GRE rating.

Being labeled A or B is a positive aspect in American education. It was expected that students who get A's and B's would rate a letter grade system high. It was further hypothesized that if those students, who get A's and B's, were denied those high marks there would be evidence to show that they rate A and B (GRE rating) over other methods.

Table 4  
Mean Scores of GRE by High Letter Grade Groups

Student's Letter Grade	Column I GRE Rating by $Ss_1$	Column II TRE Rating by $Ss_2$	Column III GRE Rating by $Ss_2$
A	8.0		9.0
B	8.0		7.64
C	6.46		7.66
D			
F			

There are two sets of data to consider in hypothesis three. One set of data is the mean scores of the control group. The control group rated A's and B's high (8.0, 8.0), and their rating was higher than C's

(6.46) for a  $r_s = .875$ . The experimental group rated GRE (letter grade) high (9.0, 7.64) but both were not higher than C's (7.66) for a  $r_s = .50$ . The experimental A students rated GRE higher than their control counterparts (9.0, 8.0), as did C students (7.66, 6.46). Both sets of data are higher than TRE ratings by  $Ss_2$ . Hypothesis Three is accepted.

Table 4 (p. 19) illustrates how GRE was rated in terms of mean scores by A, B and C students. The expectation that the mean score was correlated to grade is substantial, but not positively correlated ( $Ss_1$ ,  $r_s = .875$ ,  $Ss_2$ ,  $r_s = .5$ ). As the grade got higher, the mean score did not positively go higher in each increment.

#### Hypothesis Four.

Low letter grades and TRE ratings are negatively correlated; the lower the letter grade, the higher the TRE rating.

If a D or F appears harsh to the recipients, "is progressing" ought to be perceived as a less threatening statement. F students would then rate TRE very high and higher than D students.

Subjects' mean scores for TRE are not negatively correlated to low letter grades ( $r_s = .5$ ). Mean TRE scores did rate higher for this group compared to mean GRE scores for the same  $Ss$ . The notable group here are the D students who rate TRE 8.0. This is a good rating for any grade system and is exceeded in the ratings only by experimental group ratings of GRE by A students. F students also like TRE more than GRE (5.0 to 3.5 and 4.5). The hypothesized correlation may be poor when one considers F students probable would not rate any system high if they perceive themselves to be "F" students.

Table 5 (p. 21) illustrates how TRE was rated in terms of mean

scores by C, D, and F. students. The expectation that as the grade got lower the rating of TRE would get higher did not develop. Notice the high score by D students for example. The anticipated negative correlation was instead  $r_s = .5$ . F students did rate TRE higher than GRE but the mean score was 5.0, which is only higher than F students' ratings of GRE (3.5, 4.5). Therefore, Hypothesis Four is rejected.

Table 5

## Mean Scores of TRE by Low Letter Grade Groups

Student's Letter Grade	Column I GRE Rating by $Ss_1$	Column II TRE Rating by $Ss_2$	Column III GRE Rating by $Ss_2$
A			
B			
C		5.66	
D		8.0	
F		5.0	

Hypothesis Five.

Low letter grades and GRE ratings are positively correlated; the lower the letter grade, the lower the GRE rating.

It is recognized that students who receive low grades resent the system that labels them failures. The lower a student gets evaluated, the lower is the student's perception of such a system, as expected. Low letter grades and GRE are positively correlated ( $r_s = 1.0$ ). Lower graded students gave lower ratings to GRE. Hypothesis Five is accepted.

Table 6 illustrates how GRE was rated in terms of mean scores by C, D, and F students. It is noteworthy that both D groups had the same mean. F students who had been exposed to the less harsh TRE ( $Ss_2$ ) rated GRE the lowest of all GRE ratings. F students in the treatment group were more convinced that GRE was a poor system than were the control group students by a 3.5 to 4.5 margin.

Table 6  
Mean Scores of GRE by Low Letter Grade Groups

Student's Letter Grade	Column I GRE Rating by $Ss_1$	Column II TRE Rating by $Ss_2$	Column III GRE Rating by $Ss_2$
A			
B			
C	6.46		7.66
D	6.0		6.0
F	4.5		3.5

## Chapter 5

### Conclusions

The data collected for this study lead to the following conclusions:

1. Students in the F letter grade group rate evaluations low, irregardless of the wording of the evaluation format.
2. Task Referenced Evaluation appears to appeal to a very narrow range of graded pupils (D groups only).
3. The majority of students relate better to an ABCDF grading system, especially those who have received high grades in terms of how they evaluated Grouped Referenced Evaluation.

### Summary

This research lends support to the traditional method of grading as opposed to Task Referenced Evaluation. In and of itself, the study does not find Group Referenced Evaluation as the best method for all students. One fact is that Group Referenced Evaluation is more understandable to parents, students, and the establishment (i.e. schools of higher learning and the job market). Rinnie does indicate that any other method of evaluating is not soon forthcoming as an acceptable alternative.<sup>1</sup>

A logical extension of this study would be to compare the results of this study to investigations of student desirability of grading, per se. Longstreet and others have discussed the probability that grades are

self-defeating in the educational system for many students.<sup>2</sup> Studies may show that task referenced evaluation, group referenced evaluation, and self referenced evaluation are equally as poor in the educational system for student evaluation and that students could benefit by allowing the world of work and higher education to review their progress by some other means. This same research could be further studied by using experimental groups located in an urban school.

### Implications

The researcher's overall implications are as indicated by the following statement" A rose by any other color is still a rose. Students can recognize success and failure no matter how the evaluating report is masked. Remarks made to the researcher after the data were turned in were as follows: "What's the difference? You are going to find out if you are a failure sometime." In general the subjects rejected the idea of new evaluation methods. Given the choice some evaluators excluded letter grades and thus support Holt's approach:

"Any evaluation that is used not as a personal matter between the learner and someone trying to help him learn, but is given instead to grade and label students for someone else's purpose (colleges, employers, and anxious parents) are illegitimate and harmful."<sup>3</sup>

As students leave school they ought to be given an equal opportunity at college and work by the standards each passes for acceptance, excluding scholastic performance.

---

<sup>1</sup>C. Rinnie, "Grading and Growth: Answer to an Editorial," Educational Leadership, XVIII (January, 1975), 247-249.

<sup>2</sup>J. Longstreet, "The Grading Syndrome," Educational Leadership,

XVIII (January, 1975), 243-246.

<sup>3</sup>J. Holt, "Personalized Achievement Reporting: Grades That are Significant," The High School Journal, LI (December, 1976), 44-46.



## APPENDIX A

## GRE MIDTERM REPORT

FORSYTH SENIOR HIGH SCHOOL

## MIDTERM REPORT

Name \_\_\_\_\_

Grade    9    10    11    12

This report will indicate to the student and his parents the level of achievement that the student has accomplished for the first half of the 4<sup>th</sup> quarter in the 1977-1978 school year.

	AREA OF STUDY	GRADE AND COMMENTS	TEACHER INITIALS
1.			
2.			
3.			
4.	U.S. HISTORY	A-	R.S.
5.			
6.			

## APPENDIX B

## TRE MIDTERM REPORT

## FORSYTH SENIOR HIGH SCHOOL

## MIDTERM REPORT

Name \_\_\_\_\_

Grade    9    10    11    12

This report will indicate to the student and his parents the level of achievement that the student has accomplished for the first half of the 4<sup>th</sup> quarter in the 1977-1978 school year.

	AREA OF STUDY	GRADE AND COMMENTS	TEACHER INITIALS
1.			
2.			
3.			
4.	U.S. History	Completed Class Goals	R.S.
5.			
6.			

## APPENDIX C

## MIDTERM EVALUATION SURVEY

1. On my midterm I was given:

- a) a letter grade
- b) a task referenced grade

2. Usually in this class my grade is:

- a) A
- b) B
- c) C
- d) D
- e) F

3. I would rate the letter grade system on a scale of one (poor) to ten (excellent) as:

1   2   3   4   5   6   7   8   9   10

4. I would rate the task referenced system on a scale of one (poor) to ten (excellent) as:

1   2   3   4   5   6   7   8   9   10

Do not respond to the above item if you did not receive a task referenced evaluation.

THANK YOU FOR YOUR TIME

## BIBLIOGRAPHY

- Bornscher, Joan H. "The Grade Contract Revisited," Foreign Language Annals, IX (April, 1976), 166.
- Campbell, P. "Add a New Dimension to Your Grading System," American Vocational Journal, LI (December, 1976), 44-46.
- Hansen, J. Merrell. "Personalized Achievement Reporting: Grades That are Significant," The High School Journal, LXI (March, 1977), 255.
- Holt, J. "Personalized Achievement Reporting: Grades That are Significant," The High School Journal, LXI (March, 1977), 255.
- Ladas, Harold. "Grades: standardizing the Unstandardized Standard," Phi Delta Kappan, LVI (February, 1974), 200.
- Leary, J. "Assessing Pupil Progress: New Methods are Emerging," Educational Leadership, XVIII (October, 1975), 25-27.
- Longstreet, W. "The Grading Syndrome," Educational Leadership, XVIII (January, 1975), 243-246.
- Meyer, Mary Coeli. "A New Look at Grading Systems," School and Community, LXII (March, 1975), 26.
- Michaels, James. "A Simple View of the Grading System," Teaching Sociology, IV (January, 1976), 198-203.
- Newman, B. and R. Wise. "The Responsibilities of Grading," Educational Leadership, XVIII (January, 1975), 253-256.
- Poole, A. "A Teacher-Pupil Dilemma: Student Evaluation and Victimization," Adolescence, XI (Fall, 1976), 341.
- Power, Marian. "The Grading Syndrome," Journal of Reading, XIX (April, 1976), 568.

Rinnie, C. "Grading and Growth: Answer to an Editorial," Educational Leadership, XVIII (January, 1975), 247-249.

Rogers, Erleen J. "Meeting Student Needs Through the Levels Program and Grade Weighting," The Clearinghouse, LVI (January, 1976), 217-220.

Simon, S. B., H. Kirschenbaum, and R. Napier. Whad-ja-get? New York: Hart Publishing Co., 1971.

Simon, Sidney B. "Personalized Achievement Reporting: Grades That are Significant," The High School Journal, LXI (March, 1977), 255.

Walling, D. "Designing a Report Card That Communicates," Educational Leadership, XVIII (January, 1975), 256-258.

Wise, R., and B. Newman. "The Responsibilities of Grading," Educational Leadership, XVIII (January, 1975), 253-256.