A STUDY OF THE INFLUENCE OF CERTAIN SELECTED FACTORS ON THE RATINGS OF SPEECH PERFORMANCES

Thesis for the Degree of Ed. D.

MICHIGAN STATE COLLEGE

Emil R. Pfister

1955

This is to certify that the

thesis entitled

A Study of the Influence of Certain

Selected Factors on the Ratings of Speech Performances

presented by

Emil R. Pfister

has been accepted towards fulfillment
of the requirements for

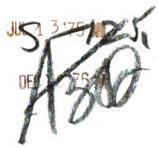
Ed.D. degree in Higher Education

Milo C. Mentzer
Major professor

Date May 17, 1955

O-169

X.V.



A STUDY OF THE INFLUENCE OF CERTAIN SELECTED FACTORS ON THE RATINGS OF SPEECH PERFORMANCES

By
Emil R. Pfister

A THESIS

Submitted to the School of Graduate Studies of Michigan State College of Agriculture and Applied Science in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

School of Education

1955

ACKNOWLEDGMENTS

The writer wishes to express his appreciation to the many people who helped him throughout the development of this study.

The encouragement and guidance came primarily from Dr. Milosh Muntyan, Dr. Wilson B. Paul, and Dr. Harry W. Sundwall. Their cooperation, sympathetic attitude, and helpful suggestions are deeply appreciated.

Sincere appreciation also should be given to the juniors, seniors and faculty of the Department of Speech and Drama, Central Michigan College of Education, who cooperated in the collection of data during the 1952-53 academic year.

Furthermore, acknowledgment is given to those of the staff of Michigan State College and the writer's colleagues at Central Michigan College of Education, especially Dr. Wilbur Moore and Dr. Karl Pratt, who were kind enough to answer questions and offer suggestions during the progress of the investigation.

Invaluable help in setting the data up so that it could be tabulated by IBM equipment was given the writer by Dr. Willard Warrington of the Board of Examiners and Mr. Frank Martin, Supervisor of Tabulation, Michigan State College.

This dissertation is dedicated to my wife, Frances Pfister, whose patience, confidence, and devotion have fostered the continuous concentration necessary to the successful completion of this study.

VITA

Emil Robert Pfister

candidate for the degree of

Doctor of Education

Final Examination: May 6, 1955

Thesis: A Study of the Influence of Certain Selected Factors on the Ratings of Speech Performances.

Outline of Studies: Major subject -- Higher Education
Cognate Field -- Speech

Biographical Items:

Born: January 3, 1913, Chicago, Illinois

Undergraduate Studies: Central Michigan College of Education, 1931-35

Graduate Studies: Master of Arts, University of Michigan, 1939
Columbia University, Summer, 1947
Denver University, Summer, 1948
Michigan State College, 1949-55

Experience: Speech Teacher and High School Principal, Kingston, Michigan, 1935-40
High School Principal and Director of Debate, Clare, Michigan, 1940-45
Assistant Professor of Speech, Central Michigan College of Education, Mount Pleasant, Michigan, 1945-53
Associate Professor of Speech, Central Michigan College of Education, Mount Pleasant, Michigan, 1953-

Member of: Pi Kappa Delta; Kappa Delta Pi; Speech Association of America; American Association of University Professors; National Society for the Study of Communication; American Forensic Association; National Education Association

A STUDY OF THE INFLUENCE OF CERTAIN SELECTED FACTORS ON THE RATINGS OF SPEECH PERFORMANCES

By
Emil ROBERT Pfister

AN ABSTRACT

Submitted to the School of Graduate Studies of Michigan State College of Agriculture and Applied Science in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

School of Education

1955

Approved Milosly Mentyan

Emil R. Pfister

Thesis Abstract

This study was designed to determine whether any statistically significant relationships existed between the ratings given by speech evaluators and (1) their academic speech training, (2) their acquaintanceship with the speaker, (3) their experience with the rating scale, and (4) their sex in relation to the sex of the speaker.

The five hundred and forty-nine speakers who participated in this project were freshmen enrolled in Fundamentals of Speech classes at Central Michigan College of Education during the 1952-53 academic year. The fifty-five evaluators (speech faculty members and juniors and seniors who were speech majors or minors) compiled a total of 4392 ratings. Precautions were taken and controls were employed with respect to speaker, speech, audience, and occasion with a view toward making these ratings comparable.

The Evaluator's Rating Scale devised for this study employed ten criteria based on a study of existing speech rating instruments. Appropriate tests of reliability and validity were made. All of the data obtained from these rating scales were transferred to punch cards which permitted sorting and tabulating by IBM methods.

(Thesis Abstract Continued)

The data were analyzed by appropriate procedures to discover the role played by each of the four selected factors under investigation. Differences of the means were computed for groups that were comparable in all respects except the factor being studied. The "t test" for significance of the difference of the means was applied and coefficients of correlation were computed.

The findings of this research led to the conclusion that the academic speech training of the evaluator influences his ratings. Undergraduate evaluators with majors or minors in speech gave significantly higher ratings than did evaluators with advanced degrees in speech. Furthermore, scores given by pairs of undergraduate evaluators had a higher correlation than did scores given by undergraduate-graduate pairs of evaluators. Pairs of evaluators with advanced degrees in speech had the highest correlation.

The investigation, in itself, provided inconclusive results with respect to the influence of acquaintanceship on the ratings of speech performances. However, the results of this study tend to substantiate the findings of previous research, i.e., that evaluators who are acquainted with the speakers give them higher ratings than do evaluators who are unacquainted with these speakers.

(Thesis Abstract Concluded)

In this particular study the experience of the evaluator with the rating scale employed was found to have no significant influence upon the scores given. However, all the evaluators had a certain minimum of speech training and had rated speeches previously.

The literature and data of this study support the contention that male and female evaluators rate male and female speakers differently:

(1) Female student evaluators gave higher ratings to both male and female speakers than did male student evaluators.

(2) Female student evaluators gave higher ratings to male speakers than they gave to female speakers.

(3) Male student evaluators gave higher ratings to female speakers than they gave to male speakers.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

CHAPTER I

INTRODUCTION

The speech teacher cannot escape responsibility for the evaluation of the oral performance of his students. His educational philosophy may make academic marks seem undesirable, or he may be disturbed by the influence of subjective factors which impair the validity of such evaluation. Nevertheless, the practical necessities of the learning situation, as well as customary institutional procedures, require that he evaluate the speech competency of his students.

Research has shown that students believe that rating speeches is of primary importance in a Fundamentals of Speech course. Graunke[1], who administered student judgment questionnaires to 1,024 Fundamentals of Speech students in four different universities, secured data showing that oral work was consistently judged by students to be of more value than reading assignments and written work.

---

[1]Dean F. Graunke, "The Use of Student Opinion in the Improvement of Instruction in Speech Fundamentals," (unpublished Master's thesis, The University of Nebraska, Lincoln, 1951), pp. 94-126.

Reid[2], in discussing computation of final grades for beginning speech classes, advocates giving two or three times as much weight to oral work as to written work.

Hollister[3] points out the need for good judging when he says:

> The question of judging contests is one of importance to every teacher of public speaking, for it influences his faith in contests, his spirit in classroom work, and the tone of public speaking in the school.

Evaluation of speech must of necessity involve some subjective judgments. Nevertheless, as Pelsma[4] pointed out, every attempt should be made to improve the fairness and accuracy of these judgments since they are used as a basis for instructing and guiding the student as well as determining his status.

The crucial importance of accurate evaluation of the student is illustrated by a recent regulation of Central Michigan College of Education[5] regarding the demonstration

---

[2]Leon D. Reid, Teaching Speech in High School. Columbia, Missouri: Artcraft Press, 1952, p. 191.

[3]R. D. T. Hollister, "Faculty Judging," Quarterly Journal of Public Speaking, 3:235, July, 1917.

[4]J. R. Pelsma, "Standardization of Grades in Public Speaking," Quarterly Journal of Public Speaking, 1:268, October, 1915.

[5]Bulletin, 1952-53 Sessions. Central Michigan College of Education. Mount Pleasant, Michigan, 1952, p. 74.

of speech competency by candidates for teaching certificates. The regulation prescribes completion of the course, Fundamentals of Speech (Speech 101), with achievement of at least a "C" in the course. This "C" is interpreted to mean "average" skill and facility in communicating information to a group of persons.

A. The Problem

Statement of the problem. This research is designed specifically to study critics' ratings of speech performances by students in Fundamentals of Speech classes at Central Michigan College of Education. The purpose of the study is to examine the role which certain factors play in the evaluation of speech competency.

The experiment will attempt to secure evidence in answer to four questions:

(1) Do ratings given by college juniors and seniors who are speech majors or minors differ significantly from those given by members of the speech faculty?

(2) Do ratings given by the evaluators who know a speaker differ significantly from those given by the evaluators who do not know the speaker?

(3) Do the ratings given by the evaluators who have had experience with the rating scale differ significantly from those given by the evaluators who have had no experience with the rating scale?

(4) Do the ratings of speakers of each sex differ significantly according to the sex of the evaluator?

Importance of the study. This study is of particular concern to the students and faculty of Central Michigan College of Education. The significance that speech performance ratings have for all freshmen on that campus has already been explained.

Jones[6], who studied the current practices in the beginning speech courses of 318 colleges, found that rating charts were used in the majority of them. Thus this study may be of general interest to a number of colleges in the United States.

However, it is not the purpose of this experiment to determine the relationship of these factors in all schools but rather particularly in the situation at Central Michigan College of Education. Before any adaptations to other colleges are made, one must first determine the extent to which their students and faculties are comparable to those at Central Michigan College of Education.

Naturally the characteristics of the evaluator who uses the rating scale are of primary consideration. As

---

[6]Horace Redman Jones, "The Development and Present Status of Beginning Speech Courses in the Colleges and Universities in the United States," (unpublished Doctor's dissertation, Northwestern University, Evanston, 1952), 216 pp.

Hudgins[7] points out, one way to increase the reliability of the evaluation is, of course, to reduce the variability among the evaluators. To do this, evidence must first be secured regarding the extent of influence that certain factors, other than the speech performance itself, have on the ratings.

The four questions being considered in this experiment have been selected because they have not been answered by previous research. Furthermore, they involve factors that can be controlled in the ordinary classroom situation.

The possible concrete results in terms of action may be seen by briefly considering the significance that each of the four factors might have when selecting evaluators:

(1) If ratings given by upperclassmen who are speech majors or minors tend to differ significantly from those given by members of the speech faculty, speakers who are to be compared ought to be rated exclusively by students, exclusively by faculty members, or by a like number of each. If, on the other hand, this difference in the academic speech training of the evaluators plays no significant role, one may select speech judges at random in this respect.

(2) If evaluators who know a speaker tend in general to rate him higher, or lower, than the evaluators who do not know the speaker, the teacher must be sure that either each

---

[7]Clarence V. Hudgins, "The Validity of Speech Tests," Volta Review, 45:271-2, May, 1943.

evaluator has no acquaintance with the speaker or that a like number of each speaker's acquaintances are used as evaluators. However, if acquaintanceship plays no role in this regard, then this factor need not be considered in securing evaluators.

(3) If evaluators who have had experience with the rating scale that is being used make more reliable evaluations than those who do not have such experience, teachers must be sure that all evaluators receive practice experiences. However, if experience with the rating scale makes no significant difference, mere verbal instruction may suffice; and any time spent in practice will be wasted.

(4) If evaluators rate speakers of the same sex differently than speakers of the opposite sex, this factor must be taken into consideration when securing evaluators. If, on the other hand, neither men nor women evaluators show an appreciable sex-tied favoritism in their rating of speakers, the evaluators may be secured at random without regard to sex.

This study has been predicated on the assumption that these sources of uncertainty in the evaluation of speech performance warrant careful investigation. There are, of course, other important factors such as the social background, intelligence, and physical health of the speech evaluators which have not been explored in this research.

B. Definition of Terms

Selected factors. Factors may be considered as certain characteristics of the raters. In this study the characteristics being considered are:

(1) The academic speech training of the evaluator. The ratings given by students who are in either their third or fourth year of college will be compared with the ratings given by faculty members who have advanced degrees.

(2) The acquaintanceship of the evaluator with the speaker. The ratings given by evaluators who are acquainted with the speakers will be compared with the ratings given by evaluators who are not acquainted with the speakers.

(3) The experience of the evaluator with the rating scale. The ratings given by evaluators experienced with the rating scale will be compared with the ratings given by these same evaluators before they had experience with the rating scale.

(4) The sex of the evaluator in relation to the sex of the speaker. The ratings given female speakers by male evaluators will be compared with the ratings given the same speakers by female evaluators. Likewise, the ratings given male speakers by male evaluators will be compared with the ratings given the same speakers by female evaluators.

Influence. Influence may be assumed to exist whenever

a statistically significant relationship is established between ratings given in any one of the above categories.

Ratings. Ratings in this study are the judgments of the "intergroup speech projects" as expressed by critics on the Evaluator's Rating Scale.[8]

Intergroup speech project. This is a phrase used at Central Michigan College of Education to designate an expository speech of approximately three minutes duration. It is delivered before a group of fifteen freshmen who are members of various sections of the Fundamentals of Speech course. In each audience there are two evaluators, one a student and the other a faculty member, who make independent ratings of each speaker. Each student evaluator is a college junior or senior who is also a speech major or minor and who has been approved by the speech faculty as a competent student. The faculty evaluators are members of the Speech Department of Central Michigan College of Education.

Fundamentals of Speech. This class, bearing the college catalog designation of Speech 101, is a two semester-hour course required of all freshmen on campus. They may register for it either the first or second semester. Approximately six hundred students take this course each year.

---

[8]See Appendix A.

C. Summary

In the light of the defined terms the problem may be regarded as an attempt to determine what statistically significant relationships, if any, exist between the judgments expressed by critic-judges on the Evaluator's Rating Scale and: (a) their academic speech training, (b) their acquaintance with the speaker, (c) their experience with the rating scale, and (d) their sex in relationship to the sex of the speaker.

This study is of particular concern to the students and faculty of Central Michigan College of Education where speech competency is a prerequisite to candidacy for a teaching certificate. However, it may also have implications for other comparable institutions that have similar programs.

CHAPTER II

REVIEW OF THE LITERATURE

In order to explore the literature pertinent to the present study, the writer not only read published and unpublished research in the speech field,[1, 2, 3] but also examined psychological, sociological, and pedagogical writings.[4, 5, 6]

---

[1]Lester W. Thonssen and Elizabeth Fatherson, Bibliography of Speech Education. New York: H. W. Wilson and Company, 1939. 800 pp.

[2]Lester W. Thonssen, Mary Margaret Robb, and Dorthea Thonssen, Bibliography of Speech Education - Supplement 1939-48. New York: H. W. Wilson and Company, 1950. 393 pp.

[3]Franklin H. Knower, Table of Contents of the Quarterly Journal of Speech (1915-1952); Speech Monographs (1934-1952); and the Speech Teacher (1952) with a Revised Index Compiled Through 1952. Columbia, Missouri: Speech Association of America, 1953. 61 pp.

[4]Walter S. Monroe, editor. Encyclopedia of Educational Research, Revised Edition. New York: The Macmillan Company, 1950. 1520 pp.

[5]Alice F. Moench and others, editors. The International Index to Periodicals Devoted Chiefly to the Humanities and Sciences. New York: H. W. Wilson Company, Vols. I-XII, 1913-1953.

[6]Isabell Towner and Ross Carpenter, editors. The Education Index; A Cumulative Author and Subject Index to a Selected List of Educational Periodicals, Books, and Pamphlets. New York: H. W. Wilson Company, Vols. I-VIII, January, 1929 - June, 1953. Also Education Index Monthly Check-List, July, 1953 - April, 1954.

The bibliographies compiled by Rosenberg[7] were consulted to discover the masters' and doctoral theses completed before 1945 which might have some bearing upon the problem. More recent studies were listed by Knower[8] and Auer.[9] The latter even included this study.[10]

The literature relating to this study will be reviewed under two categories: (1) speech rating in general, and (2) studies which give specific consideration to any of the four factors selected for investigation in the present study.

A. Speech Rating in General

Although much has been written about the use of rating scales in general,[11] little has been published on the rating of oral performances in Fundamentals of Speech classes.

---

[7]Ralph P. Rosenberg, "Bibliographies of Theses in America," Bulletin of Bibliography, 18:181-82, September-December, 1945.

[8]Franklin H. Knower, "Graduate Theses--An Index of Graduate Work in Speech," Speech Monographs, 21:108-35, June, 1954.

[9]J. Jeffery Auer, "Doctoral Dissertations in Speech: Work in Progress, 1954," Speech Monographs, 21: 136-41, June, 1954.

[10]Ibid., p. 141

[11]Carter V. Good, A. S. Barr, and Douglas E. Scates, The Methodology of Educational Research. New York: D. Appleton-Century Crofts, 1941, pp. 424-37.

Symonds[12] points out that group rating is more reliable than is individual judgment:

> A single observation is unreliable, a single rating is unreliable, a single test is unreliable, a single measurement is unreliable, a single answer to a question is unreliable. Reliability is achieved by heaping up observations, ratings, tests, questions, measures... An adequate rating requires the judgment of several raters in several situations at several different times. Reliable evidence must be multiplied evidence.

Rugg[13] recommends the use of pooled or averaged ratings of not less than three independent raters. In each instance it is assumed that the several raters are all competent to rate and that the reliability of pooled ratings tends to increase according to the Spearman-Brown formula.[14]

Holcomb[15] found that, although most judges take careful notes and have a number of definite points on which to judge, they do have personal standards which vary widely from one judge to another.

---

[12]Percival M. Symonds, Diagnosing Personality and Conduct. New York: D. Appleton-Century Company, 1931, p. 5.

[13]Harold O. Rugg, "Is the Rating of Human Character Practicable?" Journal of Educational Psychology, 12:425-38, November, 1921.

[14]Joy P. Guilford, Psychometric Methods. New York: McGraw-Hill Book Company, Inc., 1936, p. 421.

[15]Martin J. Holcomb, "The Critic-Judge System," Quarterly Journal of Speech, 19:28-38, February, 1933.

Knower,[16] who has done a great deal of research in the field of speech evaluation, says:

> The objectivity of observational evaluation is entirely a matter of the objectivity of raters. Although the standards of evaluation in this process are ostensibly subjective, it remains a fact that such judgments may be as accurate, or even more accurate, than an arbitrarily assigned score derived from items on an objective paper and pencil test.

B. The Four Factors

The almost complete lack of research in the area outlined by this study, namely the four factors which may affect the rating of speech performances, indicates the need for this work to be done. Furthermore, where studies have been conducted, as in the area of sex influences, the evidence is inconclusive and even contradictory.

1. The academic speech training of the evaluator. West and Larsen[17] experimented with students in the required freshman course in speech in the State University of Iowa. Students ranked their classmates, and their "class ratings" were compared with "grades" given by the instructor. They reported:

---

[16]Franklin H. Knower, "What is a Speech Test?" Quarterly Journal of Speech, 30:485-93, December, 1944.

[17]Robert West and Helen Larsen, "Some Statistical Investigations in the Field of Speech," Quarterly Journal of Speech, 7:375-82, November, 1921.

> The relation between the combined judgment of the class on each speaker and the instructor's judgment on each speaker, or the correlation between "class ratings" and "grades" computed on about 300 cases, is .453. Assuming that the instructor's grade is made upon a reasonable basis, one would say that comparatively just marks could be given a student of speech by getting a rating from the class.

This conclusion embodies as one of its basic assumptions the acceptance of the standard postulated by Rugg[18] as a test of significance:

> The experience of the present writer in examining many correlation tables has led him to regard correlation as 'negligible' or 'indifferent' when r (coefficient of correlation) is less than .15 to .20; as being 'present but low' when r ranges from .15 or .20 to .35 or .40; as being 'markedly present' or 'marked' when r ranges from .35 or .40 to .50 or .60; as being 'high' when it is above .60 or .70.

Knower[19] investigated the extent of agreement between students and instructors in their rating of student speakers. He had instructors and students rate thirty-three speakers. The ratings given by students were correlated, by the rank order method, with the raw score given by the instructors. In light of these correlations he concludes:

> Since the correlations were consistently higher, with one exception, between the students' ratings and the instructors' ratings than between the ratings of the instructors, we have a more objective criterion of effective public speaking in the average of a number of student scores than we have in the scores assigned by one instructor.

---

[18]Harold O. Rugg, Statistical Methods Applied to Education. New York: Houghton Mifflin Company, 1917, p. 256.

[19]Franklin H. Knower, "A Suggestive Study of Public Speaking Rating-Scale Values," Quarterly Journal of Speech, 15:30-41, February, 1929.

Anderson[20] made a study of the ratings given by 169 students to their classmates in a basic communications course, at that time called Written and Spoken English, at Michigan State College. The student speakers were in eight different rooms. Each of these eight groups was rated by three faculty members as well as by their fellow classmates. By comparing these ratings she came to the conclusion that the students were more in agreement as to the ratings the speaker should get while the faculty varied more in their judgments. However, this is not necessarily a valid comparison of faculty raters with student raters since the faculty used a rating scale listing five traits while the students evaluated only upon one of these five traits.

Gibbs[21] investigated the degree of consistency in evaluations made by faculty members and students listening to recordings of three minute speeches. The students were classmates of the speakers. According to this study student evaluators place more speakers in the below average classification than do faculty evaluators. However, the evaluations were made only of voice and articulation; and, since

---

[20]Mary Margaret Anderson, "An Analysis of Some of the Sources of Variation Involved in Rating Speeches," (unpublished Master's thesis, Michigan State College, East Lansing, Michigan, 1945), 19 pp.

[21]David Elmore Gibbs, "A Study of Reliability and Variation of Critical Rating of Speech by Trained and Untrained Observers," (unpublished Master's thesis, The University of Washington, Seattle, 1948), 87 pp.

none of the students had any courses in this field, training may have been the important factor here. There is no evidence to indicate whether the results would be the same if the students were juniors and seniors who had speech training.

Andregg[22] made an analysis of the ratings on six traits: "thinking, knowledge, initiative, cooperation, organizing ability, and expression." These ratings were made by students and instructors on the performance of student officers attending the Air Command and Staff School at the Air University. The officers participated in the planning of the tactical and strategic air operations; rated each other's performances; and were rated by their instructors, also officers, who devoted full time to observation and rating.

The study showed that students and instructors rated most reliably on "expression." Also students rated their fellow staff officers more leniently than did instructors. However, the situation of rating officers on general characteristics in the Air Command and Staff School may not be entirely comparable to rating freshmen in Fundamentals of Speech by faculty members and upperclassmen who are speech majors and minors.

---

[22]Neal Berry Andregg, "A Critical Study of Graphic Rating Scales," (unpublished Doctoral dissertation, Michigan State College, East Lansing, Michigan, 1951), 138 pp.

2. The acquaintanceship of the evaluator with the speaker. Seedorf[23] conducted a study to find out how much agreement there is among individuals in their response to an oral interpretation of literature. Among other things she answered the question: How does acquaintance with a fellow-classmate's quality of work affect the amount of agreement among judges? She states:

> The correlation of the mean scores of each member of the two groups of student-judges, the acquainted and the unacquainted, for one group of readers, were .958 and .887 respectively, indicating that when evaluated by fellow students of approximately the same degree of training, the readers received about the same rank whether given by fellow classmates or by students who were not classmates.

Knight[24] analyzed ratings of 1948 public school teachers of one school system made by the supervisors under whom the teachers were working. He concluded:

> The factor of acquaintance operates to make ratings more lenient, i. e., increases the over-rating, and to make ratings less critical and less analytical, i. e., increases the influence of the halo of general estimate. In a way it is literally true to say of a judge's estimate: "His judgment is of doubtful validity because he has known this man too long."

---

[23]Evelyn H. Seedorf, "An Experimental Study in the Amount of Agreement Among Judges in Evaluating Oral Interpretation," Journal of Educational Research, 43:10-21, September, 1949.

[24]Frederic B. Knight, "The Effect of the Acquaintance Factor upon Personal Judgments," Journal of Educational Psychology, 14:129-42, March, 1923.

Henrickson[25] made a study of one hundred and seventy-nine students in Fundamentals of Speech courses, eighty-one from three classes at the University of Montana and ninety-eight from four classes at Iowa State Teachers College. He asked them to rate their classmates at the end of a semester or quarter on: (1) how well they knew the person; (2) how well they liked the person; (3) how good they thought the person was as a speaker. From a study of these data he came to the conclusion that the better known students are apparently liked better and are judged to be somewhat better speakers.

3. The experience of the evaluator with the rating scale. Carroll[26] conducted an experiment in rating musical selections played on phonograph records. He used two sections of a class in educational psychology as raters; and each section rated the selections according to (1) volume, (2) expression, (3) quality, (4) melody, (5) harmony, and (6) rhythm. One section, the control group, rated three records. Some weeks later they rated the same three records again. The second section, the drill group, rated the same three records at the same times as the control

---

[25]Ernest H. Henrickson, "The Relation Among Knowing a Person, Liking a Person, and Judging Him as a Speaker," Speech Monographs, 7:22-25, 1940.

[26]Robert P. Carroll, "Practice in Rating," Journal of Experimental Psychology, 14:299-302, June, 1931.

group. However, between the first and second rating periods, this second section was given new records to rate three times a week. A great deal of discussion relative to method of rating was done in the drill group between ratings. Carroll criticized his research in that, due to the absences, there were changes in each group. Also he did nothing to equate the two groups. However, he concludes that "in general, the results of the experiment seem to indicate that by training in subjective ratings individuals may improve; that their ratings may more nearly agree and that they may become more reliable."

Thompson[27] in his investigation to determine the accuracy of typical speech rating techniques collected data from eleven classes in Fundamentals of Speech. Both speakers and raters were members of these classes. The raters were freshmen taking their first college course in speech and had little or no formal training in rating speeches. The procedure was to divide the class into two groups of raters. Both groups listened to the same speeches, but one used rating scales while the other used letter grades. The data were analyzed to find the variance for each method. Then, on the assumption that the system which produces the least variance

---

[27]Wayne N. Thompson, "An Experimental Study of the Accuracy of Typical Speech Rating Technique," (unpublished Doctor's dissertation, Northwestern University, Evanston, 1943), 204 pp.

is the best, the investigator concluded that a rating scale is superior the first day of the experiment but that after practice (or fatigue) letter grades are more accurate.

4. The sex of the evaluator in relation to the sex of the speaker. There have been many studies conducted and much has been written, often quite contradictory, in the field of sex differences. A few of these are somewhat relevant to the aspect of sex differences being considered in this study.

Anastasi,[28] who has done considerable writing in the field of sex differences says:

> It is apparent that the effectual environment of the two sexes are fundamentally diverse from an early age. Under such conditions, we should expect pronounced variation in the emotional and intellectual development of the two sexes.

Lehman and Witty[29] caution that in collecting data from males and females to determine variability the same age-levels only should be compared. He states:

> One source of error is hasty generalization due to the inclusion of all or nearly all age-levels in the formulation of conclusions. Because of the irregular development of numerous human characteristics, one may expect to find that at a certain age-level girls will be more variable than boys in some regards, and at an earlier or later age the reverse may be true.

---

[28]Ann Anastasi, Differential Psychology: Individual and Group Differences in Behavior. Macmillan Company, New York, 1937. p. 386.

[29]Harvey G. Lehman and Paul A. Witty, "Sex Differences: Some Sources of Confusion and Error," American Journal of Psychology, 42:140-47, January, 1930, p. 143.

Symonds[30] studied the differences of areas of interest according to sex. Fifteen areas of human interest were ranked by 784 high school boys, 857 high school girls, 276 college men, 387 college women, 73 men graduate students, and 111 women graduate students in order of interest for reading or discussion. By taking average ranks for the various groups he noted differences by sex for different maturity levels. He found a greater difference in the area of interest between college men and college women than he found between high school boys and high school girls or between men and women doing graduate work.

Carter[31] made an investigation of the assignment of marks by teachers of beginning algebra to determine whether or not teachers tend to favor one sex and whether the sex favored tends to be determined by the sex of the teacher. Nine classes, five taught by men and four taught by women, were used. In these classes there were 135 boys and 100 girls. Intelligence, achievement, and personality of these students were measured by standardized tests; and the relationship between teachers' marks and (1) intelligence,

---

[30]Percival M. Symonds, "Changes in Sex Differences in Problems and Interests of Adolescents with Increasing Age," Journal of Genetic Psychology, 50:83-89, March, 1937.

[31]Robert Scriven Carter, "Non-intellectual Variables Involved in Teachers' Marks," Journal of Educational Research, 47:81-95, October, 1953.

(2) achievement, and (3) personality was determined by computing the correlation coefficient. The results showed that the relationship between teachers' marks and intelligence was higher for the boys than for the girls. Also that the relationship between teachers' marks and achievement scores was higher for the boys than the girls. However, the relationship between teachers' marks and personality was higher for the girls than for the boys. Marks given by women teachers correlated higher with personality test scores than did the marks given by men. Conversely, marks given by men correlated higher with achievement and intelligence than did marks given by women.

Douglas and Newman,[32] who made a study of achievement and marks of 3366 students in four Minnesota high schools, say:

> In the light of the data of this and other investigations, it seems probable that marks are determined by factors other than achievement, especially marks assigned by women teachers, and that these influences result in the slight overrating of girls generally and the peculiar underrating of boys by women teachers.

However, this refers to marks in English, history, and mathematics given to high school students by faculty members and may not be comparable to ratings given in a college speech course by college students or faculty members.

---

[32]Harl R. Douglas and Olson E. Newman, "The Relation of High School Marks to Sex in Four Minnesota Senior High Schools," School Review, 45:481-88, April, 1937, p. 288.

Fifty members of the faculty at Northwestern University ranked one hundred and four students, fifty-three men and fifty-one women, by classifying them into ten different groups according to estimated intelligence. These were correlated with scores received by the students on a battery of intelligence tests. As Webb[33] reports:

> Each group showed some partiality to the opposite sex in estimating intelligence; that is, the men gave evidence of placing a slightly higher value on the intelligence of women than they do that of men. The women appear to do the same thing in regard to the men.

The writer[34] made a survey of the attitudes of 227 intercollegiate debaters towards debate judges. He found that, as far as the feelings of debaters were concerned, the sex of the judge made a difference in more than half of the cases. Only 39 per cent of the men and 49 per cent of the women felt that the sex of the debate judge had no effect upon the ratings they received. Both men and women debaters preferred male over female judges. However, this preference for male judges was slightly more pronounced among women debaters (45 per cent) than among men debaters (40 per cent). On the other hand 21 per cent of the men debaters thought

---

[33]L. W. Webb, "The Ability of Men and Women to Judge Intelligence," School and Society, 20:251-54, August 23, 1924.

[34]Emil R. Pfister, "A Survey of Attitudes Toward Debate Judges," Forensic of Pi Kappa Delta, 39:102-03, May, 1954.

female judges rated them higher while only 6 per cent of the women debaters thought they were rated higher by female judges.

Knower[35] administered one form of the Smith and Thurstone Attitude toward Prohibition Scale before a speech and gave an equated form after the presentation. He concluded that the delivery of speeches produces a change of attitude statistically significant and that women speakers are more influential with a male audience than are men speakers. Similarly, a female audience is influenced more by men speakers than by women speakers.

Graunke[36] found that female instructors gave higher ratings than did the male instructors. However, these ratings were not broken down to determine whether female instructors rated both male and female students higher than did the male instructors.

Penland[37] made a study of the ratings given to eighty-seven university sophomores, fifty-three women and thirty-four men. These students read orally and were rated by both male

---

[35]Franklin H. Knower, "Experimental Studies of Changes in Attitudes," Journal of Social Psychology, 6:315-44, August, 1935.

[36]Graunke, op. cit., p. 102.

[37]Virgil Darrell Penland, "An Experimental Study to Measure Effectiveness in Oral Reading by Means of a Rating Scale Technique," (unpublished Doctor's dissertation, The University of Southern California, Los Angeles, 1948), 177 pp.

and female judges. He found one "probably significant difference," i.e. that female judges tended to be more "severe" in rating women performers in this field of oral reading.

C. Summary

Speech rating in general. A survey of the literature that deals with speech rating in general indicates that:

(1) Group rating is more reliable than individual judgment.

(2) Reliability of pooled ratings increases as the number of competent raters is increased.

(3) Personal standards vary widely among judges.

The four factors. Although no research identical to this experiment has been conducted, studies have been made that are related to the four factors being considered in this study:

(1) Five studies compare ratings given by faculty members with ratings given by students. One experiment concludes that rating by a group of students is more accurate than by a single faculty member. Two studies point out that ratings given by students are more in agreement than are ratings given by faculty. One study indicates that student raters give more lenient ratings than do the faculty while another indicates that faculty give the more lenient ratings. However, all five of these studies use freshmen evaluators who are the speakers' classmates. These ratings may not be

equivalent to those given by speech majors and minors in their junior or senior year of college.

(2) Three studies consider the factor of acquaintance with the speaker. Here, too, the students used as evaluators were classmates of the speakers. All three studies agreed that evaluators acquainted with the speakers were more lenient than evaluators unacquainted with the speakers.

(3) Only two research studies could be found that were concerned with the evaluator's experience with the speech rating scale. The first of these studies found that training improved the reliability of the rater. However, it should be noted that this experiment was conducted by rating music on phonograph records and that coaching as well as practice was used. The second study used classmates to rate speeches. It concluded that a rating scale is superior to letter grades the first day, but after that letter grades are more accurate.

(4) The studies regarding the relationship between ratings given a speaker and the sex of the evaluator are inconclusive. This survey concurs with an earlier report by McNemar and Terman[38] regarding variability in mental traits between sexes:

---

[38]Quinn McNemar and Lewis M. Terman, "Sex Differences in Variational Tendency," Genetic Psychology Monographs, 18:8, February, 1936.

Research has not proved either the presence or absence of a sex difference in variability with respect to psychological traits. There are few problems in psychology on which investigations that would appear to be comparable have yielded results so discordant.

CHAPTER III

PROCEDURE

The data for this study were collected during the 1952-53 academic year at Central Michigan College of Education, Mount Pleasant, Michigan.

Six hundred and four people cooperated to make this experiment possible. They may be divided into two groups:

(1) There were the speakers, five hundred and forty-nine of the five hundred and ninety-eight freshmen enrolled in Fundamentals of Speech classes.[1] Three hundred and five of these were in the nineteen sections which were taught during the first semester and two hundred and forty-four were in the nineteen sections taught during the second semester. Each of these students gave two different three minute informational speeches before audiences which averaged about fifteen people, most of whom were unacquainted with the speaker. Thus the freshmen gave a total of 1098 three minute informational speeches. Furthermore, since each of these speeches was given before two different audiences, the students compiled a total of 2196 speech performances.

---

[1]The forty-nine freshmen excluded from this experiment were those who, because of some reason such as illness, were unable to participate in all four of the intergroup speech projects.

(2) There were the fifty-five evaluators, forty-six students and nine faculty members.[2] The student evaluators were juniors and seniors who were speech majors or minors while the faculty evaluators were members of the Department of Speech. Each of the 2196 speech performances was rated by at least one student and one faculty member. These raters sat in the audience, worked independently, and used a standard rating scale.

Therefore, by having various pairs of evaluators, one student and one faculty, rate the 2196 speech performances, a total of 4392 ratings was collected. As will be explained later in this chapter, precautions were taken so that these ratings would be comparable.

A. Devising the Evaluator's Rating Scale

The Evaluator's Rating Scale[3] that was used in this experiment was devised by the writer who employed the following procedure:

(1) A study was made of the speech rating scales

---

[2]See Appendix B.

[3]See Appendix A.

which have appeared in speech textbooks and periodicals published in the United States.[4-14]

---

[4]Arthur W. Cable, "A Criticism Card for Class Use," Journal of Speech Education, 12:186-88, April, 1926.

[5]J. Stanley Gray, "Objective Measurement for Public Speaking," Journal of Expression, 2:20-26, March, 1928.

[6]Wilmer E. Stevens, "A Rating Scale for Public Speakers," Quarterly Journal of Speech Education, 14:223-32, April, 1928.

[7]Alice J. Bryan and Walter H. Wilke, "A Scale for Measuring Speaking Abilities," Psychological Bulletin, 33:605-06, October, 1936.

[8]Harry G. Barnes, "Appendix," Speech Handbook. Iowa City: Privately printed, 1936. 138 pp.

[9]Helen L. Ogg and Ray K. Immel, "Speech Criticism Chart," Speech Improvement. New York: F. S. Crofts and Company, 1936. 190 pp.

[10]Elwood Murray, The Speech Personality. New York: J. B. Lippincott Company, 1944. pp. 271-391.

[11]Wilhelmina G. Hedde and William N. Brigance, "A Score Sheet for Judging Speeches," American Speech. New York: J. B. Lippincott Company, 1946. pp. 581-82.

[12]Alice Evelyn Craig, The Speech Arts. New York: The Macmillan Company, 1947, p. 256.

[13]A. Craig Baird and Franklin H. Knower, "Appendix D," General Speech. New York: McGraw-Hill Company, 1949, p. 294.

[14]Karl F. Robinson, Teaching Speech in Secondary School. New York: Longmans, Green and Company, 1951. pp. 123-28.

(2) A survey was made of the literature regarding the construction of speech rating scales.[15-22]

(3) Taking into consideration the conclusions from previous research conducted in the field of rating scale construction, the writer devised a rating instrument. This instrument incorporated the elements common to other speech rating scales that had been used by others with some satisfaction in the past. This was revised and refined in the light of suggestions offered by faculty members of the Speech Department as well as members of the advisory committee for this thesis.

---

[15]J. B. Miner, "The Evaluation of a Method for Finely Graduated Estimates of Ability," Journal of Applied Psychology, 1:123-33, June, 1917.

[16]Max Freyd, "The Graphic Rating Scale," Journal of Educational Psychology, 14:83-102, February, 1923.

[17]Percival M. Symonds, "Notes on Rating," Journal of Applied Psychology, 9:188-95, June, 1925.

[18]Paul H. Furfew, "An Improved Rating Scale Technique," Journal of Educational Psychology, 17:45-48, January,

[19]Percival M. Symonds, "Rating Methods," Diagnosing Personality and Conduct. New York: D. Appleton-Century Company, 1931. pp. 41-121.

[20]Lee Norvelle, "Development and Application of a Method for Measuring the Effectiveness of Instruction in a Basic Speech Course," Speech Monographs, 1:41-65, 1934.

[21]Alice J. Bryan and Walter H. Wilke, "A Technique for Rating Speeches," Journal of Consulting Psychology, 5:80-90, March-April, 1941.

[22]Isabel Kincheloe, "On Refining the Speech Scales," English Journal, 34:204-07, April, 1945.

(4) The writer presented this speech rating scale to his colleagues at a departmental staff meeting of the speech faculty of Central Michigan College of Education where it was discussed and received unanimous approval.

(5) The last two steps, (a) that of introducing the rating scale to the evaluators and (b) that of investigating its reliability and validity, will be discussed later.

B. Collecting the Data

Before conducting the experiment it was essential to secure the cooperation of the faculty members of the speech department. During September, 1952, several Speech Department staff meetings were held previous to the registration of students. At one of these the writer outlined a plan for conducting this experiment in evaluating oral performances of students in Fundamentals of Speech classes. The members of the speech faculty were not only willing to cooperate but they also liberally contributed ideas, time, and effort.

The project also required the cooperation of the juniors and seniors who were on either a speech major or speech minor curriculum. When asked if they would be willing to serve as evaluators, their response indicated that in general they were eager to have the experience as a background for preparation as future teachers of speech.

Scheduling the evaluators. Scheduling the evaluators from the speech faculty was accomplished with little difficulty since the instructors of Fundamentals of Speech agreed that no class sessions of the course were to be held during the weeks that the intergroup speech projects were scheduled. This policy freed the faculty to serve as evaluators. The fact that each student missed two class sessions during that week could be justified because each student was having the experience of giving the same speech before two different audiences. Furthermore, he was getting the evaluations of four well qualified evaluators.

Securing qualified student evaluators required more effort. The first step was to compile a list of the seventy students who were either speech majors or minors and who were also either juniors or seniors.[23] This list was duplicated and copies were sent to each member of the speech faculty in order to determine (1) the number and type of speech courses that each student had taken, (2) professors with whom he had done speech work, and (3) whether the professor regarded the student as qualified to serve as an evaluator.[24]

---

[23]See Appendix C.

[24]See Appendix C.

Meanwhile, letters signed by the Head of the Department of Speech and Drama were sent to all juniors and seniors who were speech majors and minors.[25] These letters explained the intergroup speech project, solicited student cooperation, and included a student evaluator's preference report blank.[26]

When these blanks were filled out and returned, the juniors and seniors were assigned groups to evaluate. These assignments were made according to the student's availability and preference. Then each student was sent a letter informing him of the time or times that he was scheduled to serve as an evaluator.[27]

Preparing the speakers. At the sixth meeting of the Fundamentals of Speech classes during the fall as well as during the spring semester, the speech sections were given the following assignment:

> You are to prepare a three minute informative speech and deliver it on the week of _____. You will be scheduled to speak before two different audiences composed of students from other Speech 101 classes. You will be rated in each of the performances by a student who is a Junior or Senior and a Speech major or minor. You will also be rated in each of the performances by a member of the speech faculty.

---

[25]See Appendix D.

[26]See Appendix E.

[27]See Appendix F.

At the seventh meeting of the class a sheet was given to each student on which he could list the dates and times that he preferred as well as those when he could not speak.[28] This helped in scheduling students for speech performances. Each instructor was given a schedule on which were listed the date, time, and room that each student was assigned. After the teacher read this aloud and the student wrote down his speech schedule for that week, the sheets were posted outside the speech secretary's office so that any student might double check his speaking assignments.

During each semester every student enrolled in Fundamentals of Speech classes was expected to participate in two Intergroup Speech Projects. First semester students gave their first project speeches the week of October 27-31, 1952, and their second project performances the week of January 12-16, 1953. The second semester students gave their first project speeches the week of March 9-13, 1953, and their second series of speeches the week of May 11-15, 1953.

Preparing the evaluators. The first problem in preparing the evaluators was to familiarize them with the rating scale without any indoctrination that would make this experiment sterile. However, the evaluators had to have verbal

---

[28]See Appendix G.

agreement regarding what was to be rated. As Wilke[29] says:

> The first difficulty which anyone attempting to rate individuals runs up against is the matter of attaching exact meanings to the terms used on the rating scale. Many previous users of rating devices have urged the use of careful definitions to establish unequivocal meaning.

According to Monroe,[30] research upon the problem of increasing the agreement among judges when rating scales are used discloses that an adequate definition of what is being rated is crucial.

Symonds[31] attempts to outline the method by which such definition is attained:

> Particular attention must be paid to the definition of the items in the scale. On this hinges much of the success or failure of ratings in general. One of the most potent factors causing unreliability of ratings is ambiguity in the meaning of items on the scale. Thus in every rating scale the items should be defined in some way. There are several possible ways of doing this. One, perhaps the least satisfactory, is to give synonyms of the original term. Another is a short paragraph amplifying the descriptive title. Another method is to ask a question which not only limits the meaning of the term but somehow helps the rater to see the problem of rating more clearly.

Furfey[32] conducted research which indicated that reliability could be increased not only by increasing the

---

[29]Walter H. Wilke, "A Subjective Measurement in Speech: A Note on Method," Quarterly Journal of Speech, 21:55, February, 1935.

[30]Monroe, op. cit., p. 961.

[31]Symonds, op. cit., p. 84.

[32]Furfey, op. cit., p. 92.

number of judges but also by increasing the number of judgments which each judge makes. He explains:

> This is easily accomplished by analyzing the trait to be rated into several sub-traits, by having the judges rate all these sub-traits separately and then combining these separate ratings into a final score. This is quite comparable to the process of measuring intelligence by measuring separately a number of abilities which are believed to correlate highly with intelligence and then combining the separate results into a final score.

This subdividing of traits may be overdone, of course, but the need for more specific items cannot, according to Freeman,[33] be ignored:

> Frequently it is held that the uniqueness of the individual personality pattern renders futile any analysis into elements which, when isolated and measured, lose their meaning. Because this view is at variance with canons of scientific procedure, it should be examined very critically. There is a middle ground somewhere, and this we must find before real progress in personality assay is made.

The students and faculty who agreed to serve as evaluators were given copies of the Evaluator's Rating Scale[34] to study two weeks before the intergroup speech projects were scheduled to begin. The week prior to the projects the evaluators met twice and discussed the question: "What is meant by the various items on this rating scale?" Student-faculty committees were set up on each of the four major divisions: (1) "Thought," (2) Language," (3) "Voice," and (4) "Action."

---

[33]Graydon LaVern Freeman, The Energetics of Human Behavior. Ithaca, New York: Cornell University Press, 1948. p. 17.

[34]See Appendix A.

Students acted as committee chairmen while faculty members served as resource persons. Only those items that were submitted by the committee and agreed upon unanimously by the evaluators were accepted as the official interpretation of the criteria used. These criteria were then mimeographed and distributed so that each evaluator would have a copy of the interpretation of the rating scale.[35] Thus an attempt was made to reduce the variables inherent in interpreting the rating instrument.

The speech faculty members who had considered and discussed the intergroup speech project early in the semester compiled a list of instructions regarding how the project should be carried out so that the procedure would be consistent in all sections.[36] These also were mimeographed and sent to all the evaluators.

Thus an effort was made to prepare the evaluators so that they would understand and appreciate the meaning, use, and purpose of the rating scale. This was in accord with the advice offered by Strang:[37]

> Only by taking into consideration the way in which the rating is used, the harm that may result from superficial or inaccurate rating, and the service which the

---

[35]See Appendix H.

[36]See Appendix I.

[37]Ruth Strang, "Seven Ways to Improve the Rating Process," Occupations, 29:107-10, November, 1950.

rating may perform in preventing the individual from getting into situations in which he is likely to fail, can the rater appreciate the importance of the rating.

Description of the experimental setting. Directions given in "Instructions to Evaluators"[38] were followed carefully since it was essential to the success of this investigation that certain conditions be kept constant. To aid in achieving this objective, precautions were taken to see that several controls operated and that all of the speakers gave the same type of speeches under similar conditions before paired judges. No exceptions were considered.[39]

Accordingly, these procedures were followed:

(1) Only college freshmen enrolled in Fundamentals of Speech participated as speakers. They gave the same length speeches (approximately three minutes) with the same general purpose (to inform).

(2) All audiences were similar, being composed of approximately fifteen speakers from various sections of the class, and two evaluators, one a member of the speech faculty and the other a college junior or senior majoring or minoring in speech.

(3) Each pair of evaluators followed identical instructions, heard the same speeches at the same time, used

---

[38]See Appendix I.

[39]Where an exception occurred, the rating scales were kept separate and were not used in this study.

the standard rating scale, and had previously agreed upon the interpretation of that rating scale.

(4) Each speaker gave two intergroup speech project information talks. Both of these were given in the same room at the same time of day and the same day of the week, exactly nine weeks apart.[40] Each speaker also had the same audience and the same pair of judges listen to both of his speeches.[41]

Additional evaluators. When pairs of evaluators heard the same speakers give their second series of information speeches during the week of January 12-16, a third evaluator was present in sixteen of the sections. Each of these additional sixteen evaluators was also either a member of the speech faculty or a junior or senior who was a speech major or minor. This was done in order to be able to study the correlation of scores given by two different faculty members, or two different upperclassmen, who heard the same speech at the same time.

---

[40]During the first semester, 1952-53, the first intergroup speech project was conducted during the week of October 27-31, 1952, and the second project the week of January 12-16, 1953. The second semester the first project was March 9-13, and the second, May 11-15, 1953.

[41]Since the second series of intergroup speech projects were scheduled for corresponding times and days, there was not a great deal of difficulty in securing the same evaluators. However, in such cases where substitute evaluators were necessary, the ratings were not considered in this experiment.

The additional evaluators, sitting in with the paired evaluators and rating the speakers, made another comparison possible. Ratings given by evaluators hearing the speakers at the same time could be compared with the ratings given by the evaluators hearing these speakers give the same speech at a different time.

Checking rater-speaker acquaintance. Immediately preceding the fourth intergroup speech project, May 11-15, an extra form[42] was added to the rating scale in order to determine whether or not the rater was acquainted with the speaker. A similar form[43] was given to each speaker so that he could indicate the extent of his acquaintance with the raters.

C. Tabulating the Data

The fact that this experiment was designed to study four separate factors and that the data consisted of nearly forty-five hundred rating scales, each filled out with twenty-five specific items of information, made machine tabulation a practical necessity. This need was met by the use of the IBM equipment.[44]

---

[42]See Appendix J.

[43]See Appendix K.

[44]IBM equipment, manufactured by International Business Machines, 590 Madison Avenue, New York City, New York, is available for research at Michigan State College.

Mechanical tabulation. Mechanical tabulation was facilitated by the use of a special punch card.[45] This punch card made it possible to record sixty separate items on each card by use of a zero through nine code.[46]

Four steps had to be taken in order to convert the raw data on the rating scales into a form which could be handled by IBM methods:

(1) Data on the rating scales were reduced to a numerical code.[47]

(2) The data were then entered on the punch cards by a trained IBM operator.

(3) The cards were sorted by a mechanical sorter.

(4) The data were then assembled by an electronic tabulator.

Organizing the data. First the data were arranged in tables designed to facilitate determining how well each student performed in comparison with his fellow classmates as well as how much improvement he had shown during the nine weeks between the first and second series of intergroup

---

[45]This punch card was designed by Francis B. Martin, Supervisor of Tabulation, Michigan State College.

[46]See Appendix L.

[47]See Appendix M.

speech projects.[48] This arrangement of the data, although useful in computing grades for the students, could not be used to answer the four primary questions being considered in this study.

Secondly, the data were arranged into sixteen categories, taking into consideration the academic speech training of the raters and their sex in relationship to the sex of the speaker. These categories consisted of two major divisions, male speakers and female speakers, each broken down into eight sub-groups:

(1) Scores given by male faculty evaluators serving with male student evaluators.

(2) Scores given by the male student evaluators serving with the above male faculty evaluators.

(3) Scores given by female faculty evaluators serving with female student evaluators.

(4) Scores given by the female student evaluators serving with the above female faculty evaluators.

(5) Scores given by male faculty evaluators serving with female student evaluators.

(6) Scores given by the female student evaluators serving with the above faculty evaluators.

---

[48]These tables, consisting of forty-two pages, have not been included in the appendix of this thesis because of their bulkiness. The writer has a duplicate copy available for anyone's use.

(7) Scores given by female faculty evaluators serving with male student evaluators.

(8) Scores given by the male student evaluators serving with the above female faculty evaluators.

Comparisons of these scores and their statistical significance are presented in the following chapter.

Thirdly, in order to determine the influence which experience with the rating scale had upon the ratings, the data were arranged so that the scores given by pairs of evaluators during the first intergroup speech project could be compared with the scores these pairs of evaluators gave the same speakers during the second intergroup speech project. These were then treated statistically as will be explained later.

Lastly, in order to consider whether evaluators who were acquainted with the speakers whom they rated tended to give scores significantly higher or lower than the evaluators who were not acquainted with these same speakers, the Evaluator's Acquaintanceship Check Sheet was used.[49] Scores given by evaluators who indicated that they were unacquainted with the speakers were compared with scores given by evaluators who were acquainted with these same speakers.

Statistical procedure. The available literature describing the principles and methods of population parameters

---

[49]See Appendix J.

and sample statistics is far too extensive to summarize in this study. However, certain citations are included in an attempt to provide examples of typical authoritative support which is available concerning the mathematical methods used in this study.

An example of calculation of the standard deviation from original scores is given by Garrett.[50] He also demonstrates how to find the limits in any normal distribution which will include a given percentage of cases.[51] This was especially useful in allocating grades according to the normal probability curve.

In order to determine the significance of the difference in the means of scores given by evaluators influenced by one factor compared with judges influenced by another factor, the "Student's t" test was used.[52]

Coefficients of correlation, symbolized by "r," were computed by the product-moment method.[53] This is described

---

[50]Henry E. Garrett, Statistics in Psychology and Education. New York: Longmans, Green and Company, 1947. p. 63.

[51]Ibid., pp. 197-208.

[52]A full account of this test and the table for its use will be found in Ronald A. Fisher's Statistical Methods for Research Workers, London: Oliver and Boyd, Ltd., 1941, pp. 116-17. Its originator published anonymously under the pseudonym "Student."

[53]The coefficient of correlation, "r," is often called the "Pearson r" after Professor Karl Pearson who developed the product-moment method.

in detail by Snedecor.[54] Its importance in the determination of the reliability of an evaluation instrument (such as a rating scale) was pointed out by Good and others:[55]

> Correlation has an extensive use in connection with the critical study of tests and other instruments. The correlation of two series of measure that are supposed to represent the same thing (such as two applications of a standard test, or of comparable forms of it), is known as the coefficient of reliability.

The writer was fortunate to have at his disposal electric calculators.[56] These were most useful when computing correlation coefficients.

Further references to statistical methods are made in the chapter presenting the analysis of the data obtained during the course of the investigation.

D. Summary

After a year of planning and experimentation the writer devised an instrument to measure speech proficiency, the Evaluator's Rating Scale. Then during the academic year, 1952-53, with the cooperation of Central Michigan College of Education's juniors and seniors who were speech

---

[54]George W. Snedecor, Statistical Methods. Ames, Iowa: The Iowa State College Press, 1946, pp. 123-41.

[55]Good, Barr, and Scates, op. cit., p. 607.

[56]Techniques for the efficient operation of these machines are given in Katharine Pease's Machine Computations of Elementary Statistics. New York: Chartwell House, Incorporated, 1949, 208 pp.

majors or minors, the speech faculty, and the freshmen in Fundamentals of Speech classes, the experiment was conducted. Five hundred and forty-nine freshmen prepared two speeches and gave each speech twice. Each speech was approximately three minutes long and its general purpose was expository.

Two evaluators, one an upperclass student and the other a faculty member, were in each audience and rated each speech performance. Thus 4392 ratings were collected. Furthermore, a check was made of rater-speaker acquaintanceship.

Although some variability was unavoidable, every possible effort was made to have sufficient controls operating regarding speaker, speech, audience, and occasion so that the ratings would be comparable.

The data were transferred from the rating scales onto punch cards and IBM methods for sorting and tabulating were employed. Then with the use of electric calculators the data were dealt with statistically. The formulas used and the organized presentation of the findings will appear in the next chapter.

CHAPTER IV

ANALYSIS OF THE DATA

The analysis of the data gathered in this experiment considers (1) the rating scale itself, particularly its validity and reliability, (2) the distribution of scores and their practical application to a marking system, and (3) the statistical relationships between each of the four factors and the ratings given by the evaluators.

A. The Rating Scale

Two important considerations of any measuring instrument are its validity and reliability. Validity means the degree to which any device or technique measures that which it is designed to measure. As stated by Cook:[1]

> A test is said to have high validity when it measures effectively the property it purports to measure. A measure of validity of a test is secured by computing a coefficient of correlation between scores on the test and an outside criterion.

Reliability refers to the consistency with which an instrument measures whatever it does measure. It is defined by Thorndike:[2]

---

[1]Walter W. Cook in the Encyclopedia of Educational Research (Walter S. Monroe, ed.), New York: The Macmillan Company, 1950, p. 1473.

[2]Robert L. Thorndike in the Encyclopedia of Educational Research (Walter S. Monroe, ed.), New York: The Macmillan Company, 1950, p. 1016.

> The reliability of measurement has to do with the precision of a measurement procedure. Measurement in education is a process of estimating the amount of some quality or attribute possessed by individual objects or specimens. These estimates are usually expressed in numbers (scores) which correspond more or less accurately to the amount of the quality or trait in question.

Validity. In this research validity is held to be the degree to which the rating scale actually measures speaking skill. However, it is difficult to determine the validity of a speech rating scale because this requires some acceptable measure of the trait being rated as a basis for comparison. One of the commonly accepted measures of speaking skill is the critical response of the listener.

In discussing speech rating methods Monroe and others[3] point out:

> The problem of validity may be viewed first of all qualitatively. On logical grounds the audience response constitutes the ultimate practical criterion of the effectiveness of any speech. This granted, it follows that to the extent to which the judgments recorded by means of a rating scale are reliable, they are also valid.

Remmers,[4] who made a study of students' ratings of their teachers, states:

---

[3]Allan Monroe, Hermann H. Remmers, and Elizabeth Venemann-Lyle, "Measuring the Effectiveness of Public Speech in a Beginning Course," Studies in Higher Education, XXIX, Bulletin of Purdue University, 37:14, September, 1936.

[4]Hermann H. Remmers, "Reliability and Halo Effect of High School and College Students' Judgments of Their Teachers," Journal of Applied Psychology, 18:621, October, 1934.

> The problem of validity of judgments is hardly pertinent. While reliability may be defined as the accuracy with which a measuring instrument measures whatever it does measure, validity is defined as the extent to which the instrument measures what it purports to measure. Since it is student judgments that constitute the criterion, reliability and validity are in this case synonymous.

While the writer believes that this use of the word "synonymous" is inaccurate, he does agree with Carp[5] who, in discussing the validity of a speech rating form, points out:

> Agreement by experts is an accepted technique in establishing validity and it is therefore plausible to maintain that validity and reliability may be derived from the same index of agreement among judges.

Kelley[6] has treated validity of rating scales similarly saying:

> If competent judges appraise Individual A as being as much better than Individual B as Individual B is better than Individual C, then it is so, and there is no higher authority to appeal to.

Symonds[7] in discussing the validity of ratings says, "In a certain sense there is nothing more valid than a judgment." He goes on to point out that all our knowledge has its origin in observation and in interpretations made of observations.

---

[5]Bernard Carp, A Study of the Influence of Certain Personal Factors on a Speech Judgment. New Rochelle, New York: The Little Print, 1945, p. 113.

[6]Truman Lee Kelley, The Influence of Nurture Upon Individual Differences. New York: The Macmillan Company, 1926, p. 9.

[7]Symonds, op. cit., p. 108.

In the present study, since the student and faculty evaluators had discussed and agreed upon the speaking skills being evaluated, their ratings of the speakers were used as the criterion. Hence the validity of these ratings is determined by inference from the reliability of the ratings.

However, a second method, that of comparison with some other measure of speaking skill, was possible. All of the speakers were members of Fundamentals of Speech classes at the time that they participated in the intergroup speech projects. At the end of the course each student was given a mark (A, B, C, D, or E) by his speech instructor. This mark was to be regarded as indicative of the student's speech effectiveness in general. Each student also was given a letter mark derived from the total score received by adding the ratings given by the four evaluators. By using the method of random sampling,[8] a comparison was made of the marks that the students received in the intergroup speech projects with the marks they received for general effectiveness of speech.

As indicated in Table I, eighty-four per cent of the students received the same mark from their speech teacher as from the evaluator. Thus they disagreed on the marks of only sixteen per cent of the students. Of this sixteen per cent

---

[8]The method of random sampling used was that described by Everett F. Lindquist in "The Technique of Random Selection," Statistical Analysis in Educational Research, Boston: Houghton Mifflin Company, 1940, pp. 24-29.

TABLE I

A COMPARISON OF MARKS GIVEN BY EVALUATORS WITH MARKS GIVEN THE SAME STUDENTS BY TEACHERS

	Mark given by Evaluator of Speech Project	Mark given by Teacher of Speech Class		Percentage of Students with These Marks
Students with Identical Marks from both Evaluator and Teacher	A	A		2
	B	B		24
	C	C		38
	D	D		19
	E	E		1
			Total	84
Students who were Marked Lower by the Teacher than by Evaluator	A	B		0
	B	C		2
	C	D		1
	D	E		1
			Total	4
Students who were Marked Lower by the Evaluator than by the Teacher	B	A		1
	C	B		3
	D	C		7
	E	D		1
			Total	12

Note: In no case did the mark given by the evaluators of the intergroup speech project differ two degrees (i.e., A to C or C to A, etc.) from the mark given by the teacher.

the speech teacher gave four per cent of the students lower marks and twelve per cent of the students higher marks than did the evaluators of the intergroup speech projects.

Reliability. Reliability may be expressed by the extent to which two independent measurements will yield the same quantitative score. In the present study it was assumed that if the speech rating scale is a trustworthy device it should give approximately the same results when employed by evaluators having a certain minimum background of speech courses. Hence a calculation of the coefficient of correlation of the rating by pairs of judges should furnish an index of reliability of the scale.

In order to determine this coefficient of reliability the writer computed the correlation of the scores given by each pair of judges who rated a group of speakers. This correlation was extended by the use of the Spearman-Brown formula to include all judges.[9]

Since no machine exists for measuring speaking skills, any evaluative system involves some sort of human fallibility. This is substantiated by Shen:[10]

---

[9]Guilford, op. cit., p. 421. Also: E. L. Clark, "Spearman-Brown Formula Applied to Ratings of Personality Traits," Journal of Educational Psychology, October, 1935. pp. 552-55.

[10]Eugene Shen, "The Reliability Coefficient of Personal Ratings," Journal of Educational Psychology, 16:232, April, 1925.

> The reliability of mental tests is usually measured by correlation between results of two comparable tests. By analogy, the reliability of personal ratings may be evaluated by a correlation between ratings by two comparable judges. By pairing elements of two tests such that they are similar in difficulty and type, an author can to a certain extent insure the comparability between his tests. But the comparability of judges is much more precarious; it is entirely beyond the control of the investigator except by a meager selective function that he may fallibly exercise. On account of this uncontrollable variability of judges, a correlation between two judges is a very crude approximation of the reliability of either. The reliability of a judge thus crudely evaluated often varies considerably according to the judge with whom he happens to be correlated.

In this study the coefficient of reliability, when correlating ratings given by student evaluators with those given by faculty evaluators, was .61 for the first semester and .62 for the second semester. However, when additional evaluators participated, the coefficient of reliability was .68 for the student evaluators and .72 for the faculty evaluators. This, according to Slawson,[11] indicates very high reliability for the use of a scale rating personal traits.

B. The Distribution of Scores

A study of the distribution of scores is not only essential in order to measure the speech proficiency and

---

[11]John Slawson, "The Reliability of Judgments of Personal Traits," Journal of Applied Psychology, 6:161-71, April, 1922. Also Symonds, op. cit., p.95

degree of improvement made, but also worthwhile to provide background material for understanding the factors affecting these scores.

The scores of the speakers in each of the four intergroup speech projects were treated separately. Since the rating scale was constructed with a hundred points maximum, the highest possible total score that could be given to a student by adding the four evaluations given in any intergroup speech project would be four hundred. Actually during the 1952-53 academic year no one received a score over 371 or under 115. The range, mean, and standard deviation of the total scores in each of the intergroup speech projects may be seen in Tables II and III.

Since Fundamentals of Speech is a required course for all freshmen on the campus of Central Michigan College of Education, academic grades for each intergroup speech project were computed according to institutional policy.[12] This was done by plotting a curve and assigning marks as follows:

(1) "C's" were given to all of the scores within the range of a point one-half standard deviation below the mean and a point one-half standard deviation above the mean.

---

[12]Central Michigan College of Education, Faculty Handbook; A Summary of the More Important Policies, Regulations, and Procedures. Mt. Pleasant, Michigan, 1953, p. 58.

TABLE II

RANGE, MEAN, AND STANDARD DEVIATION OF TOTAL SCORES RECEIVED BY STUDENTS PARTICIPATING IN THE INTERGROUP SPEECH PROJECTS

GROUP I*

1952-53

	First Intergroup Speech Project	Second Intergroup Speech Project
Range	166-348	181-371
Mean	259	287
Standard Deviation	36	35

*Group I is comprised of the 305 students who participated in the intergroup speech projects first semester.

TABLE III

RANGE, MEAN, AND STANDARD DEVIATION OF TOTAL SCORES RECEIVED BY STUDENTS PARTICIPATING IN THE INTERGROUP SPEECH PROJECTS

GROUP II*

1952-53

	First Intergroup Speech Project	Second Intergroup Speech Project
Range	115-334	200-355
Mean	257	279
Standard Deviation	28.5	27.8

*Group II is comprised of the 244 students who participated in the intergroup speech projects second semester.

(2) "B's" were given to all of the scores between one half and one and a half plus standard deviations from the mean.

(3) "D's" were given to all of the scores between one half and one and a half minus standard deviations from the mean.

(4) "A's" were given to all of the scores on the plus end of the curve beyond one and a half standard deviations from the mean.

(5) "E's" were given to all of the scores on the minus end of the curve beyond one and a half standard deviations from the mean.

In a normal bell shaped curve this method would mean 38.30 per cent "C's," 24.17 per cent each for the "B's" and "D's," and 6.68 per cent each for the "A's" and the "E's."[13]

Group I. (First Semester Students). In order to avoid confusion of terms "first semester" and "second semester" with "first intergroup speech project" and "second intergroup speech project," first semester students will be referred to as Group I and second semester students as Group II. As indicated earlier in this study, Group I as well as Group II had two intergroup speech projects. There are three

---

[13]Harry W. Sundwall, "Normal Curve Score Probabilities." East Lansing: Michigan State College, 1950. (Mimeographed).

significant facts to notice regarding the distribution of scores of this Group I.

(1) From Table II it may be noticed that the range of scores in the second intergroup speech project, 181-371, was greater on both the lower and upper level than the range of scores in the first intergroup speech project. The mean rose from 259 in the first intergroup speech project to 287 in the second intergroup speech project with a mean improvement of 28 points per student. By applying the "t test" this was found to be statistically significant at the .01 level of confidence.[14] On the basis of this calculation the difference of 28 points between these mean gains could happen by chance less than once in a hundred times.

(2) Inspection of Table IV shows that 305 students in Fundamentals of Speech class participated in all of the performances in the experiment during the first semester. As indicated in this table, 80.7 per cent of these received a higher score while 19.3 per cent either received the same or a lower score in the second intergroup speech project. The greatest gain made by a student was 115 points while the greatest loss was 54 points.

(3) Table V demonstrates that it was necessary for a student to have a higher score in the second intergroup

---

[14]Garrett, op. cit., pp. 189-93.

TABLE IV

DIFFERENCES BETWEEN FIRST AND SECOND INTERGROUP SPEECH PROJECT SCORES

GROUP I

Point Differences between Scores in First and Second Speech Project	Number of Students Making This Gain (or Loss)	Per Cent of the Total Group
111 through 120	1	0.3
101 " 110	3	1.0
91 " 100	1	0.3
81 " 90	2	0.7
71 " 80	11	3.6
61 " 70	12	3.9
51 " 60	27	8.9
41 " 50	33	10.8
31 " 40	37	12.1
21 " 30	47	15.4
11 " 20	38	12.4
1 " 10	34	11.1
Total with higher scores in second speech project	246	80.7
- 9 through 0	28	9.6
-19 " -10	13	4.3
-29 " -20	9	2.9
-39 " -30	6	1.9
-49 " -40	0	0.0
-59 " -50	3	1.0
Total with lower scores in second speech project	59	19.3

TABLE V

TOTAL SCORES AND ACADEMIC RATINGS RECEIVED BY STUDENTS PARTICIPATING IN THE INTERGROUP SPEECH PROJECTS

GROUP I

Academic Rating Given	First Intergroup Speech Project	Second Intergroup Speech Project
A	331 or above	345 or above
B	278 - 330	304 - 353
C	241 - 277	270 - 303
D	187 - 240	220 - 269
E	186 or below	219 or below

speech project in order to receive the same mark as was given in the first intergroup speech project. This follows because grades were determined on the basis of the normal curve and the overall group improved. Thus:

(a) Scores between 331 and 344, which were equal to "A's" in the first project, were valued as "B's" in the second.

(b) Scores of 278-303, which were "B's" in the first project, were "C's" in the second.

(c) Scores in the "C" range in the first project, 241-269, were given "D" grades in the second project.

(d) Scores of 187 to 219 that had been "D" scores became "E's."

This was true because, as indicated earlier in this chapter, the grades were allotted according to standard deviations from the mean; and the mean of the second intergroup speech project was 28 points higher than the first.

Group II. (Second Semester Students). The findings that were evidenced by the data on Group I were substantially the same for Group II. They are as follows:

(1) There was a statistically significant improvement shown in the total scores received by Group II speakers in the second intergroup speech project when compared with the scores received in the first project. By comparing data in Table II with that in Table III one can see that the mean

score of the first intergroup speech project of Group II was 257 compared with 259, the mean score of Group I. The mean score of the second intergroup speech project of Group II was 279 compared to Group I's 287. Thus the mean improvement in points was 28 for Group II and 22 for Group I. Furthermore, it may also be noted from Table III that the range of scores was 115-334 in the first project and 200-355 in the second project. The same phenomena occurred with both groups, namely, the scores in the second intergroup speech projects were consistently higher than those in the first intergroup speech projects.

(2) As presented in Table VI, 195 (79 per cent) of the 244 Group II speakers who participated in the experiment had a higher score in the second project than they had in the first. This 79 per cent is comparable to the 80.7 per cent of Group I students who also made higher scores in the second projects.

(3) The results in Group II (presented in Table VII) were similar to those found in Group I, i.e., the over-all improvement gains necessitated higher scores by the speaker in the second intergroup speech project in order to maintain the same letter mark he earned in the first intergroup speech project.

C. Findings Regarding the Four Factors

The academic speech training of the evaluator. There was a statistically significant difference between the rating

TABLE VI

DIFFERENCES BETWEEN FIRST AND SECOND INTERGROUP SPEECH PROJECT SCORES

GROUP II

Point Differences between Scores in First and Second Speech Project	Number of Students Making This Gain (or Loss)	Per Cent of the Total Group
91 through 100	2	0.8
81 " 90	1	0.4
71 " 80	5	1.6
61 " 70	5	1.6
51 " 60	15	6.1
41 " 50	25	10.3
31 " 40	31	13.1
21 " 30	33	13.5
11 " 20	44	18.0
1 " 10	34	13.9
Total with higher scores in second speech project	195	79.0
- 9 " 0	21	8.6
-19 " -10	15	6.1
-29 " -20	7	2.9
-39 " -30	4	1.6
-49 " -40	2	0.4
Total with lower scores in second speech project	49	21.0

TABLE VII

TOTAL SCORES AND ACADEMIC RATINGS RECEIVED BY STUDENTS PARTICIPATING IN THE INTERGROUP SPEECH PROJECTS

GROUP II

1952-53

Academic Rating Given	First Intergroup Speech Project	Second Intergroup Speech Project
A	315 or above	355 or above
B	270 - 314	294 - 334
C	243 - 269	265 - 293
D	200 - 242	225 - 264
E	199 or below	224 or below

given by students, who were in either their third or fourth year of college, and the ratings given by faculty members with advanced degrees. This is indicated in Table VIII.

In rating 305 speakers (called "Group I" in this study) during the first semester of the 1952-53 academic year, the faculty evaluators gave a mean score of 62.08. The student evaluators, hearing the same speeches in the same room at the same time as the faculty evaluators, gave a mean score of 68.94. Thus the students' ratings averaged 11.1 per cent higher than the instructors' ratings. This is statistically significant at the five per cent level of confidence. In other words, this phenomenon of student evaluators rating 11.1 per cent higher than the faculty evaluators could happen by chance only five in a hundred times.[15]

In rating the 244 speakers (called "Group II" in this study) during the second semester of the 1952-53 academic year, the faculty evaluators gave a mean score of 58.89 while the student evaluators gave a mean score of 69.95. Thus, the students' ratings averaged 19.9 per cent higher than the instructors' ratings. This is statistically significant at the one per cent level of confidence.

The acquaintanceship of the evaluator with the speaker. Each of the twenty-one pairs of evaluators participating in

---

[15]Everett Franklin Lindquist, Statistical Analysis in Educational Research. Boston: Houghton Mifflin Company, 1940. p. 72.

TABLE VIII

A COMPARISON OF THE MEAN SCORES GIVEN
BY FACULTY AND STUDENT EVALUATORS
RATING THE SAME SPEAKERS

Group I*	
Mean Score Given by Student Evaluators	68.94
Mean Score Given by Faculty Evaluators	62.08
Difference between Student and Faculty Rating	6.86
Group II**	
Mean Score Given by Student Evaluators	69.95
Mean Score Given by Faculty Evaluators	58.89
Difference between Student and Faculty Rating	11.06

*N = 305

**N = 244

the fourth intergroup speech project, May 11-15, filled out an Evaluator's Acquaintanceship Check Sheet.[16] However, only eight cases occurred where one speaker was well known by a faculty evaluator and not known by a student evaluator while another speaker in the same group was well known by this student evaluator and not known by the faculty evaluator. These eight cases are presented in Table IX. It can be seen here that the students who knew the speakers gave a mean score two points higher than did the faculty members who did not know the speakers. However, when these same faculty members knew a speaker, they gave him a mean score seven points higher than did the student evaluators who were unacquainted with these speakers. Thus in both cases evaluators who knew the speakers rated them higher than did evaluators who did not know the speakers. This was especially true of faculty members.

Furthermore, as indicated in Table I speech teachers gave the same mark to eighty-five per cent of their students as was given by the evaluators. However, when these marks did differ, the teacher gave a higher rating than did the evaluators in three of every four cases. Since the teacher was acquainted with all of his students and the evaluators were acquainted with only about ten per cent of these speakers, one may assume that acquaintanceship was a positive

---

[16]See Appendix J.

TABLE IX

SCORES GIVEN WHEN THE STUDENT EVALUATOR IS ACQUAINTED AND THE FACULTY EVALUATOR IS UNACQUAINTED WITH THE SPEAKER COMPARED WITH THE SCORES GIVEN WHEN THE FACULTY EVALUATOR IS ACQUAINTED AND THE STUDENT EVALUATOR IS UNACQUAINTED WITH THE SPEAKER

Code for Pair of Evaluators	Scores Given When: Students Know Speakers	Faculty Do Not Know Speakers	Difference Acquainted Minus Unacquainted	Scores Given When: Faculty Know Speakers	Students Do Not Know Speakers	Difference Acquainted Minus Unacquainted
071-232	71	67	4	74	68	6
062-440	82	81	1	55	69	-14
042-102	73	65	8	48	64	-16
071-483	75	78	-3	75	63	12
071-493	70	81	-11	75	63	12
042-271	91	82	9	80	78	2
032-182	83	86	-3	86	81	5
011-412	62	65	-3	72	65	7
			2			7

*The coefficient of correlation (r) between the evaluators who know the speakers and those who do not know the speakers is .62.

factor in securing a higher rating.

The experience of the evaluator with the rating scale. In order to determine whether experience with the rating scale improved the reliability of the ratings, coefficients of correlation were computed on (1) the scores given by the raters when they first used the rating scale in October, and (2) the scores given by the same raters when, after some experience, they used the rating scale again in January. These coefficients of correlation of ratings given by pairs of judges in October were compared with those of ratings given by the same pairs of evaluators rating the same speakers three months later. As indicated in Table X, there was no evidence of a significant difference between the experienced and inexperienced evaluators. Ten pairs of evaluators showed a mean increase in correlation of fourteen points. However, nine pairs of evaluators showed a mean decrease in correlation of fifteen points. The correlations of the ratings given by two pairs of evaluators remained substantially the same.

The sex of the evaluator in relation to the sex of the speaker. A comparison was made of the mean scores given female speakers and male speakers by four combinations of judges, i.e., (1) a male faculty member judging with a male student evaluator, (2) a female faculty member judging with a female student evaluator, (3) a male faculty member judging

TABLE X

A COMPARISON OF THE CORRELATION OF RATINGS GIVEN BY STUDENT AND FACULTY EVALUATORS JUDGING THE SAME GROUP OF STUDENTS THREE MONTHS APART

Code[X]	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
r* of October Ratings	.42	.40	.64	.37	.94	.52	.76	.66	.81	.61	.48	.62	.47	.76	.51	.74	.52	.56	.76	.82	.47
r* of January Ratings	.82	.57	.59	.76	.74	.39	.35	.77	.82	.44	.38	.63	.54	.75	.44	.54	.52	.83	.81	.82	.40
Increase in r*			.05		.20	.13	.41			.17	.10			.01	.07	.20					.07
Decrease in r*	.40	.17		.39				.11	.01			.01	.07					.27	.05		

[X]Each letter stands for a pair of evaluators, one student and one faculty member, who rated a group of speakers in October and then rated the same group of students, giving different speeches, in January.

*The symbol "r" is used to designate the coefficient of correlation.

with a female student evaluator, and (4) a female faculty member judging with a male student evaluator. This comparison was made for both first semester (Group I) and second semester (Group II) speakers. The data are outlined in Tables XI, XII, XIII, and XIV.

The greatest difference between ratings given by pairs of evaluators occurred when a female faculty evaluator and a female student evaluator rated together. This was true in rating both male and female speakers and occurred in Group II as well as in Group I. Rating female speakers, the female student evaluators gave an average of 9.66 points higher the first semester (Table XI) and 16.06 points higher the second semester (Table XIII) than did the female faculty evaluators. Rating male speakers, the female student evaluators gave an average of 10.61 points higher the first semester (Table XII) and 14.22 points higher the second semester (Table XIV) than did the female faculty evaluators.

The least difference between ratings given by pairs of evaluators occurred when a female faculty evaluator and a male student evaluator rated together. In rating first semester students the female faculty and male student evaluators differed 3.15 points in rating female speakers and only .21 in rating male speakers.

TABLE XI

A COMPARISON OF THE MEAN SCORES GIVEN
BY FACULTY AND STUDENT EVALUATORS
RATING FEMALE SPEAKERS
GROUP I

Sex of Faculty Evaluator	Sex of Student Evaluator	Mean Score Given by Faculty	Mean Score Given by Students	Difference of the Mean Scores
Male	Male	62.40	67.85	5.45
Female	Female	63.66	73.22	9.66
Male	Female	63.00	69.95	6.95
Female	Male	65.57	68.72	3.15
All Four Combinations of Judges Listed Above		63.66	69.94	6.28

TABLE XII

A COMPARISON OF THE MEAN SCORES GIVEN
BY FACULTY AND STUDENT EVALUATORS
RATING THE MALE SPEAKERS
GROUP I

Sex of Faculty Evaluator	Sex of Student Evaluator	Mean Score Given by Faculty	Mean Score Given by Students	Difference of the Mean Scores
Male	Male	57.20	62.50	5.30
Female	Female	63.06	73.67	10.61
Male	Female	58.06	67.89	9.83
Female	Male	64.68	64.89	.21
All Four Combinations of Judges Listed Above		60.75	67.24	6.49

TABLE XIII

A COMPARISON OF THE MEAN SCORES GIVEN
BY FACULTY AND STUDENT EVALUATORS
RATING FEMALE SPEAKERS
GROUP II

Sex of Faculty Evaluator	Sex of Student Evaluator	Mean Score Given by Faculty	Mean Score Given by Students	Difference of the Mean Scores
Male	Male	64.39	70.43	6.04
Female	Female	56.98	73.04	16.06
Male	Female	57.14	68.84	11.70
Female	Male	59.47	71.46	11.99
All Four Combinations of Judges Listed Above		59.50	70.94	11.44

TABLE XIV

A COMPARISON OF THE MEAN SCORES GIVEN
BY FACULTY AND STUDENT EVALUATORS
RATING MALE SPEAKERS
GROUP II

Sex of Faculty Evaluator	Sex of Student Evaluator	Mean Score Given by Faculty	Mean Score Given by Students	Difference of the Mean Scores
Male	Male	62.26	66.39	4.13
Female	Female	55.09	69.31	14.22
Male	Female	57.14	68.84	11.70
Female	Male	58.81	67.26	8.45
All Four Combinations of Judges Listed Above		58.33	67.95	9.62

D. Summary

An analysis of the data indicates that:

(1) The coefficient of reliability of the rating scale ranged from .61, when used by pairs of evaluators comprised of one student and one faculty member, to .72, when used by pairs of faculty members exclusively.

(2) A degree of validity was evidenced in that eighty-four per cent of the students received the same marks from their speech teacher as from the evaluators.

(3) Although the distribution of scores showed a range of 115 to 371 points, a consistent pattern of higher scores during the second intergroup speech projects was evidenced both semesters.

(4) The students' ratings averaged 11.1 per cent higher than the instructors' ratings the first semester and 19.9 per cent higher than the instructors' ratings the second semester.

(5) In general, when faculty members knew a speaker, they gave him a mean score seven points higher than did the student evaluators who were unacquainted with the speaker.

(6) When the marks that the speech teachers gave their students differed from the marks given these students by evaluators, the speech teacher who knew the speakers involved gave them higher ratings seventy-five per cent of the time.

(7) Ten pairs of evaluators showed a mean increase in correlation of fourteen points between October and January

ratings. However, nine pairs of evaluators showed a mean decrease in correlation of fifteen points during the same period.

(8) The greatest difference between ratings given by pairs of evaluators occurred when a female faculty evaluator and a female student evaluator rated together.

(9) The least difference between ratings given by pairs of evaluators occurred when a female faculty evaluator and a male student evaluator rated together.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

In this concluding chapter the principal findings of the research will be summarized and the implications for education noted. Although these will be pertinent primarily to the evaluation of students in Fundamentals of Speech at Central Michigan College of Education, they may also have significance in comparable appraisal situations elsewhere. In the light of this investigation additional problems which merit study have been disclosed.

A. Principal Findings

In Chapter IV of this thesis some of the interesting facts regarding the rating scale and the distribution of scores have been presented as incidental to the main objectives of the investigation.

The principal findings relate to the questions raised in Chapter I regarding the role of the four selected factors in the evaluation of speech performance: the academic speech training of the evaluator, the acquaintanceship of the evaluator with the rating scale, the experience of the evaluator with the rating scale, and the sex of the evaluator in relation to the sex of the speaker. Taking these up one by one the results are as follows:

The academic speech training of the evaluator. Ratings given by college juniors and seniors who were speech majors or minors differed significantly from those given by members of the faculty who had advanced degrees in speech. Pairs of student evaluators as well as pairs of faculty evaluators had a higher correlation of scores than did pairs of evaluators comprised of one student and one faculty member. Furthermore, the fact that students gave a mean score 15.5 per cent higher than did the faculty evaluators cannot be easily attributed to chance when one considers: (1) that a total of 4392 comparable ratings were studied, (2) that in all cases student and faculty evaluators heard the speeches at the same time, and (3) that there was a rather consistent pattern of student evaluators giving higher ratings than faculty evaluators, ranging from a mean difference of 11.1 per cent the first semester to 19.9 per cent the second semester.

The acquaintanceship of the evaluator with the speaker. Ratings given by the evaluators acquainted with the speaker differed somewhat, but not significantly, from those given by the evaluators unacquainted with the speaker. Although the results are not statistically significant, they are in agreement with the results of published research by Knight[1] and

---

[1]Knight, op. cit., p. 141.

Henrickson,[2] i.e., that evaluators who are acquainted with the speakers give them higher ratings than do evaluators who are unacquainted with these speakers:

Thus, although in general student evaluators gave higher ratings than did faculty evaluators, this relation was reversed when the faculty evaluators were acquainted with the speakers. Furthermore, when the mark given by the teacher of the Fundamentals of Speech class differed from the mark computed from the four evaluations, the teacher's mark was higher than the evaluators' mark seventy-five per cent of the time. Since the teacher was better acquainted with most of his speech students than were the evaluators, one may assume that the factor of acquaintanceship influenced these ratings.

The experience of the evaluator with the rating scale. Ratings given by the evaluators before they had experience with the rating scale did not differ significantly from those given by the evaluators after they had experience with the rating scale. A possible explanation of this may be:

(1) All of the evaluators had some speech training before using the Evaluator's Rating Scale for the first time. This scale evaluated on the customary criteria so there was little new or different even when using the rating scale for the first time. Furthermore, the interpretations of the

---

[2]Henrickson, op. cit., p. 124.

criteria were agreed upon at two discussion meetings previous to the first intergroup speech project.

(2) Student and faculty evaluators did not compare ratings at any time. Thus an evaluator did not know whether he was giving higher or lower scores than his colleague and could not, of course, profit by experience and adjust to his fellow evaluator's ratings.

The sex of the evaluator in relation to the sex of the speaker. Ratings of speakers of each sex differed significantly according to the sex of the evaluator. Specifically these differences are:

(1) Female student evaluators gave higher ratings to both male and female speakers than did male student evaluators.

(2) Female student evaluators gave relatively higher ratings to male speakers than they gave to female speakers.

(3) Male student evaluators gave relatively higher ratings to female speakers than they gave to male speakers.

(4) Male and female faculty evaluators gave ratings substantially the same. However, female faculty evaluators slightly favored male speakers, and male faculty evaluators slightly favored female speakers.

Furthermore, when two female evaluators rated together, they differed from each other more than did any other combination of evaluators. This was true probably because, in

general, the most lenient ratings were given by female student evaluators and the most severe ratings by the female faculty evaluators.

B. Educational Implications

Significance for Central Michigan College of Education. This study has important significance for Central Michigan College of Education. From the analysis of the data it is evident that, if one wishes to have comparable scores when rating speech performances, evaluators ought not be chosen at random. This is true even when the raters have had training in college speech courses.

Two rules may be formulated for the selection of evaluators for future intergroup speech projects.

(1) Since ratings given by upperclassmen, who are speech majors and minors, tend to differ significantly from the scores given by members of the speech faculty, speakers who are to be compared ought to be rated exclusively by students, exclusively by faculty members, or by a like number of each.

(2) Since the student evaluators of both sexes tend to evaluate male speakers differently than they evaluate female speakers, this factor must be taken into consideration when assigning evaluators if the scores are to be comparable. In order to get more meaningful scores, an equal number of male and female student evaluators should evaluate each speaker.

Significance for education in general. To make broad general statements from a specific study often results in inaccuracies. However, one available criterion for the acceptance of a conclusion is the consistency of results from different investigations. A survey of the experimental literature on differences of evaluators according to the factor of sex reveal certain major findings. These findings are so frequently repeated by different investigations as to suggest a valid basis for a generalization. The conclusion that Carter[3] arrived at in studying marks assigned by teachers of beginning algebra, the findings that Douglas and Newman[4] announced after studying the marks of students in four Minnesota high schools, the results of the study by Webb[5] on rating intelligence, and the conclusions of the study made by the writer[6] regarding the attitudes of intercollegiate debaters towards debate judges, all substantiate the findings of this study, i.e., that female evaluators rate differently than do male evaluators. Thus it may be concluded that the sex of the evaluator is a factor to be taken into consideration when attempting to make objective and valid evaluations not only in speech but in education in general.

---

[3]Carter, op. cit., p. 93.

[4]Douglas and Newman, op. cit., pp. 86-87.

[5]Webb, op. cit., pp. 252-53.

[6]Pfister, op. cit., pp. 102-03.

Furthermore, the fact that juniors and seniors who were speech majors or minors gave significantly higher ratings than did faculty evaluators may quite logically have implications for other areas of education. However, these cannot be stated conclusively until further research is done.

C. Suggestions for Further Study

The possible means of analyzing the data collected in this study have not been exhausted. Since the pertinent items have been recorded on IBM punch cards, they could easily be sorted, arranged, and tabulated in various ways to furnish evidence in answer to other questions.

One of these questions might be: Is speech skill evaluated by techniques describing specific factors more reliable than a Gestalt "general effectiveness" rating? This could be partially answered by further statistical study of the 4392 ratings secured in this experiment. The scores given on the four traits (thought, language, voice, and action) could be compared to the scores given the same speakers on "general impression and communicativeness."

Another question might be: What role, if any, does the "halo effect," described by Thorndike,[7] have on these ratings of speech performances? This could be answered by a

---

[7]Edward L. Thorndike, "Constant Error in Psychological Rating," Journal of Applied Psychology, 4:25-29, March, 1920.

study of the sub-traits to determine the extent of this constant error.

Although this study found a significant difference between ratings given by student and faculty evaluators, it did not attempt to make a study of the complexity of factors inherent in the labels "student" and "faculty." In an institution such as Central Michigan College of Education some "students," especially juniors and seniors who are speech majors or minors, do supervised teaching in the secondary schools while some "faculty" are often students in graduate courses in speech and education leading to a higher degree. Although the students and faculty in this study differ in their amount of academic speech training, a further study might examine whether this is the most distinguishable characteristic.

Furthermore, in a few years an interesting follow-up study on the evaluators might be made. Since most of the evaluators either are or will be teachers, the factor of age and teaching experience could be considered and perhaps find answers to such questions as: (1) Do teachers who rate severely become more lenient as they get older and more experienced? and (2) Do teachers who give lenient ratings become more severe raters as time goes on?

More data regarding the student evaluators are available in the files of the Student Personnel Division at Central Michigan College of Education. This could be used

with the data already tabulated on the IBM cards to determine whether there is any substantiation for the hypothesis that students who make better semester grades in speech courses are more reliable raters than their fellow students.

If it is at all possible, a study making use of some other experimental design should investigate specifically the role that the factor of acquaintanceship has upon the ratings of speech performances. There is, of course, a great deal of difficulty involved in conducting accurate research where emotional reactions are concerned. Nevertheless, if the reaction (favorable or unfavorable) of the acquaintanceship could be considered, results may be more meaningful.

D. Summary

Although there was much concomitant learning incident to this study, the research was primarily concerned with determining the influence of four selected factors on the ratings of speech performances.

The findings of this research led to the conclusion that the academic speech training of the evaluator influences the ratings that he gives. Undergraduate evaluators with majors or minors in speech gave significantly higher ratings than did evaluators with advanced degrees in speech. Furthermore, scores given by pairs of undergraduate evaluators had a higher correlation than did scores given by student-faculty pairs of evaluators. Pairs of evaluators with advanced

degrees in speech had the highest correlation.

The investigation of the influence that the factor of acquaintanceship has in rating speech performances was inconclusive. However, the results of this study tend to substantiate the findings of previous research, i.e., that evaluators who are acquainted with the speakers give them higher ratings than do evaluators who are unacquainted with these speakers.

This study found that the experience of the evaluator with the particular rating scale did not significantly influence the ratings. However, all the evaluators had a certain minimum of speech training and had rated speeches previously.

The literature and data of this study support the contention that male and female evaluators rate male and female speakers differently. This is evidence that the factor of sex does influence the ratings of speech performances.

Probably the most important conclusion that the writer has arrived at is a philosophic one. Through the experiences gained from planning, executing, and evaluating this study, he realizes that by attempting to secure answers to specific questions one raises many more questions that need to be answered. Thus the research started in this thesis may continue indefinitely.

BIBLIOGRAPHY

A. BOOKS

B. PERIODICALS

C. BULLETINS AND MONOGRAPHS

D. UNPUBLISHED MATERIALS

BIBLIOGRAPHY

A. BOOKS

Anastasi, Ann. Differential Psychology; Individual and Group Differences in Behavior. New York: The Macmillan Company, 1937. 615 pp.

Baird, A. Craig, and Franklin H. Knower. General Speech: An Introduction. New York: McGraw-Hill Book Company, 1949. 500 pp.

Barnes, Harry G. Speech Handbook. Iowa City: Privately printed, 1936. 138 pp.

Beach, Ann F., and others. Bibliography on the Use of IBM Machines in Science, Statistics, and Education. New York: International Business Machines Corporation, January, 1954. 60 pp.

Brownell, William A., and others. The Forty-Fifth Yearbook of the National Society for the Study of Education: Part I, The Measurement of Understanding. Chicago: University of Chicago Press, 1946. 338 pp.

Campbell, William Giles. Form and Style in Thesis Writing. New York: Houghton Mifflin Company, 1954. 114 pp.

Carp, Bernard. A Study of the Influence of Certain Personal Factors on a Speech Judgment. New Rochelle, New York: The Little Print, 1945. 122 pp.

Craig, Alice Evelyn. The Speech Arts. New York: The Macmillan Company, 1947. 499 pp.

Fisher, Ronald Aylmer. The Design of Experiments. New York: Hafner Publishing Company, 1951. 244 pp.

Freeman, Graydon LaVern. The Energetics of Human Behavior. Ithaca, New York: Cornell University Press, 1948. 344 pp.

Friederich, Willard J. and Ruth Wilcox. Teaching Speech in High Schools. New York: The Macmillan Company, 1953. 487 pp.

Garrett, Henry E. Statistics in Psychology and Education. New York: Longmans, Green and Company, 1947. 487 pp.

Good, Carter V., A. S. Barr, and Douglas E. Scates. The Methodology of Educational Research. New York: D. Appleton-Century Crofts, 1941. 890 pp.

Guilford, J. P. Psychometric Methods. New York: McGraw-Hill Book Company, 1936. 566 pp.

Hedde, Wilhelmina G., and William Norwood Brigance. American Speech. New York: J. B. Lippincott Company, 1947. 596 pp.

Hollingworth, Harry L. The Psychology of the Audience. New York: American Book Company, 1935. 232 pp.

Jordan, Arthur Melville. Measurement in Education. New York: McGraw-Hill Book Company, 1953. 533 pp.

Kelley, Truman Lee. Scientific Method: Its Function in Research and in Education. New York: The Macmillan Company, 1932. 233 pp.

_____. The Influence of Nurture upon Individual Differences. New York: The Macmillan Company, 1926. 49 pp.

Knower, Franklin H. Table of Contents of the Quarterly Journal of Speech (1915-1952); Speech Monographs (1934-1952); and the Speech Teacher (1952) with a Revised Index Compiled Through 1952. Columbia, Missouri: Speech Association of America, 1953. 61 pp.

Laird, Donald A. The Psychology of Selecting Men. New York: McGraw-Hill Book Company, 1927. 269 pp.

Lindquist, Everett Franklin. Statistical Analysis in Educational Research. Boston: Houghton Mifflin Company, 1940. 266 pp.

Monroe, Walter S. (ed.). Encyclopedia of Educational Research. Revised edition. New York: The Macmillan Company, 1950. 1520 pp.

Murray, Elwood. The Speech Personality. New York: J. B. Lippincott Company, 1944. 565 pp.

Ogg, Helen L. and Ray K. Immel. Speech Improvement. New York: F. S. Crofts and Company, 1936. 190 pp.

Pease, Katharine. Machine Computations of Elementary Statistics. New York: Chartwell House, Inc., 1949. 208 pp.

Reid, Loren D. Teaching Speech in High School. Columbia, Missouri: Artcraft Press, 1952. 301 pp.

Robinson, Karl F. Teaching Speech in Secondary School. New York: Longmans, Green and Company, 1951. 438 pp.

Rugg, Harold O. Statistical Methods Applied to Education. New York: Houghton Mifflin Company, 1917. 410 pp.

Seward, Georgene H. Sex and the Social Order. New York: McGraw-Hill Book Company, 1946. 301 pp.

Snedecor, George W. Statistical Methods. Ames: The Iowa State College Press, 1946. 485 pp.

Symonds, Percival M. Diagnosing Personality and Conduct. New York: D. Appleton-Century Company, 1931. 602 pp.

Terman, Lewis M., and Catharine Cox Miles. Sex and Personality. New York: McGraw-Hill Book Company, 1936. 600 pp.

Thonssen, Lester W., and Elizabeth Fatherson. Bibliography of Speech Education. New York: H. W. Wilson and Company, 1939. 800 pp.

Thonssen, Lester W., Mary Margaret Robb, and Dorthea Thonssen. Bibliography of Speech Education - Supplement 1939-48. New York: H. W. Wilson and Company, 1950. 393 pp.

Weaver, Andrew T., Gladys Borchers, and Donald K. Smith. The Teaching of Speech. New York: Prentice-Hall, Incorporated, 1953. 565 pp.

B. PERIODICALS

Anderson, Gordon V., and Royal B. Embree. "Appraisal of the Individual," Review of Educational Research 18:157-74, April, 1948.

Bartlett, Neil R. "A Punched-Card Technique for Computing Means, Standard Deviations, and the Product-Moment Correlation Coefficient and for Listing Scattergrams," Science, 104:374-75, October 18, 1946.

Bingham, Walter V. "Halo, Invalid and Valid," Journal of Applied Psychology, 23:221-28, April, 1939.

Brownell, William A. "Some Neglected Safeguards in Control-Group Experimentation," Journal of Educational Research, 27:98-107, October, 1933.

Bryan, Alice J., and Walter H. Wilke. "A Scale for Measuring Speaking Abilities," Psychological Bulletin, 33:605-06, October, 1936.

_____. "Audience Tendencies in Rating Public Speeches," Journal of Applied Psychology, 26:371-81, June, 1942.

_____. "A Technique for Rating Speeches," Journal of Consulting Psychology, 5:80-90, March-April, 1941.

Cable, Arthur W. "A Criticism Card for Class Use," Quarterly Journal of Speech Education, 12:186-88, April, 1926.

Carroll, Robert P. "Practice in Rating," Journal of Experimental Psychology, 14:299-302, June, 1931.

Carter, Robert Scriven. "Non-intellectual Variables Involved in Teachers' Marks," Journal of Educational Research, 47:81095, October, 1953.

Cason, Hulsey. "An Annoyance Test and Some Research Problems," Journal of Abnormal and Social Psychology, 25: 224-36, July-September, 1930.

Castore, George F., and William S. Dye. "A Simplified Punch Card Method of Determining Sums of Products," Psychometrika, 14:243-50, September, 1949.

Champney, Horace, and Helen Marshall. "Optimal Refinement of the Rating Scale," Journal of Applied Psychology, 23:323-31, 1939.

Clark, Edward L. "Spearman-Brown Formula Applied to Ratings of Personality Traits," Journal of Educational Psychology, October, 1935, pp. 552-55.

DiVesta, Francis J., and James H. C. Roach. "An Analysis of a Procedure Used in the Teaching and Rating of Oral Expression," Journal of Educational Psychology, 42:468-78, December, 1951.

Dorcus, Roy M. "Some Factors Involved in Judging Personal Characteristics," Journal of Applied Psychology, 10:502-18, December, 1926.

Douglas, Harl R., and Olson E. Newman. "The Relation of High School Marks to Sex in Four Minnesota Senior High Schools," School Review, 45:481-88, April, 1937.

Drushal, J. Garber. "An Objective Analysis of Two Techniques of Teaching Delivery in Public Speaking," Quarterly Journal of Speech, 25:561-69, December, 1939.

Dudycha, George J. "A Note on the 'Halo Effect' in Ratings," Journal of Social Psychology, 15:331-33, May, 1942.

Englehart, Max D. "Suggestions with Respect to Experimentation under School Conditions," Journal of Experimental Education, 14:225-44, March, 1946.

Freyd, Max. "The Graphic Rating Scale," Journal of Educational Psychology, 14:83-102, February, 1923.

Furfey, Paul H. "An Improved Rating Scale Technique," Journal of Educational Psychology, 17:45-48, January, 1926.

Gauger, Paul W. "Comparison of the Abilities of High School Speech Students and Speech Experts in Rating a Speech Performance," Journal of Educational Research, 42:209-17, November, 1948.

Gilkinson, Howard, and Franklin H. Knower. "A Study of Standardized Personality Tests and Skill in Speech," Journal of Educational Psychology, 32:161-75, 1941.

Gray, J. Stanley. "Objective Measurements for Public Speaking," Journal of Expression, 2:20-26, March, 1928.

Hart, Hornell N., and Elmer Olander. "Sex Differences in Character as Indicated by Teachers' Ratings," School and Society, 20:381-82, September 6, 1924.

Hawthorne, Joseph W. "An Attempt to Measure Certain Phases of Speech," Journal of General Psychology, 10:399-414, April, 1934.

Hayworth, Donald. "Tests and Measurements in Public Speaking," Quarterly Journal of Speech, 21:272-75, November, 1935.

Holcomb, Martin J. "The Critic-Judge System," Quarterly Journal of Speech, 19:28-38, February, 1933.

Hollister, Richard. "Faculty Judging," Quarterly Journal of Public Speaking, 3:235-41, April, 1917.

Holm, James N. "A Progressive Form of the Speech Criticism Blank," Speaker of Tau Kappa Alpha, 22:12-13, March, 1938.

Hudgins, Clarence V. "Concerning the Validity of Speech Tests," Volta Review, 45:271-72, May, 1943.

Jurgensen, C. E. "Intercorrelations in Merit Rating Traits," Journal of Applied Psychology, 34:240-43, August, 1950.

Kincheloe, Isabel. "On Refining the Speech Scales," English Journal, 34:204-07, April, 1945.

Klein, Ruth, and William McLamers. "Standards in Public Speaking," Quarterly Journal of Speech, 22:439-42, October, 1936.

Knight, Frederic B. "The Effect of Acquaintance Factor upon Personal Judgments," Journal of Educational Psychology, 14:129-42, March, 1923.

Knower, Franklin H. "A Suggestive Study of Public Speaking Rating-Scale Values," Quarterly Journal of Speech, 15:30-41, February, 1929.

_____. "Experimental Study of Changes in Attitudes," Journal of Social Psychology, 6:315-44, August, 1935.

_____. "Psychological Tests in Public Speaking," Quarterly Journal of Speech, 15:216-22, April, 1929.

_____. "What Is a Speech Test?" Quarterly Journal of Speech, 30:485-93, December, 1944.

Kornhauser, Arthur W. "A Comparison of Raters," Journal of Personnel Research, 6:338-44, January, 1927.

_____. "A Comparison of Ratings on Different Traits," Journal of Personnel Research, 6:440-46, March, 1927.

_____. "Reliability of Average Ratings," Journal of Personnel Research, 5:309-17, December, 1926.

Lehman, Harvey C., and Paul A. Witty. "Sex Differences: Some Sources of Confusion and Error," American Journal of Psychology, 42:140-47, January, 1930.

Lewis, Helen Bruck. "Studies in the Principles of Judgments and Attitudes," Journal of Social Psychology, 11:121-46, February, 1940; and 14:229-55, August, 1941.

Lorge, Irving. "Tabulating and Test-Scoring Machines: Applications of International Business Machines to Educational Research," Review of Educational Research, 12: 550-57, December, 1942.

Mahler, W. R. "Some Common Errors in Employee Rating Practices," Personnel Journal, 26:68-74, June, 1947.

Marsh, S. E., and F. A. C. Perrin. "An Experimental Study of the Rating Scale Technique," Journal of Abnormal and Social Psychology, 19:383-99, January-March, 1925.

Miner, James Burt. "The Evaluation of a Method for Finely Graduated Estimates of Ability," Journal of Applied Psychology, 1:123-33, June, 1917.

Moore, Wilbur E. "Factors Related to Achievement and Improvement in Public Speaking," Quarterly Journal of Speech, 29:213-17, April, 1943.

Newcomb, Theodore. "An Experiment Designed to Test the Validity of a Rating Technique," Journal of Educational Psychology, 22:27-89, April, 1931.

Noll, Victor H. "Measuring the Scientific Attitude," Journal of Abnormal and Social Psychology, 30:145-54, July-September, 1935.

Pelsma, J. R. "Standardization of Grades in Public Speaking," Quarterly Journal of Public Speaking, 1:266-71, October, 1915.

Pemberton, H. Earl. "A Technique for Determining the Optimum Rating Scale for Opinion Measures," Sociology and Social Research, 17:470-72, May-June, 1933.

Pfister, Emil R. "A Survey of Attitudes Toward Debate Judges," Forensic of Pi Kappa Delta, 39:102-03, May, 1954.

_____. "Ratings, Rankings, or Both?" Speech Activities, 8:25, Spring, 1952.

Remmers, Hermann H. "Reliability and Halo Effect of High School and College Students' Judgments of Their Teachers," Journal of Applied Psychology, 18:619-30, October, 1934.

Richards, T. W., and Willis Ellington. "Objectivity in the Evaluation of Personality," Journal of Experimental Education, 10:228-37, June, 1942.

Rose, Forest H. "Training in Speech and Changes in Personality: A Study of the Effects of Beginning Speech Courses upon Personality Traits," Quarterly Journal of Speech, 26:193-96, April, 1940.

Rugg, Harold O. "Is the Rating of Human Character Practicable?" Journal of Educational Psychology, 12:425-38, November, 1921; 485-501, December, 1921; 13:30-42, January, 1922; 81-93, February, 1922.

Sapir, Edward. "Speech as a Personality Trait," American Journal of Sociology, 32:892-905, May, 1927.

Schneider, Dorothy E., and A. F. Bayroff. "Relationship between Rater Characteristics and Validity of Ratings," Journal of Applied Psychology, 37:278-80, August, 1953.

Seedorf, Evelyn H. "An Experimental Study in the Amount of Agreement Among Judges in Evaluating Oral Interpretation," Journal of Educational Research, 43:10-21, September, 1949.

Shen, Eugene. "The Influence of Friendship upon Personal Ratings," Journal of Applied Psychology, 9:66-68, March, 1925.

_____. "The Reliability Coefficient of Personal Ratings," Journal of Educational Psychology, 16:232-37, April, 1925.

Slawson, John. "The Reliability of Judgments of Personal Traits," Journal of Applied Psychology, 6:161-71, April, 1922.

Smith, Marpheus. "Group Judgments in the Field of Personality Traits," Journal of Experimental Psychology, 14: 562-65, October, 1931.

Stagner, Ross. "Judgments of Voice and Personality," Journal of Educational Psychology, 27:272-77, April, 1936.

Stevens, Wilmer E. "A Rating Scale for Public Speakers," Quarterly Journal of Speech, 14:223-32, April, 1928.

Stinchfield, Sarah M. "The Standardization of Speech Testing Material," Quarterly Journal of Speech Education, 7:360-69, November, 1921.

Strang, Ruth. "Seven Ways to Improve the Rating Process," Occupations, 29:107-10, November, 1950.

Strong, Edward K., Jr. "Weighted vs. Unit Scales," Journal of Educational Psychology, 36:193-216, April, 1945.

Symonds, Percival M. "Changes in Sex Differences in Problems and Interests of Adolescents with Increasing Age," Journal of Genetic Psychology, 50:83-89, March, 1937.

_____. "Notes on Rating," Journal of Applied Psychology, 9:188-95, June, 1925.

_____. "On the Loss of Reliability Due to Coarseness of the Scale," Journal of Experimental Psychology, 7:456-60, December, 1924.

Taylor, Harold C. "Social Agreement on Personality Traits as Judged from Speech," Journal of Social Psychology, 5:244-48, May, 1934.

Thompson, Wayne N. "Is There a Yardstick for Measuring Speaking Skill?" Quarterly Journal of Speech, 29:87-91, February, 1943.

Thorndike, Edward L. "A Constant Error in Psychological Rating," Journal of Applied Psychology, 4:25-29, March, 1920.

Tiffin, Joseph, and Wayne Musser. "Weighting Merit Rating Items," Journal of Applied Psychology, 26:575-83, October, 1942.

Webb, Louis W. "The Ability of Men and Women to Judge Intelligence," School and Society, 20:251-54, August 23, 1924.

Weber, Pearl L. "Judgment Today," Psychological Review, 44: 264-66, May, 1937.

West, Robert, and Helen Larsen. "Some Statistical Investigations in the Field of Speech," Quarterly Journal of Speech Education, 7:375-82, November, 1921.

Wilke, Walter H. "The Reliability of Summaries of Rating Scale Evaluation of Student Personality Traits," Journal of Genetic Psychology, 53:313-20, February, 1938.

_____. "A Subjective Measurement in Speech: A Note on Method," Quarterly Journal of Speech, 21:53-59, February, 1935.

C. BULLETINS AND MONOGRAPHS

Auer, J. Jeffery. "Doctoral Dissertations in Speech: Work in Progress, 1954," Speech Monographs, 21:136-41, 1954.

Carter, Gerald C. "Student Personalities as Instructors See Them," Bulletin of Purdue University: Studies in Higher Education. Lafayette: Purdue University, 1945. 46 pp.

Central Michigan College of Education. Bulletin, 1952-53 Sessions. Mount Pleasant, Michigan: Central Michigan College of Education, 1952. 256 pp.

Henrickson, Ernest H. "The Relation Among Knowing a Person, Liking a Person, and Judging Him as a Speaker," Speech Monographs, 14:22-25, 1940.

Knower, Franklin H. "Graduate Theses--An Index of Graduate Work in Speech," Speech Monographs, 21:108-35, 1954.

McNemar, Quinn, and Lewis M. Terman. "Sex Differences in Variational Tendency," Genetic Psychology Monographs, 18:1-65, February, 1936.

Monroe, Allan H. "Evaluation in Speech Education," Bulletin of the National Association of Secondary School Principals, 29:156-64, November, 1945.

_____. "The Measurement and Analysis of Audience Reaction to Student Speakers--Studies in Attitude Changes," Bulletin of Purdue University: Studies in Higher Education. Lafayette: Purdue University, 1937. 80 pp.

Monroe, Allan H., Hermann H. Remmers, and Elizabeth Venemann-Lyle. "Measuring the Effectiveness of Public Speech in a Beginning Course," Bulletin of Purdue University: Studies in Higher Education. Lafayette: Purdue University, 1936. 29 pp.

Norvelle, Lee. "Development and Application of a Method for Measuring the Effectiveness of Instruction in a Basic Speech Course," Speech Monographs, 1:41-65, 1934.

Rosenberg, Ralph P. "Bibliographies of Theses in America," Bulletin of Bibliography, 18:181-82, September-December, 1945.

Thompson, Wayne N. "A Study of the Characteristics of Student Raters of Public Speaking Performances," Speech Monographs, 13:45-53, 1946.

D. UNPUBLISHED MATERIALS

Anderson, Mary Margaret. "An Analysis of Some of the Sources of Variation Involved in Rating Speeches." Unpublished Master's thesis, Michigan State College, East Lansing, 1945. 19 pp.

Andregg, Neal Berry. "A Critical Study of Graphic Rating Scales." Unpublished Doctor's dissertation, Michigan State College, East Lansing, 1951. 138 pp.

Brandenburg, Earnest. "Evaluation in Speech Training Programs of the Armed Forces." Seattle: Washington University, 1952. 7 pp. (Mimeographed.)

Case, Keith. "An Investigation into the Backgrounds for the Study and Measurement of Personality in Speech Communication." Unpublished Doctor's dissertation, Denver University, Denver, 1948. 361 pp.

Central Michigan College of Education. "Faculty Handbook; A Summary of the More Important Policies, Regulations, and Procedures." Mount Pleasant, Michigan, 1953. 98 pp. (Mimeographed.)

Gibbs, David Elmore. "A Study of Reliability and Variation of Critical Rating of Speech by Trained and Untrained Observers." Unpublished Master's thesis, University of Washington, Seattle, 1948. 87 pp.

Graunke, Dean F. "The Use of Student Opinion in the Improvement of Instruction in Speech Fundamentals." Unpublished Master's thesis, University of Nebraska, Lincoln, 1951. 187 pp.

Jones, Horace Redman. "The Development and Present Status of Beginning Speech Courses in the Colleges and Universities in the United States." Unpublished Doctor's dissertation, Northwestern University, Evanston, 1952. 216 pp.

Martin, W. L. "An Experimental Study of Seventeen Terms Used in Speech Evaluation." Spokane: Whitworth College, 1953. 16 pp. (Mimeographed.)

Moses, E. R., Jr. "A Survey of Speech Tests in Thirty American Universities and Colleges." Columbus: Ohio State University, 1942. 15 pp. (Mimeographed.)

Mueller, Henry L. "A Study of Speech Examiners' Ratings." Unpublished Doctor's dissertation, Teachers' College, Columbia University, New York City, 1952. 114 pp.

Penland, Virgil Darrell. "An Experimental Study to Measure Effectiveness in Oral Reading by Means of a Rating Scale Technique." Unpublished Doctor's dissertation, University of Southern California, Los Angeles, 1948. 177 pp.

Sundwall, Harry W. "Normal Curve Derived Score Probabilities." East Lansing: Michigan State College, 1950. (Mimeographed.)

Thompson, Wayne N. "An Experimental Study of the Accuracy of Typical Speech Rating Techniques." Unpublished Doctor's dissertation, Northwestern University, Evanston, 1943. 204 pp.

APPENDIX

A. EVALUATOR'S RATING SCALE

B. NAMES AND CODE NUMBERS OF EVALUATORS

C. LETTER TO SPEECH FACULTY

D. INVITATION TO STUDENT EVALUATORS

E. STUDENT EVALUATOR'S PREFERENCE REPORT

F. LETTER ASSIGNING STUDENT EVALUATORS

G. SPEAKER'S PREFERENCE SHEET

H. INTERPRETATION OF RATING SCALE CRITERIA

I. INSTRUCTIONS TO EVALUATORS

J. EVALUATOR'S ACQUAINTANCESHIP CHECK SHEET

K. SPEAKER'S ACQUAINTANCESHIP CHECK SHEET

L. PUNCH CARD USED FOR TABULATING THE DATA

M. CODE USED PREPARATORY TO IBM TABULATION

N. DATA SHEET USED TO COMPUTE CORRELATIONS

EVALUATOR'S RATING SCALE

Speaker_____

Regular class section_____
day hour room

This Speech: Date_____Time_____Room_____

Directions: Circle the number in the appropriate column and also write that number in the right hand margin. Then add the column. Please double check for accuracy.

		Poor				Fair		Average		Good	Excel-lent	SCORE
I.	THOUGHT											
	Content	1	2	3	4	5	6	7	8	9	10	_____
	Organization	1	2	3	4	5	6	7	8	9	10	_____
II.	LANGUAGE											
	Vocabulary	1	2	3	4	5	6	7	8	9	10	_____
	Sentence structure	1	2	3	4	5	6	7	8	9	10	_____
III.	VOICE											
	Enunciation	1	2	3	4	5	6	7	8	9	10	_____
	Adequacy	1	2	3	4	5	6	7	8	9	10	_____
IV.	ACTION											
	Posture	1	2	3	4	5	6	7	8	9	10	_____
	Gesture	1	2	3	4	5	6	7	8	9	10	_____
V.	GENERAL											
	Impression	1	2	3	4	5	6	7	8	9	10	_____
	Communica-tiveness	1	2	3	4	5	6	7	8	9	10	_____

TOTAL _____

_____
(Evaluator's Signature)

APPENDIX B

NAMES AND CODE NUMBERS OF EVALUATORS

01(1) Twyla Jo Newhouse
02(1) Jean Mayhew
03(2) Fred Bush
04(2) Emil Pfister
05(2) Elbert Bower
06(2) Herbert Curry
07(1) Ruth Fox
08(2) Don Kilbourn
09(2) Wilbur Moore
10(2) Neil Soumela
11(1) Shirley Clark
12(1) Doris Whitcomb
13(1) Phyllis Eichhorn
14(2) J. D. Shuttleworth
15(2) James Jaska
16(1) Vivienne Jack
17(2) Don Kemp
18(2) Raymond Page
19(2) Richard Balwinski
20(2) James McLennan
21(1) Sheila Maule
22(2) David West
23(2) James Prough
24(1) Barbara Moore
25(2) Clyde Hatter
26(1) Josephine Nickora
27(1) Alma Nevins
28(2) Arthur Rice
29(2) Richard Torongo
30(1) Petrine Churchill
31(2) Robert Beckley
32(2) John Kirn
33(1) Alice Wagner
34(1) Jacqueline Robinson
35(1) Marian Sanborn
36(2) Herbert Sanford
37(2) Kenneth Downing
38(2) L. D. Foster
39(1) Anita Hoag
40(1) Betty Borman
41(2) Robert Gravelle
42(2) Virgil Scott
43(2) Royal Riggs
44(2) Jack Clary
45(4) Bernard Randolph
46(3) Jean Conklin
47(4) Jack White
48(3) Jean Caldwell
49(3) Norma Levi
50(3) Martha Fuce
51(3) Patricia Thwaits
52(3) Jean Detzur
53(3) Carol Clark
54(3) Donna Clapp
55(4) Keith Birdsall

Key

First two numbers:
- 01 - 09 Faculty evaluators
- 10 - 55 Student evaluators

Last number:
- (1) Females who judged both semesters
- (2) Males who judged both semesters
- (3) Females who judged second semester only
- (4) Males who judged second semester only

APPENDIX C

LETTER TO SPEECH FACULTY

Dear Colleague:

Below is a list of the juniors and seniors who are speech majors and minors and who are not on academic probation. They are being considered as evaluators for the Intergroup Speech Projects. Will you (1) please put your initials after each student you have had in class; (2) add a question mark after anyone who you feel may be an incompetent judge.

Emil R. Pfister

MAJORS	MINORS	MINORS
Richard Balwinski	Brian Beckley	Roger Parrish
Betty Borman	John Bilsky	James Prough
Jean Caldwell	Keith Birdsall	Bernard Randolph
Shirley Clark	Dale Brown	Arthur Rice
Jack Clary	Petrine Churchill	Royal Riggs
Jean Conklin	Carol Clark	Herbert Sanford
Jean Detzur	Donna Clapp	Virgil Scott
Kenneth Downing	Dale Edgerle	J.D. Shuttleworth
Phyllis Eichhorn	Martha Fuce	Thomas Simpson
L. D. Foster	Robert Gravelle	Art Stinchcomb
Phyllis Gordon	Clyde W. Hatter	Patricia Thwaites
Lonna Rae Hall	Anita Hoag	Richard Torongo
James Jaska	Vivienne Jack	Paul Totzke
Don Kemp	Jean Klozik	Everett Vincent
John Kirn	Betty LaLone	Russell Ward
Norma Levi	James McLennan	Mary Weber
Joyce A. McNamara	Lorna Lesnick	Jack Weir
Josephine Nickora	Robert Lucas	Joyce Wells
Richard Powell	Sheila Maule	David West
Jacqueline Robinson	Barbara Moore	Doris Whitcomb
Marian M. Sanborn	John Murchie	
Carla Snow	Alma Nevins	
Neil Suomela	Raymond Page	
Loyal Thornton		
John Trask		
Alice Wagner		
Jack White		

APPENDIX D

INVITATION TO STUDENT EVALUATORS

Central Michigan College of Education
Mount Pleasant, Michigan
October 13, 1952

Dear _____,

You have been recommended by members of the Department of Speech and Drama to serve as one of the Evaluators in the Intergroup Speech Projects this year. In these projects Freshmen in Speech 101 give three minute speeches for speech majors and minors to evaluate on a rating blank.

We should appreciate it very much if you could serve at two or three of the possible thirteen times.

If you are willing to cooperate in conducting this speech project, please answer on the enclosed preference blank. Please return this to the Speech Office (W261) no later than 5:00 p.m. this Friday, October 17.

The speech majors and minors will hold two meetings next week, Thursday, October 23, in Keeler Dining Room C to discuss the rating blank to be used. One will be from 4:00 - 5:00 p.m.; the other from 6:30 - 7:30 p.m. Since the worth of the project depends so much upon the cooperation and mutual understanding among the evaluators, you are requested to attend both of these meetings.

Cordially yours,

Wilbur E. Moore, Head
Department of Speech and Drama

APPENDIX E

STUDENT EVALUATOR'S PREFERENCE REPORT

I am willing to serve as an evaluator at two of the Speech 101 Intergroup Projects. I will be available at the following times:

(Put "P" in two or three of the blanks below to indicate preference; "X" for those times that are impossible.)

	12:05-1:55	1:05-1:55	4:05-4:55
Monday, Oct. 27			
Tuesday, Oct. 28			
Wednesday, Oct. 29			
Thursday, Oct. 30			
Friday, Oct. 31			

_____
(signed)

Please return to Speech Office on or before Friday, October 17.

APPENDIX F

LETTER ASSIGNING STUDENT EVALUATORS

Department of Speech and Drama
Central Michigan College of Education
Mount Pleasant, Michigan

October 20, 1952

Dear _____:

Thank you for filling out and returning the Student Evaluator's Preference Report. Please report to the Speech Office, W261, at the following times to be available as an evaluator of Intergroup Speech Projects:

(1) Time:_____ Date:_____

(2) Time:_____ Date:_____

(3) Time:_____ Date:_____

You will probably be asked to evaluate only twice and be an alternate the third time.

Thank you for the cooperation you have given us.

Sincerely,

Emil R. Pfister, Director
Intergroup Speech Projects

APPENDIX G

SPEAKER'S PREFERENCE SHEET
FOR
INTERGROUP SPEECH PROJECT

_____ Name

_____ Instructor

_____ Class time and days

Please mark with "P" the two dates you prefer to give your speech for the Intergroup Speech Project. Mark with an "X" the times that you are unable to participate.

	12:05-1:55	1:05-1:55	4:05-4:55
Monday, Oct. 27			
Tuesday, Oct. 28			
Wednesday, Oct. 29			
Thursday, Oct. 30			

APPENDIX H

INTERPRETATION OF RATING SCALE CRITERIA

Below is a list of the specific questions agreed upon by the evaluators. These may make for a better interpretation of the criteria used in the "Evaluator's Rating Scale for the Intergroup Speech Project."

I. THOUGHT

(A) Content
1. Is there enough material to cover the subject well?
2. Is subject and the material interesting enough to hold the attention of the listener?
3. Is the information accurate?

(B) Organization
1. Is there an adequate introduction?
2. Is the body of the speech well planned?
3. Is there continuity?
4. Is there an adequate conclusion?

II. LANGUAGE

(A) Vocabulary
1. Is pronunciation correct?
2. Is there an adequate variety?
3. Are words used correctly?
4. Is there an overuse of slang?
5. Is vocabulary suitable to audience?

(B) Sentence Structure
1. Are sentences grammatically correct?
2. Is there an adequate variety of sentences?
3. Is there an overuse of the word "and"?

APPENDIX H
(CONCLUDED)

III. VOICE

(A) Enunciation
1. Can you understand the speaker?
2. Does the speaker slur his words?
3. Is the speaker over-articulate?

(B) Adequacy
1. Does the speaker speak too loudly or too softly?
2. Is the voice pleasant?
3. Is there enough variety?
4. Is there good timing and use of pauses?

IV. ACTION

(A) Posture
1. Does he lean against anything?
2. Are his hands and feet in a comfortable, natural position?
3. Does he hold his head properly?

(B) Gesture
1. Are the speaker's actions distracting?
2. Are the speaker's actions suitable and meaningful?
3. Is there either insufficient or too much action?

APPENDIX I

INSTRUCTIONS TO EVALUATORS

OF THE

INTERGROUP SPEECH PROJECT

1. Do all judging independently. Select seats near the center of the room.

2. Try to get the meeting started as soon as possible. There will be a few participants who are on leave from other classes so please permit them to give their speeches at the beginning of the session.

3. Please furnish all information called for at the top of the rating scale.

4. All speeches are to be limited to three minutes. Do not permit any contestant to exceed this time limit.

5. If you have failed to complete the rating on a speaker by the time he has concluded, complete the scale before calling another speaker.

6. Be impersonal in your judgment. Do not let personalities temper your evaluation.

7. Rate the speakers on all ten items on the rating scale, and total the ratings.

8. You can aid the speakers by being a considerate judge. Try to refrain from showing disagreement, disgust, or disinterest in the speeches. Be an attentive listener.

9. Above all, remember that these meetings have been arranged to furnish additional speaking experience for the Speech Fundamentals students. For that reason, it is imperative that you do the best job of evaluating that you are capable of rendering. Do your judging fairly and conscientiously.

10. At the end of the session return the rating scales to the Speech Office (W261).

APPENDIX J

EVALUATOR'S ACQUAINTANCESHIP CHECK SHEET

This speech: Date_____Time_____Room____

Name of Speaker_____

Please circle the number below which is the most nearly correct answer:

I. To what extent do you know the speaker?

1	2	3	4	5
Not at all	Casually	Moderately	Very well	Intimately

II. I converse with this person:

1	2	3	4	5
Never	About once a semester	Once a month	Once a week	Almost daily

_____
Evaluator's Signature

APPENDIX K

SPEAKER'S ACQUAINTANCESHIP CHECK SHEET

This speech: Date_____Time_____Room_____

Name of Evaluator_____

Please circle the number below which is the most nearly correct answer:

I. To what extent do you know the evaluator?

1	2	3	4	5
Not at all	Casually	Moderately	Very well	Intimately

II. I converse with this person:

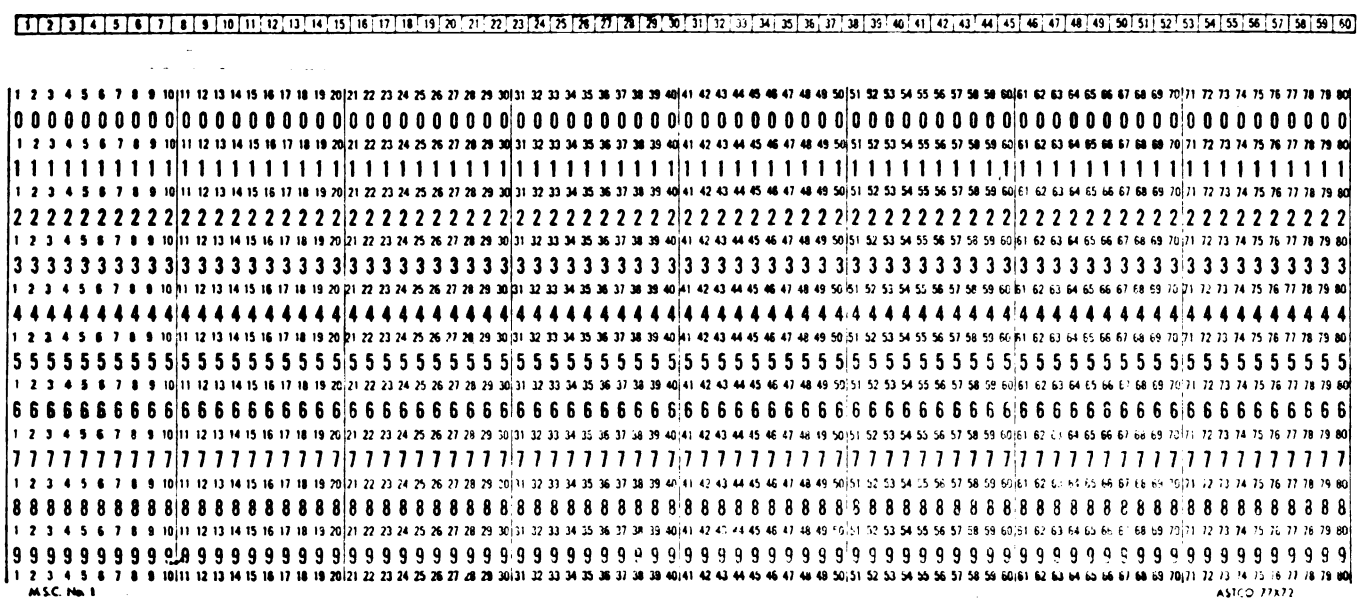
1	2	3	4	5
Never	About once a semester	Once a month	Once a week	Almost daily

_____
Speaker's Signature

APPENDIX L

PUNCH CARD USED FOR TABULATING THE DATA

BY MEANS OF IBM EQUIPMENT



APPENDIX M

CODE USED IN PREPARING THE RATING SCALES

FOR IBM MACHINE TABULATION

Column on IBM Card	Data From the Rating Scale Recorded on Card	Numbers Punched on Card	Explanation of the Number Code
1	Semester speech was given	1-2	1 for first; 2 for second
2	Sex of speaker	1-2	1 for female; 2 for male
3-4	Speaking order	01-20	01 for first; 02 for second
5	Day that speech was given	1-5	1 for Mon.; 2 for Tues., etc.
6	Time that speech was given	1-3	1 for 12 noon; 2 for 1:00, etc.
7	Room where speech was given	1-5	All in the same building
8-17	Scores on the ten traits	01-10	See Appendix A
18-27	Scores on the five criteria	10-20	See Appendix A
28-29	Total score	10-99	No 100's were given
30-32	Judge's code number	001-554	See Appendix B
33-42	Scores on the ten traits*	01-10	Comparable with columns 8-17
43-52	Scores on the five criteria*	10-20	Comparable with columns 18-27
53-54	Total score*	10-99	Comparable with columns 28-29
55-57	Judge's code number*	001-554	When not same as columns 30-31
58-59	Acquaintanceship factor*	1-5	See Appendixes J and K

*Data on the second speech given by the speaker.

DATA SHEET USED TO COMPUTE CORRELATION

Formula used:*

$$r = \frac{N\Sigma XY - (\Sigma X \times \Sigma Y)}{\sqrt{[N X^2 - (X)^2]\ [N Y^2 - (Y)^2]}}$$

$N \times \Sigma XY$ or _____ x _____ =

$\Sigma X \times \Sigma Y$ or _____ x _____ = ___

_____

Numerator =

$N \times \Sigma X^2$ or _____ x _____ =

$(\Sigma X)^2$ or _____ x _____ = ___

_____

Ans. I

$N \times \Sigma Y^2$ or _____ x _____ =

$(\Sigma Y)^2$ or _____ x _____ = ___

_____

Ans. II

Ans. I x Ans. II or _____ x _____ = _____

$\sqrt{}$ = _____

Demoninator

Numerator

_____ or _____ = _____r

Denominator

*Henry E. Garrett, Statistics in Psychology and Education. New York: Longmans, Green, and Company, 1947, p. 292.