


1994

## An Assessment of Panel vs. Individual Instructor Ratings of Student Speeches

David E. Williams  
*Texas Tech University*

Robert A. Stewart  
*Texas Tech University*

Follow this and additional works at: <http://ecommons.udayton.edu/bcca>

 Part of the [Higher Education Commons](#), [Interpersonal and Small Group Communication Commons](#), [Mass Communication Commons](#), [Other Communication Commons](#), and the [Speech and Rhetorical Studies Commons](#)

---

### Recommended Citation

Williams, David E. and Stewart, Robert A. (1994) "An Assessment of Panel vs. Individual Instructor Ratings of Student Speeches," *Basic Communication Course Annual*: Vol. 6, Article 10.  
Available at: <http://ecommons.udayton.edu/bcca/vol6/iss1/10>

This Article is brought to you for free and open access by the Department of Communication at eCommons. It has been accepted for inclusion in Basic Communication Course Annual by an authorized editor of eCommons. For more information, please contact [frice1@udayton.edu](mailto:frice1@udayton.edu), [mschlangen1@udayton.edu](mailto:mschlangen1@udayton.edu).

# **An Assessment of Panel vs. Individual Instructor Ratings of Student Speeches**

---

*David E. Williams*  
*Robert A. Stewart*

Each semester numerous students venture into our public speaking courses. Unlike most of the curriculum, these students enter a course in which their final grade will be based, partially, on a subjective evaluation of their performance ability. While instructor training and clearly defined speech presentation objectives are helpful, it is still impossible to eliminate the subjective nature of performance evaluation. Speech grading becomes even more critical when one tries to balance the expectations of several instructors teaching different sections of the same basic course.

This paper will suggest the use of a panel grading system to help combat the possibility of instructor bias and increase the amount of useful feedback provided for the student. Following a review of the most common forms of grading bias this essay will then identify precedents for the use of an instructor panel grading system. Finally, the results of an initial study will be offered along with relevant considerations for the implementation of the panel grading system.

## **PERFORMANCE EVALUATION CONCERNS**

Most public speaking instructors employ a criterion referenced measurement when assigning presentation grades.

*Assessment of Panel vs. Individual Instructor Ratings*

With criterion-referenced evaluation students will compete against their instructor's perception of what constitutes an A, B, C, D and F speech. Smythe, Kibler, and Hutchings (1973) revealed that criterion-referenced measurement is essential in communication performance courses. In a norm-referenced course, which would compare student performances against each other, a student could give a speech that would meet the criteria for a C speech, yet receive a lower grade because of being in a class of superior speakers. Frisbee (1989) noted that criterion-referenced grading allows the student to focus on course goals and possibly assist a peer without jeopardizing his/her own grade.

However, Rubin (1990) noted that instructors who use criterion-referenced grading must still be concerned with validity and reliability in performance evaluation. Rubin explained that validity refers to "how accurate and comprehensive an evaluation is" (p. 380). For example, validity may refer to whether or not a grading sheet used to evaluate speakers has all the elements on it which the instructor will be looking for. Reliability deals with consistency and dependability. The concern here is whether the instructor grades each speaker with equal rigor and according to the same criteria.

## **BIAS**

Various types of bias can reduce the validity and reliability of a performance assessment (Airasian, 1991; Rubin, 1990; Stiggins, Backlund & Bridgefords, 1985). Rubin identified several forms of bias which result from a lack of objectivity by the instructor including cultural biases, leniency, trait error, central tendency, and halo and horned effects. Leniency error refers to the tendency to be too easy or too hard (negative leniency error) in the evaluation of all performances in a class. Central tendency refers to an instructor's

*Assessment of Panel vs. Individual Instructor Ratings*

grouping of grades in a fairly tight cluster. This tendency will frequently bring down the grades of students who give superior performances while increasing the grades of inferior performances.

Halo effect and horned effect occur when an instructor is too easy or hard on a specific speaker, while trait error is the extremely harsh or lax grading on a specific component of the performance assignment (e.g., delivery, research). A study by Bohn and Bohn (1985) argued that leniency and halo effect should be of greater concern to instructors than trait errors and confirmed earlier findings (Bowers, 1964; Guilford, 1954; Gunderson, 1978) that rater training reduced overall and leniency error.

Finally, Rubin (1990) revealed that previous researchers (e.g. Miller, 1964) have warned that individual preferences and prejudices may influence an instructor's evaluation of a performance. Possibly the most likely areas of bias would be the instructor's attitude about the speaker's topic and mental disposition toward the speaker.

Another form of bias, not typically addressed in the literature, is the limited view a student receives from the feedback of only one evaluator. While the instructor may consistently apply his/her criteria for acceptable delivery to each student, how might that instructor's delivery criteria differ from those of another instructor? A student may be informed by one instructor that her delivery is acceptable while another instructor would see a need for improvement.

The limited view from a single instructor goes beyond ratings on a criteria sheet. Instructors typically provide written and/or oral feedback regarding what was done well and how to improve weaknesses in a performance. A variety of informed evaluators would discover more areas for potential improvement and provide more suggestions on how to make the necessary changes.

The use of graduate teaching assistants (GTAs) in public speaking classes adds another variable when attempting to

*Assessment of Panel vs. Individual Instructor Ratings*

improve evaluator reliability and validity. Graduate students teach a significant number of public speaking students each year. Gibson, Hanna, and Huddleston (1985) discovered that (GTAs) taught 18% of all basic communication courses.

Most teaching assistants receive some form of training but not solely on performance evaluation, although 97% of all GTAs, across disciplines, have grading responsibilities (Diamond & Gray, 1987; Parrett, 1987). Research on GTA grading practices suggests that these instructors tend to be more lenient than their faculty counterparts. Williamson and Pier (1985) found in a study of 43 basic communication course sections taught by faculty and GTAs (seven faculty members and 17 GTAs) that GTAs assigned more Bs and incompletes while instructors used more Cs and Ds.

## **PANEL GRADING**

Panel grading is suggested here as a means for further enhancing performance evaluation validity and reliability while also increasing the amount of feedback each student receives on his/her presentation. The prospect of panel evaluation is not without precedent. According to Thompson (1944) more accurate speaker ratings might be achieved with a panel of raters.

A stronger precedent is found in intercollegiate forensics competition. Forensics tournament directors and coaches recognize the importance of panel judging. During preliminary rounds of debate or individual events competition tournament directors are limited to providing only one or two judges per round. However, for elimination rounds, panels of three or five judges are assigned to evaluate the speakers.

Forensics coaches and tournament directors have recognized the importance of the decisions being rendered in elimination rounds. Panel judging is used to counter the possibility of one judge making a poor decision based on a particular bias

*Assessment of Panel vs. Individual Instructor Ratings*

or inaccurate evaluation of what is taking place in the round. Panel judging has an additional benefit of providing the student with a variety of feedback on his/her performance. The student also can compare judges' comments to determine which critiques are verified by similar statements and which critiques reflect isolated concerns or observations.

Peer evaluations provide another precedent for panel grading. Instructors frequently have students in the audience assign a grade and/or provide written or oral feedback to their peers. Book and Simmons (1980) found that student evaluators provided beneficial comments for their peers. They revealed that the feedback was perceived as helpful by the speakers, consistent with content and delivery criteria, and similar to instructor feedback.

Zeman (1986), however, noted that peer evaluators are particularly susceptible to leniency, halo, and trait errors. Barker (1969) likewise found the probable existence of a halo effect in students' evaluations of speeches. Rubin (1990) added that student ratings are higher than instructor scores, and students who are next to speak are more lenient in their scores and then become more negative after they have delivered their speech. Rubin summarizes the conflicting data regarding peer evaluation by stating, "it is NOT clear that peer evaluations are valid and reliable. The criticism given in class by peers is helpful, but their grades may not be accurate" (p. 382). Thus peer evaluation panels provide a precedent, but not a substitution, for panel grading with instructors.

## **A PRELIMINARY STUDY**

A study of 48 speeches given by students in public speaking classes was conducted to examine the effects of instructor panel grading in comparison with individual instructor grading. The researchers used students from three different public

*Assessment of Panel vs. Individual Instructor Ratings*

speaking classes and a total of five graduate teaching assistant instructors (GTAs). Each GTA had one year of teaching experience and had completed university-wide and departmental GTA training. Students from three of the five GTAs' classes were used in this study. The other two GTAs were used in grading panels, but their students were not involved in the study.

Each of the 48 student speeches was videotaped by the instructor. This was a common practice as it was required of all students in the various public speaking courses. Each student delivered an informative speech designed to provide new or useful information for the audience. The use of visual aids was optional. The student's instructor would evaluate the speech and assign a grade. This grade was recorded in the instructor's grade book and stood as the actual grade for the presentation. After grading speeches for one class, the instructor would turn the videotapes over to the designated panel of three other GTAs.

Panel raters and instructors used the same speech evaluation form for rating student speeches. The form consisted of 15 items rated on a 5-point scale, with 1 the lowest rating and 5 the highest. The items reflected criteria for the speech assignment concerning statement of purpose, organization of main points and use of support material, use of language and visual aids, and delivery. The form also included a debit item for exceeding or falling below the assigned time limit, but almost none of the speeches were affected this way; so, the item was excluded from analysis. Both panel raters and instructors used criterion-based evaluation. This was standard policy for all sections of the public speaking course.

*Assessment of Panel vs. Individual Instructor Ratings***RESULTS**

Because the items on the speech rating form were summed to derive student scores for grade determination, those scores were the unit of analysis in this study. The number of student speeches involved in this study ( $n=48$ ) was deemed too small to retain sufficient statistical power with so many possible comparisons. Means for each panelist and instructor for each class are displayed in Table 1. Scores could range from a minimum of 15 to a maximum of 75, with a theoretical midpoint of 45. Assuming the common grade scale of 90% for an A, 80% for a B, and so forth, the means generally indicate scores in the middle to high B range across raters and classes, with the exceptions being Raters B and C in Class 2 whose mean ratings represent grades of C. Inspection of item means for each Rater in each class showed consistent ratings of 4.00 or higher on the 5.00 scale. Thus, leniency may have affected ratings of these speeches across the board.

Table 1  
Means ( and Standard Deviations) on Rating Scores  
for Each Rater and Instructor within Class

Class	<i>n</i>	Rater A	Rater B	Rater C	Instructor
1	13	64.92 (4.89)	64.15 (4.36)	61.92 (6.16)	66.69 (5.51)
2	17	62.71 (4.38)	56.47 (8.23)	59.47 (7.11)	65.82 (5.63)
3	17	63.29 (7.11)	63.29 (6.79)	66.41 (5.12)	66.12 (5.29)

Note: Classes had different raters and instructors, hence, columnar means represent independent ratings.



*Assessment of Panel vs. Individual Instructor Ratings*

Table 2  
Alpha Coefficients of Reliability for Rating Scores  
For Each Rater and Instructor within Class

Class	Rater A	Rater B	Rater C	Instructor
1	.75	.61	.76	.84
2	.62	.86	.79	.82
3	.86	.82	.80	.76

*Was the rating scale reliable?* Table 2 presents alpha reliability coefficients computed for each rater within each class. Taken together the coefficients show the scale to have had moderate to moderately high reliability across multiple users and samples. Each coefficient also can be taken as an indication of intra-rater reliability within a class. The greatest similarity in reliabilities across raters was in Class 3 and the least in Class 2. In Class 1 the evaluation instrument achieved greater reliability for the instructor than for any of the panelists, while that of the instructor in Class 3 was somewhat lower than the panelists'. Since the alpha coefficient is a measure of internal consistency of items within a scale, the variation in coefficients suggests that different raters responded somewhat differently to the items. Perhaps raters differed as to the criteria they emphasized in completing the evaluations, suggesting some degree of trait error on the part of these raters.

*Was the average score across raters reliable?* One way a panel of raters could be used in evaluating student speeches would be to average their ratings with that of the instructor. The need would then arise to establish the reliability of the obtained average score. In the present case, scores assigned by the three panelists and instructor within each class were

*Assessment of Panel vs. Individual Instructor Ratings*

treated as a composite, and alpha reliability coefficients thereby computed. For Class 1 the reliability was .86. For Class 2 it was .93. And for Class 3 it was .91. Thus a form of inter-rater reliability was established for each class. In all three classes the resulting coefficients can be considered high.

*Were rating scores consistent among panel raters?* Analyses reported above revealed that the scale was reliable across users, and that combining panelist and instructor ratings would produce highly reliable average scores. Another issue concerned whether mean ratings on the same speeches by a panel of raters were statistically similar. Assuming each speech was evaluated similarly by the three panelists, it would follow that the raters' means on those evaluations would not differ significantly. Pairwise t-tests were computed to compare the means of panelists within each class. Results are reported in Table 3. In six out of nine comparisons,

Table 3  
Tests for Pair-wise Differences in Rating Scores Among  
Raters within Each Class

Class	Raters A-B	Raters A-C	Raters B-C
1	.77 (1.43)	3.00 (1.51)	2.23* (0.99)
2	6.23** (6.35)	3.24* (1.23)	-3.00** (0.71)
3	0.00 (1.58)	-3.12* (1.38)	-3.12** (0.82)

Note. Parenthetical values are standard error of the difference between the pair of means.

\*\*p<.01

\*p<.05

*Assessment of Panel vs. Individual Instructor Ratings*

pairs of panelists differed significantly in their mean ratings of the same students' speeches. Most striking is that in Class 2 all comparisons were significantly different. These findings indicate that even experienced panelists can be inconsistent in their evaluations of student speeches, and call into question the reliability results reported above. Still, it is interesting to note that Raters A and B in both Classes 1 and 3 were negligibly different in their respective average evaluations.

*Were individual panelists' mean ratings consistent with the instructor's ratings?* Results of this analysis are reported in Table 4. For this analysis, t-tests were computed to compare each panelist's mean ratings in each class with the mean ratings made by the instructor of that class. Out of nine comparisons, four were nonsignificant, showing consistency between those panelists and instructors. Two of these occurred

Table 4  
Tests for Pair-Wise Differences in Rating Scores between  
Each Rater and Respective Instructor within Class

Class	Instructor — Rater A	Instructor — Rater B	Instructor — Rater C
1	1.77 (1.16)	2.53 (1.51)	4.77** (0.99)
2	3.12** (0.78)	9.35** (1.23)	6.35** (1.29)
3	2.82 (1.20)	2.82* (1.01)	-.29 (0.88)

Note: Parenthetical values are standard error of the difference between the pair of means.

\*\*p < .01

\*p < .05

*Assessment of Panel vs. Individual Instructor Ratings*

in Class 1 and two in Class 3. For Class 2, none of the panelists was similar to the instructor in evaluating student speeches. In each of these cases, the instructor's mean rating was significantly higher than those of the panelists. The same is true for the other two significant differences. In fact, in only one comparison did the instructor have a mean rating lower than a panelist.

## DISCUSSION

While specific conclusions might be difficult to derive from this study, some tendencies were apparent. The rating form used in this study was found to have adequate reliability across classes and raters, but the panelists differed in their ratings of students in the same class. Panel members apparently varied in how they applied the criteria indicating that trait error was prevalent. Although there was a strong tendency to rate students at the top end of the rating scale, there was discrepancy among individual items. This would help explain the differences in overall mean ratings among panelists.

This study found that while some panelists were similar to instructors in evaluating the same speeches, others were significantly different. This finding could be interpreted in different ways. One interpretation suggests that the use of panel evaluators has promise and could be an effective grading practice. A second insight would hint that steps need to be taken to help insure the strongest validity and reliability possible with instructor and/or panel ratings. The third interpretation could offer that panel grading allows evaluators to make distinctions between superior and inferior performances which regular instructors do not make when assigning grades. While there seemed to be relative agreement in the performances which received the highest grades, much of the discrepancy between instructor and panel grades tended to occur

*Assessment of Panel vs. Individual Instructor Ratings*

with performances that received lower grades. In general, the panel would tend to grade weaker performances more harshly than the individual instructor. It could be possible that the panel graders are less susceptible to leniency error and therefore give more accurate grades to inferior performances.

Two other important needs seem to be emphasized by the results of this study. First, it is important to use systematic and thorough training of all raters. This will help to alleviate leniency and trait error. A second need falls into the decision making realm of the course director. While it appears that there may be some merit to the use of panel evaluators, the course director will need to determine how much emphasis to place on the instructors' grade and how much to place on the panel's evaluation.

*Suggestions for Implementing a Panel Grading System.* While evaluating the possible merits of panel grading, basic course directors also will need to determine whether such a system could be implemented in their department. Although circumstances and available resources will vary between institutions, we can offer a few frameworks which might be tailored according to specific needs.

The first means of implementing panel grading involves selecting four GTAs/instructors who would have only performance grading responsibilities, they would not teach sections of the basic course. This framework might be appropriate for departments which offer 15 or fewer sections per semester.

The selected instructors could be paired together with each duo assigned to assist in the grading of speeches from half of the sections. With this framework, each regular instructor would grade their students' performance and then the two elected instructors would also grade either the live or videotaped performance. All students would receive feedback from three evaluators and a panel grade could be determined.

Assuming that there were 15 sections of the basic course being taught, with an average of 25 students per section, one pair of selected instructors would evaluate 175 speeches

*Assessment of Panel vs. Individual Instructor Ratings*

(seven sections) and the other pair would evaluate 200 speeches (eight sections) per round of assigned speeches. With ten sections, each selected instructor would grade 125 speeches. While this is a heavy grading burden, it is balanced by the fact that the selected instructors would not have traditional instructional responsibilities and would have no duties when speeches were not being presented. The selection of panel instructors can be based on seniority or other qualities which would indicate that those individuals are among the most competent evaluators available.

While this is probably the easiest means for implementing panel grading, it has some limitations which might make it impractical for many basic course directors. Selecting four GTAs/instructors to have positions which do not involve covering classes will not be economically feasible for many departments. Arguments can be made for the improved evaluation and development of students which could result from panel grading, but these claims will probably not be enough to persuade most administrators who have budget constraints.

The perceived value of the panel instructors might also emerge as a problem. Ideally, these positions would carry a degree of esteem and be sought after by instructors or GTAs. However, if the grading is perceived as being too burdensome, these positions may not be wanted by the most qualified individuals. Furthermore, GTAs may prefer the experience of classroom instruction as opposed to only evaluating speeches.

Finally, this format could probably not be used by course directors who have more than 15 sections per semester. With additional sections the panel evaluators would become overburdened with the number of speeches to evaluate and the quality of those evaluations would likely falter. Course directors would probably not be able to assign additional instructors to panel positions. These limitations will likely prevent many course directors from being able to use this panel grading format. However, if these limitations can be avoided, this

*Assessment of Panel vs. Individual Instructor Ratings*

format would be the easiest means for implementing panel grading.

A second way basic course directors can implement a panel grading format is by assigning groups of three instructors to work together. With this format, instructors would grade their own students' speeches and the other two instructors in the trio would also be responsible for evaluating those performances. Therefore, each instructor would evaluate their own 25 students and 50 additional students.

By assigning instructors to groups of three, the process of getting all speeches graded would be easier because each instructor would know which classes they are responsible for. The trio can also coordinate schedules to make the process more efficient. Along those lines, course directors could assign different class meeting times to each of the members of the trio. For example, a trio of classes could be scheduled for Monday, Wednesday, and Friday from 8:00 to 8:50, 9:00 to 9:50 and 10:00 to 10:50. This would allow for the possibility of panel members sitting in on the other classes to which they are responsible.

This format would allow the possibility of implementing panel grading without employing instructors/GTAs who do not cover the regular instructional responsibilities of the basic course. It also allows for the possibility of panel instructors either sitting in on the classes they are responsible for or grading the speeches from videotape at their leisure. Furthermore, this format is not limited by the number of sections available. It could work equally well with 15 or 50 sections of the basic course.

The limitation to this format is that the number of speeches instructors/GTAs are required to grade is tripled. Some consideration might need to be made for the extra time required to fulfill their grading responsibilities. For GTAs, it might be possible that their service responsibilities could be reduced to compensate for their grading responsibilities. Departments which require a larger number of speech

### *Assessment of Panel vs. Individual Instructor Ratings*

assignments (four or more) may choose to reduce the number of performance assignments in favor of the greater feedback per speech.

The preceding formats offer two quite different means for implementing panel grading into the basic course curriculum. Hopefully, interested basic course directors could implement one of these or a variation of either format. However, if full implementation of a panel grading system is not feasible, course directors could consider using the second format for only one or two of the assigned speeches. This would limit the grading burden on instructors yet provide some of the benefits of panel grading.

A final alternative would limit the use of panel grading to honors sections of the public speaking course. Honors students typically seek stronger academic challenges and more thorough feedback on their work. Panel grading would provide these students with the critique and feedback they desire. If full implementation of panel grading is not feasible in all sections of the public speaking course, this might be a viable alternative as the logistical concern of developing GTA/instructor grading panels for one (or two) honors sections would be minimal.

## CONCLUSION

This study suggests at least a few issues which must be taken into consideration before implementing a panel grading system. First, leniency error presents a problem for GTAs. This is consistent with the findings of Williamson and Pier (1985). However, panel members were less susceptible to leniency error than the real instructor of students who delivered inferior speeches. Second, trait errors are a common problem in performance evaluation and they are not necessarily eliminated by the use of panel evaluations. Third, there is a dichotomy between the use of instructor and panel eval-



*Assessment of Panel vs. Individual Instructor Ratings*

uations. One can assume that the instructor should be a more reliable evaluator because he/she knows the student better. However, this relationship may cause the prevalence of leniency error. Fourth, the availability of multiple written feedback (from panelists) gives the student more information on how to improve weaknesses, but there is the possibility that this information could become contradictory. Finally, the course director would need to consider the logistical complications of developing panels of evaluators. Future studies might benefit by overcoming two limitations of the present study. First, a larger sample size would allow for more detailed analysis. Finally, future studies might attempt to have panelists evaluate live performances instead of videotaped speeches.

## REFERENCES

- Airasian, P.W. (1991). *Classroom assessment*. New York: McGraw Hill .
- Barker, L.L. (1969). The relationship between sociometric choice and speech evaluation. *The Speech Teacher*, 18, 204-207.
- Bohn, C.A. & Bohn, E. (1985). Reliability of raters: The effects of rating errors on the speech rating process. *Communication Education*, 34, 228-231.
- Book, C. & Simmons, K.W. (1980). Dimensions and perceived helpfulness of student speech criticism. *Communication Education*, 29, 135-145.
- Bowers, J.W. (1964). Training speech raters with films. *Speech Teacher*, 113, 228-231.
- Diamond, R.M. & Gray, P.J. (1987). A national study of teaching assistants. In N. Van Note Chism and S.B. Warner (Eds.), *Institutional responsibilities and responses in the*

*Assessment of Panel vs. Individual Instructor Ratings*

- employment and education of teaching assistants.* (pp. 80-82). Columbus: The Ohio State University.
- Frisbee, D. (1989). Assigning course grades. Paper identified by the Task Force on Establishing a National Clearinghouse of Materials Developed for Teaching Assistant (TA) Training. ED 285 496.
- Gibson, J.W., Hanna, M.S., & Huddleston, B.M. (1985). The basic speech course at U.S. colleges and universities: IV. *Communication Education*, 34, 281-291.
- Guilford, L.P. (1954). *Psychometric methods*. New York: McGraw Hill .
- Gunderson, D.F. (1978). Video tape modules as a device for training speech raters. *The Southern Speech Communication Journal*, 43, 395-406.
- Miller, G.R. (1964). Agreement and the grounds for it: Persistent problems in speech rating, *Speech Teacher*, 13, 257-261.
- Nyquist, J.D., Abbott, R. D., & Wulff, D.H. (1989). The challenge of TA training in the 1990's. *New Directions for Teaching and Learning*, 39, 7-14.
- Parrett, J. L. (1987) A ten year review of TA training programs: Trends, patterns, and common practices. In N. Van Note & S. B. Warner (Eds.), *Institutional responsibilities and responses in the employment and education for teaching assistants: Readings from a national conference.* (pp. 67-79). Columbus: The Ohio State University.
- Rubin, R. (1990). Evaluating the product. In J.A. Daly, G.W. Friedrich, & A. L. Vangelisti (Eds.), *Teaching Communication: Theory, research and methods*, (pp. 379-401). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Smythe, M. J., Kibler, R. J., & Hutchings, P.W. (1973). A comparison of norm-referenced and criterion referenced

*Assessment of Panel vs. Individual Instructor Ratings*

measurement with implications for communication instruction. *Speech Teacher*, 22, 1-17.

Stiggins, R. J., Backlund, P.M. & Bridgeford, N. J. (1985). Avoiding bias in the assessment of communication skills. *Communication Education*, 34, 135-141.

Thompson, W.N. (1944). An experimental study of the accuracy of typical speech rating techniques. *Speech Monographs*, 11, 65-79.

Williamson, K. & Pier, J. (1985). Student grades in basic speech: A research report. A paper presented at the annual meeting of the Speech Communication Association. ED 264 632.

Wilson, R. F. & Griswold, M.L. (1985). The effects of method and comprehensiveness of training of the reliability and validity of ratings of counselor empathy. *Measurement and Evaluation in Counseling and Development*, 18, 3-11.

Zeman, J.V. (1986). *A method of using student evaluations in the basic course*. A paper presented at the annual meeting of the Speech Communication Association. ED 279 035.