


2003

## Stretching the Academic Dollar: The Appropriateness of Utilizing Instructor Assistants in the Basic Course

Paul D. Turman  
*University of Northern Iowa*

Matthew H. Barton  
*Southern Utah 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

Turman, Paul D. and Barton, Matthew H. (2003) "Stretching the Academic Dollar: The Appropriateness of Utilizing Instructor Assistants in the Basic Course," *Basic Communication Course Annual*: Vol. 15 , Article 10.  
Available at: <http://ecommons.udayton.edu/bcca/vol15/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).

## **Stretching the Academic Dollar: The Appropriateness of Instructor Assistants in the Basic Course**

---

*Paul D. Turman  
Matthew H. Barton*

Teaching the basic course has become a consistent and integral role for communication faculty across the nation. This role has become increasingly important because the ability to speak confidently in a public or small group setting has been consistently identified as one of the most important skills that college graduates need (Adler & Elmhorst, 2001). As the basic course has developed over time, a greater need to satisfy the private sector's demands has become more and more of an issue. Universities have responded in kind by increasing the enrollments in the basic communication course in order to accommodate some of these needs (Gray, 1989). Gray (1989) argues that this increased economic pressure has had a significant impact on the instructional format utilized to teach the basic course. Often an increase in class size has been a traditional solution to this problem, (Gibson, et al., 1980; Gibson, Hanna, & Huddleston, 1985) however, increasing classroom size brings with it a number of pitfalls. First, public and legislative bodies are calling for greater accountability for money spent to fund universities resulting in smaller budgets for some academic departments and continued pressure on faculty to make every student an "excellent"

speaker. Second, and more importantly for this study, because of this increase in external accountability universities are feeling the need to service more students in a single course with fewer dollars. Thus, administrators are caught between the need to teach a greater number of students with little increase in budget, while continuing to produce effective speakers.

There are no easy answers to these problems, but one common approach that institutions are using involves an increased use of graduate students (Buerkel-Rothfuss & Gray, 1990; Golish, 1999; Gray, 1989; Larenz, et al., 1992; Myers, 1998; Roach, 1997; Rushin, et al., 1997) and in some cases undergraduate students (e.g. Humbolt State University, University College of Cape Breton, University of Denver, University of Nebraska, Lincoln, Portland State University, Miami University, Hope College) to assist with instruction (e.g. grading student speeches, assisting with large lecture sections, providing feedback to students concerning speech topics, etc) in order to accommodate larger class sizes. This practice naturally begs the question; can undergraduates be effectively trained to evaluate student presentations in the basic course? While this idea has interesting promise, it is also fraught with potential peril. Perhaps two of the greatest concerns about this practice are the potential problems of rater error and speaker order effects. Thus, this investigation is designed to explore the effectiveness of utilizing undergraduate instructor assistants as speech evaluators in the basic course. In particular, this study attempts to determine whether instructor assistant (IA) grading is affected by rater error and recency and primacy effects based on the order in which students present. In addi-

tion, this study attempts to determine whether the quality of evaluative comments decreases between the first and last speakers.

## **GTA TRAINING**

Buerkel-Rothfuss & Gray (1990) argue that across all disciplines numerous institutions utilize graduate and undergraduate students to fulfill the duties of evaluating and critiquing student work at the undergraduate level. During an investigation of eight institutions, these researchers found that 53.5% of introductory courses were taught by Graduate Teaching Assistants (GTAs). Rushin, et al., (1997) indicated that for most institutions, GTAs have more one-on-one contact with undergraduates than professors and as Roach (1997) has argued the title of teaching "assistant" is deceiving, because most GTAs maintain complete control over their own courses with little or no training. Kaufman-Everett & Backlund (1980) found that 86% of the speech communication departments in their studies utilized GTAs for teaching autonomous sections of the basic course. Buerkel-Rothfuss & Gray's (1990) examination supported these conclusions indicating that most courses in speech communication were taught by GTAs with their own autonomous sections and that many were working on Masters rather than Doctoral degrees. As the use of graduate and undergraduate teaching assistants in a variety of undergraduate courses has increased, many researchers have begun to examine the impact teacher assistant training has on their effectiveness in the classroom. Rushin, et al., (1997) argued that

even though there appears to be a strong formal structure in place for GTA training which includes workshops, seminars, and courses, the experience is often brief and takes place at a superficial level. Buerkel-Rothfuss & Gray (1990) stated that "we should applaud our efforts and then redouble them. Much of our undergraduate education foundation rests on the ability of people who have had no prior teaching experience and who have only recently left the undergraduate classrooms themselves" (p. 305).

Many Basic Course Directors working with GTAs stress the importance and value of a rigorous training program for preparing them for the classroom. Of those programs measured in their study, Buerkel-Rothfuss & Gray (1990) found that the duration for training sessions ranged from one-hour to an entire semester with the average program utilizing a weeklong session prior to the start of the semester. They, however argued that it is still unclear what is appropriate to cover while training GTAs. Many programs simply address course content, grading procedures, and classroom management, while a limited number address instructional strategies for enhancing student learning (Buerkel-Rothfuss & Gray, 1990). Prieto and Altmaier (1994) suggested that most research on GTA training focuses exclusively on effects of training programs rather than more fundamental elements such as ensuring effective teaching and learning for undergraduates.

A significant concern for the training of GTAs is the development of grading practices. Allen (1998) reported that assessment decisions are extremely important in academic life. "If academics cannot grade work well, they will be viewed with sympathy or derision by their

colleagues, and in either case may have their professional competence and status called into question” (p. 241). With this in mind, there appears to be a number of variables that have been determined to impact the nature of grading including: student ethnicity (Agee & Smith, 1974; Rubin & Yoder, 1985; Young, 1998), gender (Bock, 1970; Ford, Puckett, & Tucker, 1987; King, 1998), positive leniency (Bock & Bock, 1977), halo effects (Lance, LaPointe, & Fiscaro, 1994; McKeachie, 1994; Murphy & Anhalt, 1992), and feedback strategies (Book, 1985; Clauser, Clyman, & Swanson, 1999; Loudon & Shellen, 1976). Another significant problem associated with rater error is the overall planning of the course. Foster, et al., (1990) discovered that student perceptions about the grading practices and grading scales used in assessment are notably different than the instructor intended them to be. For example, Quigley (1998) observed that because written and oral communication skills are so critical in the workforce, educators can and should take specific steps to incorporate these needs into the curriculum. Quigley explained that grading criteria needed to be “consistent with cultural expectations for public speaking” (p. 43). Additionally, when students are given oral assignments, they “benefit from clear grading criteria, structured practice, and specific feedback” (p. 48). Thus, failure to meet these steps in the planning process leads to poor instruction and little improvement in speaking skills. Other research has demonstrated that selecting a meaningful evaluation instrument (Carlson & Smith-Howell, 1995) can increase equity and accuracy of overall grading, but rater error remains a serious issue. Also, evaluator

training can help control for some grader errors (Goulden, 1990).

Finally, when training GTAs to grade effectively in the classroom, Basic Course Directors should be concerned about primacy and recency effects. For example, in 1925, Lund explored a theory that he called primacy, which referred to the notion that an idea presented first in a discussion would have a greater impact than the opposing side presented second (in Mason, 1976). Other research has since followed Lund's lead exploring the viability of his theory (Anderson & Barrios, 1961; Barnette, 1999; Bishop, 1987; Ehrensberger, 1945; Freebody & Anderson, 1986; Jersild, 1929; Krosnick & Alwin, 1987; Sato, 1990). Specifically relating to public speaking, Knower (1936) found that competitive speakers in first and last positions are more commonly ranked in intermediate positions as opposed to either high or low extremes and second to last speakers often score highest on final averages. Benson and Maitlen (1975) disputed some of Knower's findings as their research concluded that there was no significant relationship between rank and speaking position.

When training GTAs to utilize a standardized grading system for the basic course it is vital that basic course directors ensure various forms of rater error are not occurring. It is apparent that rater errors do exist for a number of reasons, and that further, there appears to be enough research supporting both primacy and recency effects. Because rater errors exist and most of the research suggests that training can help eliminate these problems, further research should be done in this area. One could reasonably argue that if graduate students are susceptible to the various forms of rater error, then

undergraduates are likely prone to make these same mistakes. Thus, if speaker order affects student evaluation, it is valuable to empirically test the effects of rater error on instructor assistant grading. Based on the above rationale the following research question was set forth:

**RQ 1: Are instructor assistants affected by the primacy and recency effects during the grading of student speeches?**

An additional challenge is ensuring that students receive the appropriate valid and reliable feedback from those that rate them during their presentations. Prieto and Altmaier (1994) suggested that most research on GTA training focuses exclusively on effects of training programs rather than on more fundamental elements such as ensuring effective assessment and development for undergraduates. One of the primary implications concerning the use of undergraduates (particularly undergraduates from majors outside the communication discipline) as raters in the basic course is whether they have the acquired skills to provide students with appropriate feedback to assist in the development of their speaking skills. Additionally, one could argue that as class size, and the number of speakers in a given class period increases; additional constraints are placed on undergraduate instructor assistants to provide effective feedback. Thus, to determine whether speaker order affects the quality of comments provided by instructor assistants the following research question was set forth.



**RQ 2: Does the order in which students speak affect the quality and reliability of speech evaluation comments from instructor assistants?**

## **METHOD**

### ***Participants***

The participants in this study consisted of 38 undergraduate instructor assistants (IAs) currently working with the basic course at a large Midwestern University. To become an IA in this university's basic course students must successfully complete the course, fill out an application and receive a strong endorsement from their previous instructor(s). Applicants are then competitively selected for the program based on their grade point average and reported performance in the classroom. Prior to the evaluation of student speeches in the classroom, instructor assistants take part in a rigorous eight week training program which focuses on evaluation of recorded presentations and speaker outlines, discussion on the value of presentation grades, and instruction on how to provide effective feedback. Overall, instructor assistants are composed of a mix of students including communication studies majors, business majors, communication studies minors and students majoring in the hard sciences (e. g., engineering, veterinary medicine, et. al.). Although instructor assistants have many important responsibilities in the course, their main role is grading student speeches. The basic course has an enrollment of approximately 550 students per semester, divided into 12 sections directed by a graduate teaching assistant (evaluation criteria, assignments and exams

are standardized across each section). In an average class, instructor assistants are responsible for 15 students and serve as graders and facilitators for these individuals based on the cooperative learning component of this standardized course.

### ***Procedures***

For this study, instructor assistants were asked to grade four ten-minute persuasive speeches selected from student speakers on the university's forensics team. All speeches were used competitively on the AFA (American Forensics Association) circuit during the 1999-2000 school year. These speeches were chosen for this study in order to ensure a consistency of high quality speeches and to ensure that the quality of the speech did not account for rater error in the event that it did occur. Three of these speeches were considered to be highly persuasive speeches (Persuasive Speaking Category) and one was considered moderately persuasive (after dinner speaking) based on the use of humor to discuss the problem. Also, to ensure the elimination of gender as a confounding variable, all speakers used in this study were female.

### ***Scales of Measurement***

Because speeches are an integral part of the pragmatic element of instruction in the basic course, it is critically important that instructor assistants receive appropriate instruction relevant to assessment. Consequently, before grading any of the speeches, trainers familiarized the instructor assistants with the criterion referenced evaluation instrument and other grading

techniques (e.g., taking copious notes, grading speeches on the same day they are given in class, etc). Instructor assistants utilized an evaluation instrument which utilizes an analytic method by which content and delivery elements are rated and then summed to generate the final score for the presentation, rather than a holistic approach (using personal judgment when determining the importance of specific traits toward the overall product). In an attempt to determine the effectiveness of each approach, Goulden (1994) found that neither the analytic nor holistic method was more effective at producing a reliable assessment of student presentations.

In addition to testing for any differences in the overall mean scores of student speeches related to speaker order, this study also measured the quality of student comments on a seven point semantic differential scale. This scale was created to analyze the quality of student comments based on a combination of the introduction/conclusion, the body and delivery. Three student coders were selected and asked to rate IA comments for each of the speakers based on a semantic differential type scale adapted from an instrument developed by Osgood, Suci, & Tannenbaum (1957). Using the stimulus statement of "What is the quality of the written feedback provided by the evaluator for this presentation" and used a 7-point scale to capture coders perceptions to the degree that each section (e. g., introduction, conclusion, body, delivery) was: good-bad, valuable-worthless, qualified-unqualified and reliable-unreliable. A semantic differential type scale was used because of its ability to accurately measure the way different individuals view the same concepts (Keyton, 2001; Neuman,

2000). To examine the validity of the scale, inter-coder reliability was computed at  $r = .76$ .

### ***Experimental Design***

Speakers were selected and taped in the regular training classroom to help simulate a typical speech day in the basic course. Speeches were then re-taped in a different order with 30 seconds between speakers. This was designed to make sure that each speaker appeared in the first, second, third and fourth position. To help maximize external validity and eliminate the potential for confounding variables, the research was conducted in four classrooms used during the training session. Each of the four groups was given the same environment, visual equipment and tape quality to help ensure a similar experience across all four groups.

To increase internal validity the independent variable (speaker order) was manipulated and the IAs were randomly assigned to one of four treatment groups. Three assistants were used to help administer the study. They were each provided with a detailed list of instructions in order to make sure that each group followed the same procedures and had the same experience. Participants were asked to watch all four speeches, evaluate them, make comments, assign final grades for each speech and return them to the primary investigator within 24 hours.

Three IAs not participating in the previous portion of the study were selected and trained as coders. These coders were then asked to use the presentation comment quality evaluation instrument to assess the quality of comments provided for each speaker.

## ***Design and Analysis***

Research question one used a 4 x 4 factorial design to measure the potential change in student speech grades. The order of the speech (either going 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, or 4<sup>th</sup>) was a between subjects design, while IA group assignment (group 1, 2, 3, or 4) is within subjects design. An analysis of co-variance (ANCOVA) was used to analyze data from the four groups based on the grade that was assigned. Research question two used a one-way analysis of variance (ANOVA) to analyze the data among the four groups on the dependent measure and the difference on scores assigned based on the quality of comments provided by the instructor assistants.

## **RESULTS**

The first research questions asked whether instructor assistants would be affected by primacy and recency effects when grading student speeches based on the order in which they gave their presentations. The findings indicated no significant difference on grades assigned to speakers based on their designated groups (Group & Speaker,  $F = 2.775$ ,  $p > .05$ ). There was a significant interaction between group and speaker, however an examination of mean scores reveals that the speaker position had no effect on the persuasive level of the other speeches. This suggests that the speech identified as moderately persuasive did not impact the grading of other speeches (1<sup>st</sup>,  $m = 89.83$ ,  $SD = 4.30$ ; 2<sup>nd</sup>,  $m = 92.87$ ,  $SD = 3.60$ ; 3<sup>rd</sup>,  $m = 89.25$ ,  $SD = 4.55$ ; 4<sup>th</sup>,  $m = 89.88$ ,  $SD =$

**Table 1**  
**Descriptive Statistics for Speaker by Group**

Speaker	Group	Mean	Std. Deviation
1.00	1.00	89.8333	4.3089
	2.00	93.6250	2.3261
	3.00	94.8750	1.9594
	4.00	93.3333	2.3979
2.00	1.00	93.4286	2.5071
	2.00	92.8750	3.6031
	3.00	93.2500	4.1662
	4.00	93.3333	1.8708
3.00	1.00	95.0000	.8944
	2.00	94.2000	2.7512
	3.00	89.2500	4.5591
	4.00	92.2222	2.3333
4.00	1.00	90.8333	4.6224
	2.00	91.7778	4.9441
	3.00	93.1250	4.0861
	4.00	89.8889	3.5158
	Total	91.4062	4.2719

3.51), because the other speaker scores did not vary more than two points from one group to the next.

The second research questions asked whether the quality of IA feedback would decrease from the first speaker to the last based on the order of student presentations (e.g. 3<sup>rd</sup> or 4<sup>th</sup>). Results indicate that no significant differences existed ( $F = .492$ ,  $p > .05$ ), suggesting

that students were likely to receive the same quality of comments from instructor assistants regardless of their position in the speaker order: 1st ( $m = 26.93$ ,  $SD = 8.87$ ), 2<sup>nd</sup> ( $m = 28.62$ ,  $SD = 9.53$ ), 3<sup>rd</sup> ( $m = 29.63$ ,  $SD = 9.96$ ), 4<sup>th</sup> ( $m = 27.84$ ,  $SD = 8.60$ ).

Table 2  
ANOVA Table

	Sum of Squares	Df	Mean Square	F	Sig.
Between groups	126.649	3	42.216	.492	.689
Within groups	10649.226	124	84.881		
Total	10775.875	127			

## DISCUSSION

This study focused on determining whether speaker order has a statistically significant effect on student speech grades and on the quality of written feedback. Two hypotheses were used to test for the presence of these relationships. Research Question one attempted to test for “speaker order effects” in the grading process. Findings show no evidence of primacy or recency effects, thus speaker order has no impact on the final grades students received during this study. These findings dispute Anderson & Barrios’ (1976) conclusions that primacy effects exist, as well as Miller & Campbell’s (1959) conclusions that recency effects exist to the extent that speaker order had no impact on final grade assignment.

However, this study is consistent with Benson & Maitlen's (1975) research, which found no significant relationship between rank and speaker position. Although their study is slightly dissimilar in that it looked specifically for primacy and recency effects in a competitive speech performance, the current findings show that students are equally evaluated regardless of the speaking order.

In addition, there are three other reasons that may help explain these findings. First, because these speeches were of such similar quality, perhaps they were not entirely representative of typical classroom speeches given in the basic course. Second, only four speeches were used in this study, which represents half the normal number of speeches delivered during a typical speech day at this university, which may not account for grader fatigue. Finally, there may be some support for the value of the criterion-referenced approach used during the IA training program (Behnke & Sawyer, 1998), resulting in higher levels of rater confidence in using the evaluation instrument.

The second research question focused more explicitly on the quality of evaluative feedback students received. This study found no evidence of differences between speaker position and the quality of comments students received from undergraduate instructor assistants. These findings suggest that students would receive the same type of feedback in terms of quality whether they were speaking in the first, last or intermediate position. These results are supportive of Louden & Shellen's (1976) findings in two ways. First, they found that judges assigned the same overall grade regardless of assessment experience, which is consistent to some extent



with this study because of the high degree of grader agreement. Second, and more importantly, because instructor assistants received the same type and amount of training, the idea that differences in feedback do not exist across similar groups is supported. There also appears to be some evidence to support other notable conclusions from this data.

First, inter-coder reliability was relatively low in this study ( $r = .76$ ). This may have been a result of a 7-point semantic differential scale, which allowed for more variability across the raters. Because such a low correlation exists, the quality of student feedback may be less uniform than these findings suggest. Inter-coder reliability at this level would indicate that it is difficult to determine whether the quality of feedback increased or declined across each of the speakers based on their placement in the speech rotation. Additionally, it is yet unclear as to whether undergraduates, especially undergraduates from disciplines outside communication, are capable of providing students with appropriate feedback. This finding suggests a greater need for more specific coder training in order to increase the strength and reliability of the coders and coding. Based on the above limitation, further research needs to be done to determine whether ranking of rater feedback would remain the same across speaker order if stronger inter-coder reliability was obtained.

Second, because instructor assistants did not have to interact with these speakers in the classroom, there may be some logic to suggest that they felt less inhibited in providing feedback and assigning overall scores. Instructor assistants were not faced with the pressures often associated with the grading process including stu-

dent reactions to presentation scores. This is one of the aspects of the grading process that might ultimately affect undergraduate raters the most. Additionally, watching speeches on videotape is not the same as a live experience in terms of the overall critical distance the mediated version provides.

Finally, because of the concern over grade inflation, the instructor assistant training program focuses on fundamental speech issues of organization and supporting materials, with a large focus on some delivery elements (like eye contact, movement and vocal disfluencies). Because instructor assistants are trained on such a straightforward criterion based level, these particular speakers were much more polished than many speakers evaluated during training and more capable than many speakers that instructor assistants might evaluate in the classroom which may have caused them to award higher scores in the classroom. Additionally, a larger number of speakers ranging from "A" to "F" performances would change the nature of these findings and better reflect the typical speaking day. Also, having more speeches would better test for instructor fatigue that is more likely to happen when more speeches are viewed at a given time. Since the literature suggests that rater errors still occur even after training, the implication is that "halo effects" and "personal relationships" (Bock & Bock, 1977) might exist which can impact student grades both positively and negatively. A further implication from this study supports Goulden's (1990) findings that training for classroom evaluators decreases rater error, and in this case, some of the consistency can be linked to adequate instruction in light of course objectives for instructor training.

A number of interesting implications emerge from this study in regards to the appropriate use of undergraduate raters and the pedagogical and institutional implications that result. Morreale et al, (1999) state that the biggest problem or frustration basic course administrators face is "maintaining consistency " across courses with multiple sections (p. 29). This study has demonstrated that an instructor assistant training program has the potential of reducing the variability that often occurs in grading across groups. More definitively, one potential implication for this finding is the utilization of the criterion-based rating scale for ensuring standardization across rater groups. By providing instructor assistants with a clearly established standardized set of criteria and then training them to utilize that criteria has a significant chance of reducing the variability that often occurs across multiple section courses.

While more research needs to be done, this study does show some promise in terms of increasing the reach and scope concerning the facilitation of the basic course. Additionally, Morreale et al, (1999) identified the maintenance of existing class size as an additional concern administrators of the basic course face. In this regard, these findings should be valuable for administrators or basic course directors who are considering the option of utilizing undergraduate graders in the basic course to alleviate some of the constraints associated with increased class loads and reduced budgets. However, as you examine the findings obtained from each of these research questions, it is important to discuss a number of implications that emerge on both a practical and pedagogical level. Although these findings suggest that undergraduates can be trained to consistently

grade across groups, they do not answer whether this practice is then appropriate for the college classroom or the basic course. A number of student, parent, and institutional issues begin to emerge as a result. Should undergraduates be placed in the position to evaluate their fellow students? Should parents feel their children are obtaining the best education available when undergraduates with limited knowledge of the field are involved in providing guidance for student presentations? Is the quality of the institution ultimately impacted by using undergraduates in multi-section courses? At this point, each of these broader questions is at stake and further research is needed to provide answers to these questions.

## REFERENCES

- Adler, R.B., & Elmhurst, J.M. (1999). *Communicating at work* (6<sup>th</sup> ed.). Boston: McGraw-Hill.
- Agee, W.H., & Smith, W.L. (1974). Modifying teachers' attitudes towards speakers of divergent dialects through inservice training. *Journal of Negro Education*, 43, 82-90.
- Allen, G. (1998). Risk and uncertainty in assessment: Exploring the contribution of economics to identifying and analyzing the social dynamic in grading. *Assessment and Evaluation in Higher Education*, 23, 241-258.
- Anderson, N.H., & Barrios, A.A. (1961). Primacy effects in personality impression formation. *Journal of Abnormal and Social Psychology*, 63, 346-350.

- Behnke, R.R., & Sawyer, C.R. (1998). Perspectives on norm-referenced and criterion-referenced grading in introductory speech performance courses. *Journal for the Association for Communication Administration*, 27, 147-153.
- Benson, J.A., & Maitlen, S.K. (1975). An investigation of the relationship between speaking position and rank assignment in forensic competition. *Argumentation & Advocacy*, 11, 183-188.
- Bishop, G.F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, 51 (2), 220-232.
- Bock, D.G. (1970). The effects of persuadability on leniency, halo, and trait errors in the use of speech rating scales. *Speech Teacher*, 19, 296-300.
- Bock, D.G., & Bock, E.H. (1977). The effects of sex on experimenter, expectancy inductions, and sex of the rater, on leniency, halo, and trait errors in speech rating behaviors. *Communication Education*, 26, 298-306.
- Book, C.L. (1985). Providing feedback: The research on effective oral and written feedback strategies. *Central States Speech Journal*, 36, 14-23.
- Buerkel-Rothfuss, N.L., & Gray, P.L. (1990). Graduate teaching assistant training in speech communication and noncommunication departments: A national survey. *Communication Education*, 39, 292-307.
- Carlson, R.E., & Smith-Howell, D. (1995). Classroom public speaking assessment: Reliability and validity

- of selected evaluation instruments. *Communication Education*, 44, 87-97.
- Clauser, B.E., Clyman, S.G., & Swanson, D.B. (1999). Components of rater error in a complex performance assessment. *Journal of Educational Measurement*, 36(1), 29-48.
- Ehrensberger, R. (1945). An experimental study of the relative effectiveness of certain forms of emphasis in public speaking. *Communication Monographs*, 12, 94-111.
- Epstein, H.I. (1997). Scale grading. *The Journal of Irreproducible Results*, 42, 6-8.
- Ford, H. T., Puckett, J.R., & Tucker, L.A. (1987). Predictors of grades assigned by graduate teaching assistants in physical education. *Psychological Reports*, 60(3), 735-739.
- Foster, T.J., Smilowitz, M., Foster, M.S., & Phelps, L. A. (1990). Some student perceptions of grades received on speeches. *Basic Communication Course Annual*, 2, 121-142.
- Freebody, P, and Anderson, R C. (1986). Serial position and rated importance in the recall of text. *Discourse Processes*. 9(1), 31-36.
- Gibson, J.W., Gruner, C.R., Hanna, M.S., Smythe, M., & Hayes, M.T. (1980). The basic course in speech at U. S. colleges and universities: III. *Communication Education*, 29, 1-9.
- 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.

- Golish, T.D. (1999). Student's use of compliance gaining strategies with graduate teaching assistants: Examining the other end of the power spectrum. *Communication Quarterly*, 47(1), 12-32.
- Goulden, N.R. (1990). A program of rater training for evaluating public speeches combining accuracy and error approaches. *Basic Communication Course Annual*, 2, 143-183.
- Goulden, N.R. (1994). Relationship of analytic and holistic methods to raters' scores for speeches. *Journal of Research and Development in Education*, 27(2), 73-82.
- Gray, P.L. (1989). The basic course in Speech Communication: An historical perspective. *Basic Course Communication Annual*, 1, 1-15
- Jersild, A.T. (1929). Primacy, recency, frequency and vividness. *Journal of Experimental Psychology*, 12, 58-70.
- Keyton, J. (2001). *Communication research: Asking questions, finding answers*. Mountain View, CA: Mayfield Publishing Company.
- King, J.L. (1998). The effects of gender bias and errors in essay grading. *Educational Research Quarterly*, 22, 13-25.
- Knower, F.H. (1936). A study of the rank-order methods of evaluating performances in speech contests. *Journal of Applied Psychology*, 24, 633-644.
- Krosnick, J.A., & Alwin, D.F. (1987). An evaluation of a cognitive theory of response-order effects in survey

- measurement. *Public Opinion Quarterly*, 51, 201-219.
- Lance, C.E., LaPointe, J.A., & Fisicaro, S.A. (1994). Tests of three causal models of halo rater error. *Organizational Behavior and Human Decision Processes*, 57(1), 83-84.
- Lawrenz, F., Heller, P., Keith, R., & Heller, K. (1992). Training the teaching assistant. *Journal of College Science Teaching*, 22, 106-109.
- Louden, A., & Shellen, W.N. (1976). A study of ballot feedback from professional and nonprofessional debate judges. *ERIC Document*, 1-12.
- Mason, K.J. (1976). Intercollegiate individual events competitors' speaking order and rankings. Unpublished manuscript, Nebraska Wesleyan University.
- McKeachie, W.J. (1994). *Teaching tips* (9<sup>th</sup> ed.). Lexington, MA: D. C. Heath & Company.
- Morreale, S.P., Hanna, M.S., Berko, R.M. & Gibson, J.W. (1999). The basic communication course at colleges and universities. *Basic Communication Course Annual*, 11, 1-36.
- Murphy, K.R. & Anhalt, R.L. (1992). Is halo error a property of the rater, ratees, or the specific behaviors observed? *Journal of Applied Psychology*, 77(4), 494-499.
- Myers, S.A. (1998). GTAs as organizational newcomers: The association between supportive communication relationships and information seeking. *Western Journal of Communication*, 62(1), 54-73.



- Neuman, W.L. (2000). *Social research methods: Qualitative and quantitative approaches, fourth edition*. Boston, MA: Allyn & Bacon.
- Osgood, C.E., Suci, C.J., & Tannenbaum, P.H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Prieto, L.R., & Altmaier, E.M. (1994). The relationship of prior training and previous teaching experience to self-efficacy among graduate teaching assistants. *Research in Higher Education, 35*, 481-497.
- Quigley, B.L. (1998). Designing and grading oral communication assignments. *New Directions for Teaching & Learning, 74*, 41-49.
- Roach, D.K. (1997). Effects of graduate teaching assistant attire on student learning, misbehaviors, and ratings of instruction. *Communication Quarterly, 45*, 125-142.
- Rubin, R.B. & Yoder, J. (1985). Ethical issues in the evaluation of communication behavior. *Communication Education, 34*, 13-17.
- Rushin, J.W., Saix, J.D., Lumsden, A., Streubel, D.P., Summers, G., & Bernson, C. (1997). Graduate teaching assistant training: A basis for improvement of college biology teaching and faculty development? *The American Biology Teacher, 59*, 86-91.
- Sato, K. (1990). Recency effects and distinctiveness of position/order information. *Perceptual and Motor Skills, 71*, 259-266.
- Young, C. (1998, November). Intercultural adaptation in the classroom: The ethics of grading and assessing

168

*Appropriateness of Instructor Assistants*

**students with minimal proficiency in speaking English. Paper presented at the annual meeting of the National Communication Association, New York.**