Journal of the Association for Communication Administration
30(2001), 104--110

Computer-Assisted Evaluation of Speaking Competencies in the Basic Speech Course

CHRIS R. SAWYER
RALPH R. BEHNKE

WHENEVER a college-level course utilizes a number of instructors and sections, administrators responsible for ensuring the quality of that course become increasingly concerned about issues of equivalency or the extent to which students in the various sections of the course are receiving essentially the same educational experience (Hendrix, 2000; Sawyer & Behnke, 1997). Student complaints can range from a simple concern that section one students have more public speaking assignments than are required in section two, to complex issues about differences in grading standards from instructor to instructor (Colwell, 1996).

Problems of this nature are best solved by proper planning and structuring and before complaints occur (Richardson, 1999). Even the perception of unreliable implementation of grading standards is harmful to a department, since the multi-section basic speaking course is usually central, in many ways, to the general health of the entire academic program. Sawyer and Behnke (1997; 1998) have proposed using technological innovations, such as specialized software packages, as a means of eliminating the problems associated with standardization in multi-section performance courses, such as the basic course in human communication. Recent advances in software development and the trend of integrating assessment in the undergraduate curriculum (*Assessing College Student Competency*, 1994; *Criteria for Assessment of Oral Communication*, 1998; Crocker-Lakness, 1991), however, suggest a larger array of course options available to course planners and administrators. The following essay will describe these trends and will propose a number of innovative solutions to this recurring problem.

STANDARDIZING CONTENT AND METHOD ACROSS SECTIONS

While the highly interrelated issues of content and teaching method should always be major concerns for college instructors, when the basic speech performance course uses several instructors, the department as a whole becomes involved. These con-

cerns are magnified when performance sections are taught by relatively inexperienced graduate students. Consequently, a director of the basic course is often assigned to oversee course operations in order to ensure comparability across sections. In smaller programs, that administrative duty often falls to the department chair or program head. Regardless of the departmental structure, issues of content and method in the basic course become scrutinized more carefully by the administration.

Contemporary Trends

Historically, a variety of speaker evaluation forms have been used in the basic speaking course. While some forms were created by professors well-trained in measurement and evaluation, many were simply "teacher made" with little evidence of reliability or validity. Currently, standard forms, such as the *Competent Speaker Evaluation Form* (Morreale & Taylor, 1991, November), are available that have a record of desirable reliability and construct validity (Carlson, & Smith-Howell, 1995). In addition to content improvements, the contemporary trend toward using computer delivery of speech commentary or criticism has increased (Behnke & Beatty, 1977; Behnke & King, 1984; Behnke & Sawyer, 1987; Sawyer & Behnke, 1997).

The timely delivery of performance comments to students is important. Some instructors take notes during student speeches and then take them home in order to write thoughtful comments about the performances that are then given to the speakers at the next class meeting. Probably the most common tactic is to provide instructor commentary to all speakers at the end of the class period in which the speeches were given. Yet, others insist that comments and criticism are most effective when provided immediately, *during* the presentation (Behnke & Beatty, 1977). Depending on several factors, including availability of computers and various established personal teaching preferences, the rate of infusion of computerized speech criticism into contemporary practice is still an open question. However, in the case of the multi-section/multi-instructor basic public speaking course, this tool is a powerful aid in assuring validity and reliability of content and method across sections.

Historical Perspective

The idea of using computers to compile evaluations and to provide student feedback on public speaking performance is probably based upon early models of computer-assisted instruction (CAI). These early laboratories used very expensive computers reaching well beyond the budgets of most academic departments (Behnke & Derry, 1984; Derry & Behnke, 1983). With the advent of the microcomputer, approaches such as the one advocated here are inexpensive and can be executed successfully on a relatively low powered laptop computer making this system cost effective (Behnke & King, 1984).

Considerable evidence regarding the success of computer-based performance evaluation in speech communication courses has been reported (Behnke & Beatty, 1977; Behnke & Sawyer, 1998; Jurma, 1982; O'Hair, 1984). These scholars focus on the benefits of increased quality, reliability, speed, and efficiency of the method. Instead of producing feedback that is short, terse, and "canned", the computerized system actually encourages extensive, friendly and well-worded commentary. Moreover, its effectiveness is enhanced through repeated editing and is then stored for recall at appropriate times. In fact, the entire, extensive public speaking commentary file can be reviewed by an entire faculty, and the suggestions of students who have taken the course could provide further suggestions for improvement. In a sense, the feedback file represents the wisdom and experience of the whole faculty rather than relying exclusively upon the wisdom and experience of any one particular faculty member or graduate student teacher. Software of this type has been implemented since 1993 in community colleges and private and public

universities. Adoption of this technology is often associated with reductions in student complaints about speech grades, as compared to traditional paper-pencil methods. Appropriately applied, these tactics contribute substantially to continual improvement of the breadth and quality of public speaking instructional feedback that will be received by future students in the course.

Student reactions to computer-aided criticism of public speeches have been favorable indeed (Behnke & Beatty, 1977; Sawyer & Behnke, 1997; 1998). Students said that they liked being told specifically how to improve, expressed positive attitudes toward the instructor, and reported that they were more enthusiastic and self-confident about the assignment. Students often remark that the extensive commentary provided by computer-assisted evaluation enables them to understand how their speech grade was determined and how to improve performance in future assignments. Because much of the *instructor's* behavior during speech performances is obscured by computer equipment, students are less aware of the evaluation process and are able to focus on speaking to the audience.

Instructors using the method reported similar positive reactions. Specifically, they commented on the increased level of involvement they felt while listening to the student speeches and the positive comments students gave them about the process. They were impressed with being able to focus more on the presentations themselves and less on creating and writing evaluative comments during the performances. This phenomenon is sometimes called "the instructor's dilemma" because grading speeches requires engaging in two different processes that cannot be carried out simultaneously: being a good listener while, at the very same time, writing helpful, appropriate, and well-worded, critical commentary. Most of the benefits, described above, pertain regardless of the interval between performance and receipt of instructional commentary. However, the system is most effective in situations wherein short lag times between speaking behaviors, and delivery of comments intended to modify them, are desired or required. Behnke and Beatty (1977) report that only simple, preliminary exposure to the system is required for teachers.

Computerized speech criticism provides many solutions to the validity and reliability problems encountered in a multi-section, multi-instructor public speaking course. The following sections describe the nature of the selected competent speaker form, the delivery software and its availability, and the integration of the two into the recommended computerized speech criticism package.

COMPUTER ASSISTED SPEECH EVALUATION (CASE) SOFTWARE

Delivering CSEF Commentary Via CASE Software

Currently, graduate teaching assistants assigned to the basic course at Texas Christian University use specialized computer software, based on the *Competent Speaker Evaluation Form* (Morreale, Moore, Taylor, Surges-Tatum, Hulbert-Johnson, 1993) to evaluate student speeches and to provide instructional feedback. This software title, Computer-assisted Speech Evaluation (CASE), is a version of the document-modeling package described by Sawyer and Behnke (1998) and runs as part of a software suite called *Intelligent Questionnaire with the WriteOne* (Performance Guild Associates, 1995). All graduate assistants assigned to the basic course are given instruction on the use this package as part of a one-day intensive training session conducted by the course director. Using the *Competent Speaker* evaluation criteria is a major component of the training session. At the conclusion of this one-day workshop, teaching assistants practice operating the CASE software by evaluating video recordings of student speeches. Because the speeches used during the practice evaluation have been evaluated by experts using the

CSEF, comparisons between the expert ratings and those of teaching assistants lead to helpful discussions of the speech evaluation process and its role in instruction.

Procedures for conducting in-class speech evaluations follow the general pattern established in previous computerized speech criticism studies (Behnke & King, 1984; Behnke & Sawyer, 1987) especially those using CASE software (Sawyer & Behnke, 1998). A desktop microcomputer and laser printer, located in the rear of the basic course laboratory, permit teaching assistants to observe student speeches from a position behind the audience. Speeches are presented from the front of the lab room. Speakers are permitted use a podium for their notecards. Presentation times are recorded for each presenter. Consequently, aside from the presence of the evaluator's station, the lab environment is very similar to those in which traditional speech evaluations are conducted.

CASE software is based on "intelligent questionnaire" technology, that is, it prompts the operator to answer questions about an event while that event is being observed, stores the observations, and then assembles a document or report reflecting the stored remarks. Both analytical with holistic evaluations (Goulden, 1994) are supported by the CASE intelligent questionnaire and, in the current version of the software, both are employed. For example, the teaching assistant must provide a holistic evaluation of the student's topic selection when CASE displays the following prompt: "Rate how well <<Student's First Name>> chose and narrowed the topic." CASE then requires the operator to select one of the following descriptors by clicking on the appropriate "radio button," very poor, poor, acceptable, good, very good, or superior. A six-item rating scale of this type is permitted under the guidelines for the CSEF (Morreale, Moore, Taylor, Surges-Tatum, & Hulbert-Johnson, 1993, p. 48) and these descriptors correspond to the competency levels of unsatisfactory (very poor, poor), satisfactory (acceptable, good), and excellent (very good, superior). Each descriptor is weighted in accordance with the Rasch analysis of the CSEF scoring system (Tatum, 1991, November). Although an overall score for each speech is computed from the sum of CSEF items, instructors may choose from among several grading options including a percentage of competencies mastered by the student.

Occasionally, special situations emerge in which instructors decide that some students should receive substantially higher or lower scores than those generated by the software. Consequently, instructors have the option of overriding the numeric evaluation. In one case, a student flagrantly violated the CSEF standard for language by using hate speech and profanity during a presentation. In the judgment of the instructor, the inappropriateness of the student's language warranted a failing grade for the assignment. Consequently, the instructor was able to override the CASE parameters and gave the student a greater penalty than the software was originally designed to assign. Because the performance evaluation was constructed in "real time", the instructor was able to document each specific case of inappropriate language and compose a statement that justified lowering the student's grade. Later, copies of this evaluation, retrieved from the instructor's hard drive, helped to successfully fend off a frivolous grade dispute brought by the student.

Because CASE displays additional prompts that require evaluators to defend their holistic assessments either from prepared lists of possible explanations or to state a rationale for their assessments in free response format, analytical grading procedures are also supported. For example, if the speaker receives a rating of "poor" for topic selection, CASE will display the following screen asking the grader to justify the evaluation by clicking one or more checkboxes:

Why was <<Student's First Name>>'s topic selection unsatisfactory? Was it ...

- ❑ Inconsistent with the speech purpose?
- ❑ Inappropriate for time constraints?
- ❑ An example of poor audience analysis?

Each checkbox option represents a statement derived from the CSEF criteria. Under CSEF performance standards, a speaker's topic selection will receive an unsatisfactory rating if the topic is deemed inappropriate for the audience, the purpose, or the time constraints of the assignment (Morreale, et al, 1993). In the hypothetical case above, suppose that the speaker was assigned to present an informative speech but prepared a persuasive one instead and could not cover the topic adequately within the time limits of the assignment. Consequently, the chief faults in topic selection are inconsistency with the speech purpose and violation of time constraints. Consequently, the student will receive the following statements on the CASE output:

> "Your topic was inconsistent with the purpose of the speech and could not be treated adequately in the time limits for this assignment."

Ratings of acceptable to very poor open additional screens that prompt the evaluator to selected corrective advice for the faults detected in a student's performance. These statements, which vary in length from one sentence to a full paragraph of text, also appear on the student's evaluation sheet. It is plain to see that the volume of evaluation statements produced by the intelligent questionnaire exceeds what most teachers would feel comfortable writing by hand, especially during student performances.

Course Record Keeping and Database Management

In addition to the capacity of providing extensive instructional feedback to students immediately upon conclusion of the grading session, intelligent questionnaires, such as CASE, have the capacity to assist instructors with the difficult tasks of record keeping and standard setting. Observations of student performances recorded on CASE are stored on the computer hard drive and are useful during student-instructor conferences and tutorial sessions. Intelligent questionnaires store the statements justifying an evaluation as well as the numeric, and CASE also generates numeric grades based on how well students demonstrate mastery over the CSEF criteria. These scores are stored along with the other evaluation records and exported to a database file, such as Excel, FoxPro, or FileMaker.

Because intelligent questionnaires permit storage of performance evaluations in database form, differences in grading practices among teaching assistants can be identified and remediated by the course director. For example, the effects of halo and leniency rating errors on performance are well-documented (Woehr & Huffcutt, 1994). A wide difference among lab grades is an oft-cited source of student complaint in basic communication courses (Buerkel-Rothfuss & Gray, 1990; Hawkins, 2000, March). Lab assistants, whose speech evaluations are restricted to either the high or the low ends of the grade scale, should be retrained with a view to correcting their particular rating errors.

CONCLUSIONS

Aside from the advantages in the arenas of grading, record keeping, and establishing performance standards, recent advances in computer software promise additional intelligent questionnaire benefits. Educators have recommended the use of student portfolios in courses featuring public speaking (Jensen & Harris, 1999). In this instructional strategy, a student's work during a course is compiled and reviewed periodically through-

out the academic term. Students often find these activities valuable learning and self-discovery experiences, especially as assessments of their communication competence accumulate over time. This emerging self-perception of one's own ability and growth encourages continued work and improvement. An intelligent questionnaire designed specifically for the purpose of compiling a comprehensive record of student progress would be a welcomed relief for course directors who must manage the intense record keeping required by course portfolios.

Recently, the European Community has funded a major effort to establish international standards for intelligent questionnaires (http://www.epros.ed.ac.uk/iqml). These standards will permit the development of cross-platform software packages that will enable communication scholars to construct data warehouses by supporting an array of data collection techniques, including telephone, e-mail, and web-based surveys in addition to instructor generated performance evaluations. In addition, the next generation of intelligent questionnaire users will have the capacity to seamlessly access on-line information sources without prior knowledge of the database's architecture. These improvements will dramatically enhance record keeping for both teaching and research purposes.

REFERENCES AND NOTES

Chris R. Sawyer, (Ph.D., University of North Texas, 1992), is Assistant Professor in the Department of Speech Communication at Texas Christian University, PO Box 298045, Fort Worth, Texas 76129.

Ralph R. Behnke, (Ph.D., University of Kansas, 1966) is Professor in the Department of Speech Communication at Texas Christian University, PO Box 298045, Fort Worth, Texas 76129.

Behnke, R. R. & Beatty, M. J. (1977). Critiquing speaker behavior through immediate video display. *Communication Education, 26,* 345-348.

Behnke, R. R. & Derry, J. O. (1984). Design and development of a computerized audio/video laboratory. *New Advances in Learning Laboratory Developments Journal, 18,* 8-13.

Behnke, R. R. & King, P. E. (1984). Computerized speech criticism. *Communication Education, 33,* 173-177.

Behnke, R. R. & Sawyer, C. R. (1987). Microcomputer feedback in communication training. *Western Journal of Speech Communication, 51,* 127-135.

Buerkel-Rothfuss, N. L. & Gray, P. L. (1990). Graduate teaching assistants 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.

Colwell, G. (1996). *You won't believe this but...: Responding to student complaints.* Calgary, Alberta: Detselig Enterprises.

Crocker-Lakness, J., Manheimer, S., & Scott, T. (1991). *SCA policy on criteria for the assessment of oral communication.* Annandale, VA: Speech Communication Association.

Derry, J. O. & Behnke, R. R. (1983). Instantaneous feedback in the teaching/learning laboratory. *Educational Technology, 23,* 28-30.

Goulden, N. R. (1994). Relationship of analytic and holistic methods to raters' scores for speeches. *Journal of Research and Development in Education, 27,* 73-82.

Hawkins, K. (2000, March). *Recruiting, selecting, and orienting graduate teaching assistants for the basic course in oral communication.* A paper presented at the annual meeting of the Southern States Communication Association.

Hendrix, K. G. (2000). *The teaching assistant's guide to the basic course.* Belmont, CA: Wadsworth.

Jensen, K. K. & Harris, V. (1999). The public speaking portfolio. *Communication Education, 48,* 211-227.

Jurma, W. E. (1982). Improving discussion skills with immediate video feedback. *Journal of Business Communication, 57,* 268-269.

Morreale, S. P. & Taylor, K. P.(1991, November). *The Competent Speaker Assessment Instrument: Uses and logistics for administration.* Paper presented at the annual meeting of the Speech Communication Association.

Morreale, S. P., Moore, M. R., Taylor, K. P., Surges-Tatum, K. P. & Hulbert-Johnson, R. (Eds.). (1993). *The Competent Speaker Speech Evaluation Form.* Annandale, VA: Speech Communication Association.

National Communication Association. (1995). *Assessing college student competency in speech communication: 1994 NCA summer conference proceedings.* Annandale, VA: National Communication Association.

National Communication Association, (1995). *Criteria for assessment of oral communication.* Annandale, VA: National Communication Association.

O'Hair, H. D. (1984). Interactive video feedback during medical interviewing training. *Biomedical Communications, 12,* 22-27.

Performance Guild Associates (1995). *Intelligent Questionnaire with the WriteOne: How the inquisitive professional reports information.* Grandbury, TX: Performance Guild Associates.

Richardson, S. M. (1999). *Promoting civility: A teaching challenge.* San Francisco: Jossey-Bass.

Sawyer, C. R. & Behnke, R. R. (1997). Technological approaches for improving grading efficiency and compatibility in multi-section/multi instructor communication courses. *Journal of the Association for Communication Administration, 27,* 163-169.

Sawyer, C. R. & Behnke, R. R. (1998). New wave computer technology and the administration of speech communication performance courses. *Journal of the Association for Communication Administration, 28,* 1-6.

Tatum, D. S. (1991, November). *The competent speaker speech performance evaluation: A Rasch analysis.* Paper presented at the annual meeting of the Speech Communication Association, Atlanta, GA.

Woehr, D. J. & Huffcutt, A. I. (1994). Rater training for performance appraisal. *Journal of Occupational and Organizational Psychology, 67,* 189-205.