

2018

Assessing Evaluation Fidelity Between Students and Instructors in the Basic Communication Course: The Impact of Criterion-Based Speech Evaluation Training

T. Kody Frey

University of Kentucky, terrell.frey@uky.edu

Cheri J. Simonds

Illinois State University, Cjsimon@ilstu.edu

John Hooker

Illinois State University, jfhooker@ilstu.edu

Kevin Meyer

Illinois State University, kmeyer@ilstu.edu

Stephen Hunt

Illinois State University, skhunt2@ilstu.edu

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



Part of the [Communication Commons](#), and the [Higher Education Commons](#)

Recommended Citation

Frey, T. Kody; Simonds, Cheri J.; Hooker, John; Meyer, Kevin; and Hunt, Stephen (2018) "Assessing Evaluation Fidelity Between Students and Instructors in the Basic Communication Course: The Impact of Criterion-Based Speech Evaluation Training," *Basic Communication Course Annual*: Vol. 30 , Article 4.

Available at: <https://ecommons.udayton.edu/bcca/vol30/iss1/4>

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, mschlangen1@udayton.edu.

Research Article

Assessing Evaluation Fidelity Between Students and Instructors in the Basic Communication Course: The Impact of Criterion-Based Speech Evaluation Training

*T. Kody Frey, University of Kentucky
Cheri J. Simonds, Illinois State University
John F. Hooker, Illinois State University
Kevin R. Meyer, Illinois State University
Stephen K. Hunt, Illinois State University*

Abstract

This study¹ investigates the role of speech evaluation training in a) creating speech evaluation fidelity between instructor scores and student self-evaluation scores and b) facilitating the type and quality of written feedback on speeches by both students and instructors. The results suggest that students who undergo speech evaluation training achieve a higher level of evaluation fidelity with their instructors. Second, negative feedback by instructors and students significantly predicted the score provided on the speeches. Finally, students who received speech evaluation training provided significantly more

¹ This study is a follow-up to the thesis of the first author and was submitted for initial review under the previous editor of this journal. The authors would like to thank their coders, James Ndone and Adam Mason, for their due diligence in this project.

constructive comments on their evaluations than students who did not receive training. These findings reiterate the necessity for speech evaluation training while offering practical implications for revising training methods and preparing individuals for providing effective feedback.

Keywords: speech evaluation training, evaluation fidelity, written speech feedback, speech performance

In many institutions of higher education, students complete general education courses in order to earn their degrees regardless of their chosen field of study, and one of the tenets of general education is that students are exposed to similar learning objectives in foundational classes (Mazer, Simonds, & Hunt, 2013). Assessing the desired outcomes in these courses should be of utmost importance for teachers, learners, and administrators. Gardiner (1994) explained that faculty should aim to monitor, develop, and improve programs through continual assessment for the purposes of enhancing both student learning and understanding of expectations as well as providing institutions with evidence of both educational quality and accountability. Furthermore, research indicates that, due to these objectives, assessment is most informative when it is conducted in actual classroom contexts (Benander, Denton, Page, & Skinner, 2000; Sprague, 1993).

Assessment in the basic communication course is especially vital, given that this course aims to establish a common understanding and acquisition of skills necessary for students to progress in their education and, ultimately, their professions (Allen, 2002; Hunt, Simonds, & Hinchliffe, 2000). These foundational skills have been championed by the Association of American Colleges and Universities (AAC&U) as an essential learning outcome of general education including inquiry and analysis, critical and creative thinking, written and oral communication, information literacy, and teamwork and problem solving (AAC&U, 2010). Given the centrality of the introductory communication course for many general education programs, basic course directors are faced with the challenge of standardizing how the course is administered and graded across multiple sections by different instructors with varying backgrounds and experience levels. A level of standardization is also critical to programmatic and inter-institutional assessment efforts.

One way to address this challenge, especially when it comes to the evaluation of student speeches, is to link criterion-based grading with instructor and student

training. Stitt, Simonds, and Hunt (2003) discussed the need for basic course directors to establish speech evaluation training programs that facilitate evaluation fidelity. Evaluation fidelity represents a “shared understanding of meaning between those doing the evaluating and those being evaluated in terms of established performance criteria” (Stitt et al., 2003, p. 344). The objective of this approach is for instructors and students to come to a shared understanding of the performance expectations to earn a particular grade on a speaking assignment. To establish this shared understanding, both instructors and students need to be trained in the speech evaluation process, including strategies for providing meaningful written feedback. Simonds (1997) identified the clear communication of expectations as a subtype of clarity known as process clarity. One of the key elements of process clarity is fostering student understanding of how grades are determined.

The current study examines the veracity of the claim that training students to use criterion-based grading will improve evaluation fidelity in the basic communication course. Scholars have established that rater training is an essential component of criterion-based evaluation (Eckes, 2008; Harsch & Martin, 2013; Lumley, 2005; Weigle, 1999). In terms of speech evaluation, previous research demonstrates that a robust training program can improve evaluation fidelity among raters (Stitt et al., 2003). The current study extends the programmatic research regarding evaluation fidelity by examining the impact of speech evaluation training using criterion-based grading on instructor and student feedback.

Review of Literature

Criterion-Based Assessment

Criterion-based assessment involves the elimination of competition for grades and the use of clear performance expectations that an instructor sets in advance of an assignment (Dominowski, 2002). According to Brookhart (2013), this approach to evaluation provides “a set of criteria for students’ work that includes descriptions of levels of performance quality on the criteria” (p. 4). Criterion-based assessment relies on rubrics with clear, specific explanations of what is required of students to meet different levels of performance on an assignment. With this method of evaluation, students must develop familiarity with instructor expectations and gain extensive knowledge of the criteria used to succeed (Stitt et al., 2003). Criterion-based assessment holds numerous advantages as a pedagogical strategy including facilitating

self-evaluation, promoting students' understanding of expectations, making grades more meaningful, and promoting deep learning (Brookhart, 2013).

While each instructor may have her or his own clear set of assignment requirements, students taking the same communication course from different instructors should be able to achieve common general education learning outcomes as they are operationalized in the basic course. Thus, basic course directors and faculty should develop clear assessment measures related to these outcomes. Stitt et al. (2003) clarified this argument in the following terms:

Criterion-based assessment is one very effective way of making assessment accessible for students and instructors. Providing students with specific criteria that they must meet to obtain a particular grade should decrease students' uncertainty about teachers' expectations. Also, the development of such criteria should be useful in the training of instructors to assess student performance—assessment can be standardized based on these criteria. (p. 343)

Thus, for criterion-based assessment to be effective, faculty should attempt to intentionally and deliberately address uniformity in grading criterion across sections for major assignments and clearly communicate those criteria to students. One of the most important areas for assessment in the basic course, and perhaps one of the most relevant areas for testing specific criterion-based assessment systems for achieving this goal, is speech evaluation training (Frey, Hooker, & Simonds, 2015).

Speech Evaluation Training

One of the key learning objectives in most basic communication courses is competent speech delivery. To reliably assess if our courses facilitate this objective, instructors need to be trained to evaluate student speeches. Although she was referencing written evaluation training, Charney (1984) argued that systematic training “procedures are designed to ‘sensitize’ the readers to the agreed upon criteria and guide them to employ those standards, rather than their own” (p. 73). Students benefit from speech evaluation training to the extent that it clarifies and operationalizes how their performances will be assessed (Goulden, 1990). Similarly, speech evaluation training should help students improve their oral communication skills by promoting better understanding of desired speaking competencies

referenced in the criteria. Finally, a systematic training program allows for programmatic assessment efforts because it creates similar assignment expectations for students and instructors across multiple sections of the course.

In their seminal study on speech evaluation training, Stitt et al. (2003) tested a criterion-based speech evaluation program and found that instructors improved their rater reliability after participating in the program. This study also revealed that a group of students who were given the criteria and trained how to apply it to sample speeches demonstrated greater evaluation fidelity with instructors than a group who did not receive training. While Stitt et al. (2013) used actual student speeches, the results revealed several modifications that could be made to improve the training program including a more robust set of criteria and models of expected performance. These models of expected performance were written and performed to be exemplars of “A” quality and “C” quality speeches based on the modified criteria. Although Stitt et al. (2003) broke new ground with their work in this area, several questions remain about which specific teaching strategies might best promote evaluation fidelity. First, is mere exposure to the speech evaluation criteria sufficient to facilitate meaningful evaluation fidelity? Alternatively, is exposure to the criteria coupled with training necessary to increase evaluation fidelity between students and instructors on actual graded speeches rather than sample speeches? Additionally, what role does speech evaluation training play in the type and quality of written feedback students provide? Answering these questions will demonstrate how instructors can best use scarce class time to prepare students to perform speeches that are evaluated using criterion-based grading. While previous studies have determined that speech evaluation training can improve evaluation fidelity (Simonds et al., 2009; Stitt et al., 2003), results also indicated that instructors and students need more training in terms of providing quality written feedback.

Written Feedback

It is important that students not only understand performance expectations, but that they also critically reflect on how their performance aligns with established criteria. Students who critically reflect on high quality written feedback should experience improvements in speech performance as they progress through the course. Previous research shows high levels of evaluation fidelity for scores between instructors and the students whom they have trained (Mazer et al., 2013; Stitt et al., 2003); however, there remains a dearth of literature examining the quality of student

written feedback (LeFebvre, LeFebvre, Blackburn, & Boyd, 2015) or how that feedback can lead to students to improve on future assignments. In order to advance pedagogical content knowledge in this area, researchers must explore the quality of written speech feedback as a tool to communicate to students which aspects of their speech performance align with, exceed, or fall short of expectations expressed in the criteria (Reynolds, Hunt, Simonds, & Cutbirth, 2004).

Simonds et al. (2009) evaluated the type and nature of the feedback used by instructors on their students' evaluation forms and identified the following four specific types of comments instructors give to students: positive non-descriptive, positive descriptive, negative, and constructive (see Appendix A for descriptions of each type of feedback). Their research led to two distinct findings. First, by examining instructors' written feedback on students' evaluation forms, they found a positive, linear relationship between positive comments and student grades. Second, the results indicated that instructors do, in fact, incorporate language from the speech criteria into their written feedback. However, the feedback included mostly objective, descriptive comments rather than clarifying, prescriptive comments. Instructors often emphasized the use of comments reflecting active behaviors of the speaker rather than offering suggestions for future development. Indeed, research by Reynolds et al. (2004) clearly demonstrated that students desire more negative face-threatening comments suggesting specific methods of improvement rather than simple descriptions of their behaviors. Mazer et al. (2013) extended these findings through an assessment of written speech feedback comments made by students on self-evaluations of their own speeches. Further, Mazer et al. (2013) found that instructors often fail to train students to use speech evaluation criteria to justify the numerical scores they assign. Thus, additional research is needed to examine the effects of speech evaluation training on the types of written feedback instructors and students employ when evaluating speeches and the relationships between written feedback and numerical speech scores. Research in this area will equip basic course directors with the information they need to improve their training programs and lay the foundation for wider programmatic assessment of student learning outcomes (Mazer et al., 2013).

Hypothesis and Research Questions

Students often report positive learning outcomes and increased clarity of grading criteria when instructors utilize criterion-based grading (Topping, 1998). When

considering the previous research conducted on speech evaluation assessment, students who receive training should possess the ability to score a speech in the same manner as the instructor. Considering this argument, the following hypothesis and research questions are proposed:

H1: There will be greater evaluation fidelity (agreement between speech scores) between instructors and students when students have been trained to use speech evaluation criteria.

Because previous research suggested improvements to instructor and student training in terms of written speech feedback (Mazer et al., 2013; Reynolds et al., 2004; Simonds et al., 2009), this study advances the following research questions to examine relationships between the types of feedback employed and numerical speech scores.

RQ1: Does the type of instructor feedback predict instructors' scoring of student informative speeches?

RQ2: Does the type of student feedback predict students' self-scoring for the informative speech?

In addition, students who received the training program were compared with those who did not receive training to determine if the program made a difference in feedback type and frequency.

RQ3: Do students who receive criterion-based training use different types of feedback than students who do not?

Method

The researchers sent a call to participate to second-year graduate teaching assistants of the basic communication course at a large Midwestern university. Researchers selected this group under the assumption that second-year instructors with teaching experience and syllabus flexibility could easily incorporate the study design into their course schedules. Six instructors voluntarily agreed to participate in the study. We then randomly assigned the six participating instructors into two

independent conditions; three instructors into a control group and three into an experimental group. The three instructors in the control group informed their students that speech evaluation materials were available for them to use, yet they did not provide formal training. Instructors in the experimental group provided their students with formalized speech evaluation training (see Appendix A). Although we did not perform a manipulation check *per se*, we did communicate with instructors in both groups to ensure they were following training protocols.

Participants

Participants included all individuals enrolled in the participating instructors' basic communication course classes ($n = 84$). It was important to the researchers that the study be conducted in the context of actual classroom conditions with intact groups to allow for formative assessment of the speech evaluation program and to increase ecological validity. The redaction of student names resulted in extra protection of participant confidentiality, yet it also shielded important demographic information. All students who participated in the research through the collection of their instructor and self-evaluation forms provided informed consent to contribute information to the study.

Coding Procedures

Speech evaluation materials were collected at the close of the fall semester from all students enrolled in communication courses taught by the participating instructors ($n = 6$). To facilitate a direct comparison, instructors completed the evaluation of student speeches both during and after the speech (see evaluation form in Appendix A), while students conducted a self-evaluation using the same evaluation form after watching a video-recording of their speech. Only complete sets of speech materials, including instructor and student self-evaluation forms for the informative speech, for the experimental group ($n = 40$) and the control group ($n = 44$) were included in this study.

Speech evaluation materials were content analyzed using the objective and systematic procedures described by Kaid and Wadsworth (1989). Accordingly, the researchers analyzed the data based on the categories established in Simonds et al. (2009). To address the hypothesis and answer the research questions, a code book was designed to record the number of each type of comment (positive non-descriptive, positive descriptive, negative, and constructive) for each category of

evaluation (outline, introduction, body, conclusion, delivery, overall impression) for both the instructor and student self-evaluations (see Appendix B). Scores for each category of evaluation and total scores for each speech were recorded on a code sheet for speech evaluation (see Appendix C).

Next, two coders were trained by the researchers to implement the coding process. The coders then independently analyzed 10% of the sample sets ($n = 9$) to assess intercoder reliability for all categories. Each coder reviewed the data and placed instructor and student feedback comments into a respective category. Each of these comments were numbered and unitized to facilitate a direct comparison of discrepancies for discussion. For instructor comments, the coders achieved an agreement of 92.9% (Cohen's $\kappa = .85$). The coders achieved agreement on 93.8% of student comments (Cohen's $\kappa = .86.9$). Importantly, a coding reliability coefficient, measured with Cohen's κ , of .75 or greater is considered excellent (Fleiss, 1981; Neuendorf, 2002). Upon completion of the independent analysis for intercoder reliability, the researchers located points of disagreement, established coding rules, and repeated the process where the two coders came to 100% agreement on placement of categories. During this process, the coders discovered a few discrepancies occurred due to a misunderstanding in terms of past or present tense on instructor feedback. The coders then came to an agreement about how to code feedback tenses and the code book was modified accordingly (see Appendix B, tenses). Additionally, the coding process revealed another area of discrepancy in terms of instructor notes, which led to another modification of the code book (see Appendix B, instructor notes). One other source of discrepancy was a result of coder fatigue. The coders agreed that the remaining data would be coded in shorter time increments of no more than a three-hour time frame per coding session. The coders then divided the remaining data sets for content analysis. This iterative procedure helped determine that the categories provided by Simonds et al. (2009) fit the data well.

Results

The hypothesis posited greater evaluation fidelity between instructor and student scores on an informative speech when students received training to use the evaluation criteria. A bivariate correlation was run to address the hypothesis. Results indicated a significant, positive correlation between instructor-assigned grades and student self-scoring for the students who received criterion-based training, $r(37) = .71, p < .001$. We also observed a moderate, positive correlation between instructor-

assigned grades and student self-scoring for the untrained students, $r(42) = .58, p < .001$. We employed a Fisher r to z transformation to determine if the correlations we observed were statistically different. The transformation failed to reveal a difference, which is likely a result of a small sample size, $z = .96, p > .05$. However, an examination of the confidence intervals for both correlations reveals a smaller range for the .71 correlation, 95% CI [.50, .84], compared to the .58 correlation, 95% CI [.34, .75]. As Allen, Titsworth, and Hunt (2009) noted, smaller CIs “provide more precision and therefore result in more credible conclusions” (p. 24). Therefore, we are most confident in the results reported for the experimental group which supports the hypothesis.

The first research question asked whether the type of instructor feedback predicted instructors' scoring of student informative speeches. The four instructor feedback categories were entered as predictor variables in a multiple linear regression procedure, with instructor grade serving as the outcome variable. The four feedback categories predicted 77.1% of the variance in instructor grade, $R^2_{adj} = .759, F(4, 79) = 66.50, p < .001$. Analysis of regression coefficients revealed that three of the four feedback categories significantly predicted instructor grade. The strongest individual predictor was instructor negative comments, $\beta = -.801, t = 14.45, p < .001$, followed by constructive comments, $\beta = -.220, t = 4.03, p < .001$, and positive descriptive comments, $\beta = .131, t = 2.19, p = .03$. Squared part correlations indicated that negative comments uniquely predicted 60.5% of the variance in instructor-assigned grades, while constructive comments and positive descriptive comments uniquely predicted 4.7% and 1.7% of the variance, respectively. Positive nondescriptive comments were not a significant individual predictor. Tolerance and VIF statistics did not indicate collinearity among variables. Beta weights are in Table 1.

Table 1
Beta Weights for Instructor Grades (RQ1)

Predictor Variables	<i>B</i>	<i>SE B</i>	β
Positive Nondescriptive	.049	.043	.066
Positive Descriptive *	.517	.236	.131
Negative *	-1.479	.102	-.801
Constructive *	-.723	.179	-.220
R^2		.771	
R^2_{adj}		.759	
<i>F</i>		66.500	

Note. An * indicates a unique significant predictor variable at $p < .05$. ($n = 83$)

The second research question explored whether the type of student feedback predicts students' self-scoring for the informative speech. The four feedback categories predicted 17.0% of the variance in student self-scoring, $R^2_{adj} = .127$, $F(4, 78) = 3.99$, $p < .01$. One of the four feedback categories significantly predicted student self-scoring. The only significant individual predictor was student negative comments, $\beta = -.384$, $t = 3.37$, $p = .001$. The squared part correlation indicated that student negative comments uniquely predicted 12.1% of the variance in student self-scoring of the informative speech for the trained students. The other three feedback types were not significant individual predictors. Tolerance and VIF statistics did not indicate collinearity among variables. Beta weights are in Table 2.

Table 2
Beta Weights for Student Self Grades (RQ2)

Predictor Variables	<i>B</i>	<i>SE B</i>	β
Positive Nondescriptive	.200	.104	.224
Positive Descriptive	.253	.302	.091
Negative *	-.882	.261	-.384
Constructive	-.282	.202	-.147
R^2		.170	
R^2_{adj}		.127	
<i>F</i>		3.991	

Note. An * indicates a unique significant predictor variable at $p < .05$. ($n = 82$)

The third research question asked if students who receive criterion-based training use different types of feedback than students who do not. A series of independent samples *t*-tests were conducted, with the types of student feedback entered as the dependent variables and the student groups (trained versus untrained) as the independent variable. Levene's test for equality of variances indicated no violation of test assumptions for positive descriptive and negative student comments. Positive descriptive comments were not significantly different, $t(82) = .61, p = .53$, between the trained students ($M = 1.57, SD = 1.98$) and the untrained students ($M = 1.20, SD = 3.28$). Negative comments were not significantly different, $t(82) = .03, p = .97$, between the trained students ($M = 2.95, SD = 2.79$) and the untrained students ($M = 2.97, SD = 3.76$). Levene's test was significant for constructive and positive non-descriptive comments, so equality of variances could not be assumed, and statistical corrections were made. Constructive comments were significantly different, $t(44.49) = 5.53, p < .001$, between trained students ($M = 4.80, SD = 4.66$) and untrained students ($M = .56, SD = 1.30$). Positive nondescriptive comments were not significantly different, $t(81.82) = .69, p = .49$, between trained students ($M = 6.40, SD = 7.88$) and untrained students ($M = 7.68, SD = 9.09$).

Discussion

This study extends the line of research on speech evaluation fidelity in two ways. First, the research examined speech evaluation fidelity as it occurred between instructor and student self-assessed speech scores using speech evaluation training compared to simply making the criteria available. Second, the investigation examined the written speech feedback comments provided by students and instructors when evaluating student speeches. By understanding these relationships, researchers and educators can further develop and assess the clarity needed when administering assignments in the basic communication course.

In terms of the hypothesis, the results of the study revealed more confidence in the larger correlation observed in the experimental group compared to the control group. Analysis demonstrated that students in the control group experienced moderate, positive correlations with their instructors' ratings; however, students in the experimental group yielded strong, positive correlations between their ratings and the score provided by their instructor. Therefore, the experimental group that received the criteria and training had higher evaluation fidelity with their instructors than the control group. This has implications for both pedagogy and assessment.

The difference between conditions was exposure to formalized training, yet all students had the same access to speech evaluation resources. Though students have access to the criteria and models of expected performance long before the delivery of their first formal speech, the students' awareness of these resources alone sans training on their application does not allow them to apply the criteria in the same way as their peers who received training. The results echo the claim by Topping (1998) that by discussing and clarifying the existing criteria through training, students perceive greater clarity regarding what constitutes high-quality work. This demonstrates a need for greater emphasis on teaching students to evaluate their own speeches. Speech evaluation training provides clarity for students in terms of what is expected of them as well as how to meet those expectations and allows them to better reflect on their performance in terms of established criteria.

Basic course faculty should provide meaningful descriptive and constructive feedback that provides students with opportunities to improve over time. For research question one, the content analysis indicated that instructors relied heavily on negative comments to justify scores followed by the more meaningful constructive and positive descriptive comments. Recall that constructive comments provide future direction for improvement, whereas positive descriptive comments provide repeatable behaviors for continued strong performance. While it is encouraging that instructors are using more meaningful and instructive comments (constructive and positive descriptive) in determining scores, it is clear that they should be providing constructive comments more often.

It appears that instructor training has improved their ability to provide more negative face-threatening comments (Reynolds et al., 2004) in terms of negative and constructive comments, but there is more room for improvement. In providing training to instructors, we should be more intentional in equipping them with specific prompts to help them move from a reliance on less descriptive to include more instructive comments. Also, we can modify the training to include specific verbiage to move from positive non-descriptive to positive descriptive comments to provide future repeatable behaviors. For example, *continue to (engage in specific behaviors that move above and beyond a behavior listed on the evaluation form)* will move instructors from providing positive non-descriptive comments to the more instructive and descriptive comments. Likewise, *try, avoid, or be careful with (a specific behavior that will allow improvement over time)* will help them move from negative to constructive comments. Future instructor training should provide multiple examples of both descriptive and constructive comments for each of the behaviors listed on the

evaluation form. Researchers should explore creating an electronic speech grading platform where instructors can choose from a variety of meaningful comments associated with the level of student performance on a given behavior. These modifications to the speech evaluation training and grading program will allow us to collect more reliable assessment data in the future and to continue this programmatic research inter-institutionally. For example, the National Communication Association recently funded a group of basic course directors to begin such assessment efforts. Also, the Social Science Research Council is currently engaged in a project to create national learning standards for the basic public speaking course. Future research can begin to identify a common rubric to measure these learning outcomes that can be used to conduct inter-institutional assessment. Such assessment efforts will arm basic course directors with a strong rationale for inclusion of the course in general education programs across the country.

For research question two, our analysis revealed that students relied heavily on negative comments to justify their self-evaluation scores as well. While students appear to be comfortable in providing negative (albeit not instructive) comments, they too should be trained on how to move from less descriptive to instructive comments. In addition to the suggestions for modified instructor training above, we can modify the self-evaluation form to ask students to specifically identify elements they particularly *enjoyed* about their presentation for a continued successful performance and what specific *advice or alternatives* they could suggest for future improvement. Speech evaluation training programs must consider adding more emphasis on teaching students to provide more detailed, prescriptive feedback comments that offer suggestions for future growth and improvement.

The third research question examined the influence of criterion-based training on the types of feedback that students provide in their self-evaluations. Our analysis of the data revealed a significant difference in terms of constructive feedback. That is, students who received speech evaluation training provided more meaningful and constructive feedback than students who were not trained. While this finding supports the need for training students to use the criteria, there remains a need to improve student use of positive descriptive comments. Students should be more concerned with providing in-depth, reflexive feedback that details exactly which behaviors led to their success and what behaviors need to be replicated for continued success. This provides even more evidence that with improved training, students can be taught to provide effective feedback and to appropriately use feedback to determine scores. Again, these results will allow us to make modifications to the

training program, advance this programmatic research, and foster continued larger-scale assessment efforts.

As the basic communication course often features three or more speeches, perhaps future evaluation efforts should allow instructors more time to properly demonstrate how to use positive descriptive or constructive feedback to improve speeches early in the course. Then, students may be allowed to take responsibility of their own learning and demonstrate their knowledge of the routine criteria by applying it to their own speeches at a later point in the course. This type of assessment effort would mirror the reflective learning often available through portfolio assessment in the basic course (Hunt et al., 2000) as students will have the ability to compare their own speech feedback comments to those previously offered by their instructor.

Limitations and Suggestions for Future Research

Researchers and scholars routinely use assessment efforts to monitor and improve the basic communication course. While the current study provides valuable insight into the refinement of basic course pedagogy, one must consider some important limitations. First, the use of intact classrooms presents any number of potentially hidden confounding variables. For example, one of the instructors in the experimental group who voluntarily agreed to incorporate the study design into the class schedule taught an honors section of the basic course. The possibility exists that students within this section exhibited greater desire to learn, achieve high scores, and demonstrate their competence to the instructor in comparison to traditional classes. While students with intellectual ability like those in the honors section could feasibly enroll in any section of the course, the inclusion of the honors section serves as a potentially confounding variable.

Given this potential limitation, it is important that readers understand why we made the decision to use intact groups. Initially, we were interested in exploring how students respond to training, or the lack thereof, with the context actual communication classrooms (e.g., ecological validity). In order to meaningfully evaluate the pedagogical variables of interest, we needed the instruction to happen in a context where students worked together to learn concepts related to speech evaluation. Such an intervention certainly would not have been appropriate for a zero-history group. In other words, it would not be methodologically possible to conduct this research outside of the context of the classroom. Again, we recognize

there are limits to this approach, as there are with any research design; however, the current study yielded information that allowed us to improve our speech evaluation training. The improved evaluation training protocols can now serve as the basis for larger, programmatic assessment of student learning in the basic course. Finally, the regressions reported in answering the second research question provide information that transcend the inclusion of honors students in the experimental condition. Taken together, the results of the current study inform and advance our understanding of evaluation fidelity. Ultimately, we agree with Tincani and Travers (2017), that studies employing designs like ours should not be automatically rejected on a *prima facie* basis as such a decision would contribute to a “file drawer” effect in our discipline. Clearly, there were significant benefits to testing the interventions employed in this study in intact classes.

Moving forward, future assessment efforts should identify new ways of testing and evaluating students’ abilities to apply and understand standardized grading criteria. Future research should assess if training modifications for both instructors and students can improve the quality of feedback to move from less descriptive (positive non-descriptive and negative) to instructive (positive descriptive and constructive) comments. Also, future scholars could employ more control in ensuring as much similarity across the groups as possible (e.g., eliminating the inclusion of special sections designated for honors or comparing multiple honors sections). In addition, studies that employ larger sample sizes should allow future scholars to come to more conclusive findings regarding the influence of criterion-based training on evaluation fidelity.

Research efforts should also attempt to analyze the language students use when providing feedback, rather than simply testing for the presence of feedback. Ideally, comments should reflect the same language used in the establishment of the evaluation criteria and that language should be reflected in the score. Instead of simply categorizing remarks made by instructors or students, research should examine whether the language used truly reflects the criteria on which it is based. Finally, content analysis generally serves only to describe the available data. This type of procedure attempts to report on the identification of specific trends to provide support for findings and conclusions. Consequently, this descriptive process may conceal underlying motives for observed patterns such as those reported here. Research methods like in-depth interviews or focus groups could reveal more exhaustive information about students’ perceptions of the speech evaluation process.

A second limitation is the instructional time for training. For instructors not already incorporating speech evaluation training into their course schedules, the process may not have adequately fit within the respective class structure. Speech evaluation training involves certain activities and examples that typically require an entire class period to illustrate effectively. To meet course goals, general education outcomes, and speech requirements, fitting course content and activities into the time frame of the course is challenging. Future research should investigate ways to offer this training as an online module to protect valuable instructional time.

Overall, these findings support future facilitation of speech evaluation training for students in the basic communication course. The data suggest that instructors and students can achieve strong levels of evaluation fidelity. However, students must receive not just the rubric for criterion-based assignments but also training on how to implement it themselves through applying it to exemplars. Future speech evaluation training efforts should focus on teaching students the importance of providing effective self-evaluation feedback to clarify what is being asked of them in the assignment. In addition, future research should examine if speech evaluation training can improve the evaluation fidelity between instructors, students, and peers as well as improve students' abilities to critically reflect on speech performances through peer evaluation.

This study contributes to basic communication course scholarship by assessing the extent to which speech evaluation training affects students' understanding of universally desired speaking skills. While strong levels of evaluation fidelity reflect the positive state of speech evaluation training at our institution, assessments efforts such as this should still seek out ways of improving even the most objective and successful programs across institutions. We would like to see if this training program can be made available to other institutions with similar results. In this way, we could cross-validate the findings and further establish the relevance of a communication course in general education. We firmly believe that the stronger our assessment efforts are, the more relevant our course becomes.

Author Information

T. Kody Frey (M.A., Illinois State University, 2015) is a lecturer in the School of Information Science at the University of Kentucky.

Cheri J. Simonds (Ph.D. University of Oklahoma, 1995) is a professor and director of communication as critical inquiry in the School of Communication at Illinois State University.

John F. Hooker (Ph.D. Purdue University, 2010) is an associate professor and director of communication as critical inquiry in the School of Communication at Illinois State University.

Kevin R. Meyer (Ph.D. Ohio University, 2009) is an associate professor and graduate coordinator in the School of Communication at Illinois State University.

Stephen K. Hunt (Ph.D. Southern Illinois University, 1998) is a professor and executive director of the School of Communication at Illinois State University.

References

- AAC&U. (2010). Liberal education and America's promise (LEAP). Retrieved from http://www.aacu.org/leap/documents/Introduction_to_LEAP.pdf
- Allen, M., Titsworth, S., & Hunt, S. K. (2009). *Quantitative research in communication*. Los Angeles, CA: Sage.
- Allen, T. H. (2002). Charting a communication pathway: Using assessment to guide curriculum development in a re-vitalized general education plan. *Communication Education, 51*, 26-39. doi:10.1080/03634520216502
- Benander, R., Denton, J., Page, D., & Skinner, C. (2000). Primary trait analysis: Anchoring assessment in the classroom. *The Journal of General Education, 49*, 279-302. doi:10.1353/jge.2000.0025
- Brookhart, S. (2013). How to create and use rubrics for formative assessment. Alexandria, VA: Association for Supervision and Curriculum Development.
- Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English, 18*, 65-81. Retrieved from <http://www.jstor.org/stable/40170979>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46. doi:10.1177/001316446002000104

- Dominowski, R. L. (2002). *Teaching undergraduates*. Mahwah, NJ: Lawrence Erlbaum.
- Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, 2, 155-185.
doi:10.1177/0265532207086780
- Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement*, 5, 105-112.
doi:10.1177/014662168100500115
- Frey, T. K., Hooker, J. F., & Simonds, C. J. (2015). The invaluable nature of speech evaluation training for new basic course instructors. *Basic Communication Course Annual*, 27, 1-7. Retrieved from
<http://ecommons.udayton.edu/bcca/vol27/iss1/6>
- Gardiner, L. F. (1994). Redesigning higher education: Producing dramatic gains in student learning (Report No. 7). Washington, DC: Graduate School of Education and Human Development, George Washington University.
- Goulden, N. R. (1990). A program of rater training for evaluating public speeches combining accuracy and error approaches. *Basic Communication Course Annual*, 2, 143-165. Retrieved from <http://ecommons.udayton.edu/bcca/vol2/iss1/13>
- Harsch, C., & Martin, G. (2013). Comparing holistic and analytic scoring methods: Issues of validity and reliability. *Assessment in Education: Principles, Policy & Practice*, 20, 281-307. doi:10.1080/0969594x.2012.742422
- Hunt, S. K., Simonds, C. J., & Hinchliffe, L. J. (2000). Using student portfolios as authentic assessment. *Journal of Excellence in College Teaching*, 11(1), 57-77.
Retrieved from <http://celt.muohio.edu.libproxy.lib.ilstu.edu/>
- Kaid, L. L., Wadsworth, A. J. (1989) Content analysis. In P. Emmert & L Barker (Eds.), *Measurement of communication behavior* (pp. 197-217). New York, NY: Longman.

- LeFebvre, L., LeFebvre, L. E., Blackburn, K., & Boyd, R. (2015). Student estimates of public speaking competency: The Meaning Extraction Helper and video self-evaluation. *Communication Education, 64*, 261-279.
doi:10.1080/03634523.2015.1014384
- Lumley, T. (2005). *Assessing second language writing: The rater's perspective*. Frankfurt, Germany: Lang.
- Mazer, J. P., Simonds, C. J., & Hunt, S. K. (2013). Assessing evaluation fidelity: An examination of student comments and scores on speech self-evaluation forms in a general education communication course. *Ohio Communication Journal, 51*, 1-28.
Retrieved from www.ohiocomm.org
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.
- Reynolds, D. L., Hunt, S. K., Simonds, C. J., & Cutbirth, C. W. (2004). Written speech feedback in the basic communication course: Are instructors too polite? *Basic Communication Course Annual, 16*, 36-71. Retrieved from <http://ecommons.udayton.edu/bcca>
- Simonds, C. J. (1997). Classroom understanding. An expanded notion of teacher clarity. *Communication Research Reports, 14*, 279-290.
doi:10.1080/08824099709388671
- Simonds, C. J., Meyer, K. R., Hunt, S. K., & Simonds, B. K. (2009). Speech evaluation assessment: An analysis of written speech feedback on instructor evaluation forms in the basic communication course. *Basic Communication Course Annual, 21*, 69-94. Retrieved from <http://ecommons.udayton.edu/bcca/vol21/iss1/8>
- Sprague, J. (1993). Retrieving the research agenda for communication education: Asking the pedagogical questions that are “embarrassments to theory”. *Communication Education, 42*, 106-122. doi:10.1080/03634529309378919

- Stitt, J. K., Simonds, C. J., & Hunt, S. K. (2003). Evaluation fidelity: An examination of criterion-based assessment and rater training in the speech communication classroom. *Communication Studies*, *54*, 341-353. doi:10.1080/10510970309363290
- Tincani, M., & Travers, J. (2017). Publishing single case research design studies that do not demonstrate experimental control. *Remedial and Special Education*, *38*, 1-11. doi:10.1177/0741932517697447
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*, 249-276. doi:10.3102/00346543068003249
- Weigle, S. C. (1999). Investigating rater/prompt interactions in writing assessment: Quantitative and qualitative approaches. *Assessing Writing*, *6*, 145-178. doi:10.1016/s1075-2935(00)00010-6

Appendix A

Speech Evaluation Training for Students

1. Direct students to the Instructor Evaluation Form (included in this packet).
 - a. Point out the categories (Outline, Introduction, Body, Conclusion, Delivery, and Overall Impression).
 - b. Point out the behaviors within the categories (e.g. Attn. Getter, Relevance Statement, Credibility, Thesis, Preview).
 - c. Indicate that each category is evaluated separately.
2. Next, direct students to the Criteria for Evaluating Speeches on page 14-15 of the spiral workbook.
 - a. Point out that these are the qualitative differences between an A, B, C, and D for each of the behaviors within the categories.
 - b. Discuss the themes (or guidelines) used in developing the criteria (included in this packet).
 - c. Discuss the grading scale for each of the categories.
 - d. Discuss Types of Speech Feedback and Using Feedback/Criteria to Determine Score (included in this packet).
3. Show the C Coliseum Speech.
 - a. Ask, does this speech meet the requirements of the assignment?
 - b. Ask, what kinds of constructive comments can you provide this speaker to help her improve her performance?
4. Show the A Coliseum Speech.
 - a. Ask, what were the qualitative differences between the first and second speeches?
 - b. What kinds of positive-descriptive comments would you give this speaker?
 - c. What kinds of constructive comments might you add?
 - d. Ask, is this speech perfect? No, does a speech have to be perfect to get an A? No, but does this speech go above and beyond meeting the requirements of the assignment? Yes!
5. Direct student to the Worksheet for Evaluating Introductions on page 22 of the spiral workbook.
 - a. Have students complete the worksheet and determine the score based on the nature of the feedback provided.
 - b. Have students share their scores.
 - c. Go through worksheet together.
 - d. Determine agreed upon score.

Instructor/Self Evaluation Form: Informative Speech

Name: _____ Topic: _____

OUTLINE AND REFERENCES (10 pts.)

Purpose statement clear
Follows Outline Format
References correct/sufficient _____ pts.

INTRODUCTION (20 pts.)

Gained attention
Showed relevance of topic to audience
Established credibility
Introduced topic/thesis statement clearly
Previewed body of speech _____ pts.

BODY (30 pts.)

Main points clear
Strong evidence & supporting material
Organization effective
Language precise, clear, powerful
Transitions effective
Sources are well integrated, credible, & cited fully _____ pts.

CONCLUSION (10 pts.)

Audience prepared for conclusion
Purpose & main points reviewed
Closed speech by reference to intro./other devices _____ pts.

DELIVERY (15 pts.)

Maintained eye contact
Used voice, diction, & rate for maximum effect
Used space, movement, & gestures for emphasis _____ pts.

OVERALL IMPRESSION (15 pts.)

Topic challenging
Adapted to audience
Maintained time limits
Evidence of preparation & practice
Quality & relevance of visual aids
Was informative _____ pts.

TOTAL POINTS _____ pts

Major Themes in Developing Criteria and Grading Scale

- (D) = Present in outline or speech, but not both.
- (C) = Present—Meets requirement of the assignment.
- (B) = Logically flows, well integrated.
- (A) = Creative, unique, captivating, powerful.

Outline and References (10 Possible Points)

- 9 = A
- 8 = B
- 7 = C
- 6 = D

Introduction (20 Possible Points)

- 18 = A
- 16 = B
- 14 = C
- 12 = D

Body (30 Possible Points)

- 27 = A
- 24 = B
- 21 = C
- 19 = D

Conclusion (10 Possible Points)

- 9 = A
- 8 = B
- 7 = C
- 6 = D

Delivery (15 Possible Points)

- 13.5 = A
- 12 = B
- 10.5 = C
- 9 = D

Overall Impression (15 Possible Points)

- 3.5 = A
- 12 = B
- 10.5 = C
- 9 = D

Types of Speech Feedback

Positive Non-Descriptive: Positive non-descriptive comments say that the student did a good job but do not describe or detail how the task was accomplished. These comments generally identify which behavior is performed well, but lack any specificity. Positive non-descriptive comments will use qualitative language from the *A* or *B* criteria and indicate a skill on the behavior list. When feedback is high inference in nature, it is considered non-descriptive. For example, *effective, funny*.

Examples: Good eye contact
 Clear thesis
 Thorough development
 Excellent visual aids
 Plus marks (+)
 Happy faces (☺)
 Yes
 Very appropriate
 Letter grades (A or B)
 Funny (high inference)
 Effective (high inference)

Positive Descriptive: Positive Descriptive comments are those that say that the student did a good job, and specifically describe or detail what was liked about how the student accomplished their task (going above and beyond what is listed as a skill in the behavior set). Positive Descriptive comments will use qualitative language for the *A* or *B* criteria, identify the behavior or skill, and provide additional specificity that includes mention of a behavior or skill not listed in the behavior set. These comments transcend the requirements of non-descriptive comments and may give students some advice and/or future direction. In other words, positive descriptive comments may indicate repeatable behaviors for continued success. Positive descriptive comments are also low inference in nature. For example, *nice energy and enthusiasm in your closing remarks*.

Examples: Good job of engaging your audience through the use
 of facial expression and direct eye contact.

Nice job of incorporating full source citations into the flow of your presentation.

Your visual aids are very professionally produced and incorporated smoothly into the presentation.

Cool quote to close.

Negative: Negative comments criticize the speech without providing suggestions for improvement. These comments generally identify which behavior is present, lacking, or performed poorly, but lack any specificity (or are high inference in nature). Note to ISU coders: If the quicksheet is used, code *C* and *D* items in this category.

Examples:

- Poor eye contact
- Use APA references
- Only heard two sources
- Conclusion not stated
- Visual aids need work
- Minus marks (-)
- Check marks
- Letter grades (C, D, or F)
- No
- Neutral statements (present or completed, adequate, fine, OK, sufficient, appropriate)
- Be more effective (high inference)

Constructive: Constructive comments acknowledge the need for improvement in the speech and provide specific direction or detail on how to improve (going above and beyond what is listed as a skill in the behavior set). These comments transcend the requirements of negative comments and may give students some advice and/or future direction. In other words, constructive comments make a request of the student or ask him/her to do something different next time. These suggestions are low-inference in nature—that is, you can assume that the student would reasonably know specific behaviors to engage in based on the feedback. For example, *be confident*.

- Examples:
- You need more direct eye contact. Try using fewer note cards and gaze more directly with more of your audience.
 - Try to provide more complete information for each source. I would suggest putting complete information on your note-cards.
 - Your visual aids need to be larger and bolder. Practice incorporating them into the flow of your speech.
 - Read less.
 - Be confident.

Using Feedback/Criteria to Determine Score

“C” Speeches: will meet all of the requirements for the assignment and the criteria for a “C” speech. However, “C” speeches will contain a preponderance of constructive comments. Start by writing negative comments during the presentation and provide elaboration (constructive comments) when completing the evaluation.

“A” Speeches: will exceed the requirements for the assignment, the criteria for an “A” speech, and will contain a preponderance of positive descriptive comments. Start by writing positive comments during the presentation and provide elaboration (positive-descriptive comments) when completing the evaluation.

Notes: Use language from the criteria form to provide elaboration. Examine the relationship between the types of comments provided (constructive/positive-descript) and the score for each graded category (outline, introduction, body, conclusion, deliver, impression).

Appendix B

Code Book for Speech Evaluation

1. **Coder ID:** Refers to the number assigned to each coder. Please be sure to include your coder ID on each code sheet.
2. **Student ID:** Refers to the number assigned to each student data set (instructor and self-evaluation form) and can be found in the upper right corner of each sheet of paper. This number indicates the condition (experimental or control)- and the student identifier. Please be sure to include the dash as you record this number.
3. **Feedback Table Instructor Evaluation:** Records the number of comments (tallies) for each *type of feedback* (details provided on page 3 and 4 of this code book) provided in each *evaluation category* provided by the instructor. Please place a tally mark (/////) for each type of comment within each category.
4. **Feedback Table Student Self Evaluation:** Records the number of comments (tallies) for each *type of feedback* (details provided on page 3 and 4 of this code book) provided in each *evaluation category* provided by the student. Please place a tally mark (/////) for each type of comment within each category. Note that for determining inter-coder reliability, you will provide the corresponding number of the comment rather than the tally.

Tally (unit of analysis): Record each comment into its smallest possible unit. Consider the behavior list when separating comments. When the comment moves from one skill listed in the behavior set to another, you should separate the comments. It may be necessary to divide detailed, combined, or mixed comments into separate units. For example, an instructor may make both a positive non-descriptive and a constructive comment for a given behavior in a given category (ex. Good eye contact, but try looking at more of your audience throughout your speech—or— Attention getter is good, but try to be more creative). The rule that should be applied is when the comment crosses over from the positive categories to the negative categories, you must separate the comments. In some instances, behaviors sets are grouped with one single comment. Code this comment for each behavior set (do not divide individual qualities within the behavior set unless the instructor does so specifically).

Code marks: Sometimes an instructor may assign both a letter grade or code mark (plus or minus, check mark, happy/unhappy face) and provide a comment for

a behavior. You should separate the codes from the comments and provide a tally mark for each. If letter grades are indicated for each behavior, code *As* and *Bs* as positive non-descriptive and *Cs* or *Ds* as negative. All neutral statements (check marks, terms such as OK, adequate, present, sufficient) because they are consistent with *C* quality behaviors should be coded as negative. Please do not code marks that are intended to provide bullet points to separate comments.

Tenses: Sometimes instructors use past or present tense in their feedback. Since we cannot ascribe the instructor's intent, you should assume that past tense comments indicate what the student *did* and present tense comments as what the student *should do*.

Instructor notes: Notations made to flow the speech (e.g., a list of main points, tally marks for number of sources heard, speech time, etc.) and should not be coded.

General feedback: Refers to comments made not specifically directed to a skill or behavior. These are sometimes noted in the margin or at the bottom of the page. If the comment is related to the student's performance (ex. Good job Casey!), it should be coded under overall impression. If the comment is more personal in nature (ex. Good luck, Hope you feel better, Happy Birthday, etc.) it should not be coded.

Missing data: In some instances, data may be missing such as a score for a particular category. All missing data should be coded as 999. If, however, there are no instances of a certain type of comment in a particular category, you should leave that cell in the feedback table blank.

Score: Refers to the number of points the student received for each category. Please record the number of points awarded (score) for that category.

Total grade: Refers to the overall score the students received from the instructor or gave themselves. Please record the number of points awarded (score) for that speech. In cases where points are deducted for any reason, record the original score (without the deduction) as this better represents the student's level of performance. And, just to be sure, please calculate the overall score by adding the scores for each category.

Appendix C

Code Sheet for Instructor and Student Self-Evaluations

1. Coder ID: _____

2. Student ID: _____

(include dash to separate condition from student)

3. Feedback Table Instructor Evaluation:

	Outline	Intro	Body	Concl	Delivery	Overall
PosND						
PosDesc						
Negative						
Constructive						
Score						
Total Grade						

4. Feedback Table Student Self-Evaluation:

	Outline	Intro	Body	Concl	Delivery	Overall
PosND						
PosDesc						
Negative						
Constructive						
Score						
Total Grade						