



Received: 26 October 2016
Accepted: 23 January 2017
Published: 16 February 2017

*Corresponding author: Gina Iberri-Shea, Department of English and Fine Arts, United States Air Force Academy, 2354 Fairchild Drive Ste 6D-177, CO 80840, USA
E-mail: gina.iberri-shea@usafa.edu

Reviewing editor:
Sammy King Fai Hui, The Education University of Hong Kong, Hong Kong

Additional information is available at the end of the article

EDUCATIONAL ASSESSMENT & EVALUATION | RESEARCH ARTICLE

Adaptation and assessment of a public speaking rating scale

Gina Iberri-Shea^{1*}

Abstract: Prominent spoken language assessments such as the Oral Proficiency Interview and the Test of Spoken English have been primarily concerned with speaking ability as it relates to conversation. This paper looks at an additional aspect of spoken language ability, namely public speaking. This study used an adapted form of a public speaking rating scale originally designed for English as a foreign language (EFL) contexts. This paper seeks to evaluate the relationship between this EFL-based scale and assessment within a core US university public speaking course. The relevance of EFL assessments to those used in English medium courses provides insight as to whether language learners are being evaluated on similar public speaking constructs to their English speaking peers, and informs instruction aimed at preparing students for English medium courses. A sample of undergraduate students ($N = 44$), primarily native speakers of English, performed classroom speeches in an introductory public speaking course and were rated using the adapted public speaking rating scale. The rating scores, independent instructor evaluations, and written feedback from two raters were analyzed to determine reliability and validity, and to inform revisions of the scale. The rating scale was found to be reliable and valid for this population. Limitations of the scale and proposed expansions are discussed as the result of qualitative analysis of the data.

Subjects: Communication Studies; Education; Language & Literature

Keywords: speech rating; spoken language assessment; classroom speaking

Public speaking has become an integral part of academic life for many students. For the purposes of this paper, the description of public speaking will be limited to university student public speech that occurs in the classroom. Several undergraduate programs in the United States now include public speaking in their liberal studies or major requirements, and an interest in English public speaking has grown abroad, for example in Japan, where secondary school guidelines recommend public speaking and debate in the English as a foreign language curriculum (Carter, Goold, & Madeley, 1993). In addition, there are a growing number of post-secondary institutions in Japan that have added public

ABOUT THE AUTHOR

Gina Iberri-Shea is an associate professor of English and director of forensics at the United States Air Force Academy. Her research interests include university student public speech, debate, and language variation in academic contexts.

PUBLIC INTEREST STATEMENT

This study looks at the assessment of public speaking in mainstream US university classrooms and compares it to evaluations of public speaking in English as a foreign language contexts. This is important in determining how public speaking instruction in foreign language contexts relates to and prepares students for English medium university coursework and assessment.

speaking and debate to their language departments. According to Yamashiro (2002), many students from these courses have reported that the skills of public speaking are crucial to success in college both in Japan and North America. The requirement of extended speaking in English as a foreign language (EFL) and English as a second language (ESL) educational contexts has increased with the acceptance of communicative approaches to pedagogy (Nakamura, 1995; Turner, 2000). This suggests that investigations of the current expectations and assessments of public speaking in English speaking universities may help to inform public speaking pedagogy in EFL/ESL contexts. A number of reports have examined public speaking in EFL contexts (Izumi, 1995; Russell & Loschky, 1998; Stroud, 2001; Yamashiro, 1999; Yamashiro & Johnson, 1997), but there is little research on how the pedagogy and assessment practices in EFL/ESL speaking contexts translate to the expectations and evaluations used in university public speaking classrooms.

1. Background

The ACTFL Oral Proficiency Interview (OPI) and the TOEFL Test of Spoken English (TSE) are used broadly in academia and professional disciplines to assess foreign language speaking and EFL speaking, respectively. Public speaking is one specific aspect of academic speaking ability that is neglected by such general speaking tests as the OPI and TSE. Although demands for public speaking are often placed on language learners in academic contexts (as in communication, business, and liberal studies classes), there is not yet enough research on this type of task and appropriate measures of assessment for language learners. Because public speaking is required in many academic contexts, an assessment that reflects the public speaking expectations of mainstream classrooms would benefit these students.

Public speaking is a complex task, requiring an individual to analyze, plan, and produce language in a performance context. According to Turner and Upshur (2002), performance tasks which require extended speaking are among the greatest improvements in second language assessment. Rather than maintaining a strict focus on accuracy, performance assessments allow the evaluator to measure the actual use of language in context. Performance assessment refers to both the assessment task (public speech) and the scoring method used (Khattari, Reeve, & Kane, 1998). This type of assessment has gained increasing support in spoken language testing as researchers strive to increase authenticity and transparency in testing.

The development of an EFL public speaking course and associated rating scale was an important first step in creating an assessment that addresses public speaking specifically. Yamashiro (2002) investigated the connection between second language study and public speaking, and tested the reliability and validity of a rating scale to be used in EFL contexts. The study emphasized assessment measures that correspond to the expectations and methods of a particular academic context. The public speaking rating scale was developed by Yamashiro and Johnson (1997) as part of a curriculum project, and was used to rate the classroom speeches of Japanese university students ($N = 170$) from seven classes attending two different universities. Reliability and validity were assessed using structural equation modeling, analyzing three traits (non-verbal, verbal, and purpose) and three rating methods (teacher, peer, and self) with a 14-point public speaking rating scale. Yamashiro found the scale to be reliable and valid in this context. These findings are an important step in meeting the demands of EFL public speaking instruction and providing a bridge between assessment and language teaching by validating a test for a specific classroom purpose. It is important however, to research ways of improving the accuracy of assessment for specific tasks and evaluating the usefulness of L2 assessments. Yamashiro's study does not provide us with a connection between EFL assessment in public speaking and the expectations of mainstream public speaking courses within the United States. This is an important question for many EFL/ESL students planning to transition to credit-bearing university courses.

In order to address this question, the scale used by Yamashiro (2002) has been adapted for the current investigation to examine the relationship between an assessment designed for second/foreign language speaking classrooms and evaluation used within mainstream public speaking

classrooms. The purpose of this study is to (a) explore the constructs underlying the assessment of public speaking and (b) evaluate the extent to which Yamashiro's adapted public speaking scale adequately measures the language abilities emphasized in a university public speaking context. Specifically, this study addressed the following research questions:

- (1) What is the relationship between scores on the adapted public speaking rating scale and grades assigned by a public speaking instructor?
- (2) To what extent does the adapted rating scale address the concerns of a public speaking instructor and experienced raters of public speaking?

2. Method

2.1. Participants

Participants in this study were undergraduate students enrolled in a university in the southwestern United States. There were 44 students from two intact public speaking classes. The two classes were taught by the same teacher, a professor in the Communication Department with over 40 years of teaching experience. The classes met twice a week for 15 weeks during a fall term. There were 14 males and 30 females ranging in age from 18 to 44 with a mean age of 21 and a standard deviation of 4.7 years. Sixteen of the students were communication majors, with the remainder from a variety of other majors (English, Liberal Studies, Math, Parks and Recreation, Social Work, Latin American Studies, Hotel and Restaurant Management, Psychology, Sociology, History, Construction Management, and Education). Two of the participants spoke languages other than English as their first language (Spanish and German), and 11 students had studied other languages in addition to English (Spanish, French, Latin, Japanese, German, and American Sign Language). There were 22 students in each of the two classes. The majority of students reported that they were taking the class to fulfill a requirement ($n = 39$) and most had little or no speech experience ($n = 37$). There were two raters, both graduate students in an English program at the same university. Both of the raters had 3 years of experience in teaching and evaluating student public speaking at the university level.

2.2. Instruments

The prompt used for this task was taken from the course textbook (Osborn & Osborn, 2002, p. 378) and consisted of an outline of major elements expected in the speaking assignment. It was reviewed during regular class time as part of a 3-week instructional unit. Students also received a copy of the instructor's rubric (see Appendix A), which contained a breakdown of presentation elements required at percentage grade intervals and space for instructor comments. This feedback sheet was developed by the instructor and had been used in previous semesters of teaching public speaking. The rating scale used in this study was adapted from the 14-point EFL public speaking rating scale used by Yamashiro (2002). The original 14-point scale was altered for this study by removing three points addressing non-verbal behavior. The rationale for this decision was that the focus of the current study was on verbal presentation skills only, and that raters would not be able to address non-verbal behavior because they based their evaluations on audio-recorded speeches.

The 11-point public speaking rating scale was previously piloted with 15 pre-recorded speeches to assess inter-rater reliability. Reliability refers to the consistency of measurement (Bachman & Palmer, 1996) and is here determined by the relationship between the scores assigned by two raters on the same speech samples. Inter-rater reliability was calculated using Cronbach's α (Bailey, 1998) on each point of the rating scale. Raters participated in a short training session in order to familiarize themselves with the rating scale and optimize consistent usage. This session included reading over and discussing the rating scale descriptors to ensure that the categories made sense and that the descriptors were relevant and transparent. Audio recordings of three practice speeches were then rated and any differences were discussed, resulting in small changes to the wording of the rating scale descriptors (see Appendix B for rating scale and final descriptors). Each rater then evaluated

Table 1. Inter-rater reliability for pilot sample using Cronbach's α

Rating scale	Inter-rater reliability
Point 1 Projection	.98
Point 2 Pace	.92
Point 3 Intonation	.87
Point 4 Diction	.92
Point 5 Introduction	.86
Point 6 Body	.84
Point 7 Conclusion	.92
Point 8 Topic Choice	.93
Point 9 Language Use	1.0
Point 10 Vocabulary	.97
Point 11 Purpose	.98
Total score	.93

the 15 pilot speeches independently. Analysis of the scores shows an overall inter-rater reliability of $\alpha = .93$ (see Table 1), showing that the public speaking rating scale was used reliably in this public speaking context.

2.3. Data collection

2.3.1. The speaking task

Data was collected from two sections of an introductory public speaking class. The researcher attended each class informally 4 times prior to the study so that the additional presence would not cause unnecessary anxiety. Study procedures and questionnaires were approved by the Institutional Review Board through an expedited review, and students filled out demographic questionnaires and consent forms explaining the general purpose of the study. The students had performed two speeches in their class earlier in the semester and were accustomed to the speeches being videotaped for self-evaluation. For this reason, the use of a recording device was not considered invasive and none of the students appeared to be concerned or inhibited by its presence. The students delivered one 4–6 min persuasive speech on a topic of their choice. Each speech considered in this analysis was audio-recorded during performance assessments as scheduled by the instructor in the class syllabus. Data were collected over 6 class periods.

2.3.2. Instructor and rater evaluations

The researcher audio-recorded the student speeches and collected a copy of the instructor's graded evaluations and written comments for each presentation. The two raters used the 11-point public speaking rating scale to score the recorded speeches. In addition to the rating scale scores for each speech, comments were collected from the two raters to inform issues of validity. Raters were asked to provide written feedback on the speech that they thought would enhance the evaluation, note any problems or limitations encountered while using the scale for this population, and comment on any salient features that helped distinguish ability level among the speech performances.

The instructor and one rater also participated in a workshop replicating scale development methods used in North and Schneider (1998). In this workshop the instructor and rater were asked to listen to two of the pilot speeches and discuss which was better and why. The goal of this session was to identify language used to describe aspects of public speaking, in order to compare these descriptors to the rating scale categories. The discussions were transcribed as notes by the researcher. Rating sheets, feedback, and any additional notes made by the raters and the researcher were gathered throughout the assessment process.

2.4. Data analysis

2.4.1. Correlational analysis

Each of the 44 audio-recorded speech samples were rated using the adapted public speaking rating scale. Validity corresponds to the degree that the inferences made from the assessment scores reflect the constructs or skill of interest, in this case public speaking proficiency. In order to assess the validity of the rating scale, the rating for each speech was compared with the instructor's percentage grade evaluation of each performance. The 5-point scale was treated as interval data and a Pearson correlation was used.

2.4.2. Qualitative analysis

A qualitative review of the rater comments, teacher feedback, and research notes was conducted in order to establish an empirical (data-driven) revision of the rating scale for future use. Each of the comments on the instructor evaluation sheets, the written notes made by each rater, and the research notes from the instructor-rater workshop, were compared to the rating scale descriptors to discover aspects of public speaking not considered in the rating scale. In order to systematically evaluate the feedback, comments made by the raters reflecting problems with scale use were separated from those directed at the speech performance, and feedback on non-verbal behavior was removed from the analysis. Next, each instructor and rater comment was independently compared to the 11-point public speaking rating scale descriptors in order to determine whether the comment was captured by one of the descriptors, or not addressed in the 11-point scale. Those comments that were addressed in the rating scale were coded by descriptor heading. A list was generated of all comments unrelated to the scale descriptors. Following scale-making procedures similar to those used in previous studies of rating scale development for specific tasks (Turner, 2000; Turner & Upshur, 2002; Upshur & Turner, 1999), these comments were placed on cards and raters were asked to place the cards in piles under categories identified in the communication literature (see Appendix C), as well as an additional pile for comments that did not fit into one of the categories. This sorting process elicited five prominent categories, not previously included in the scale: credibility, distracting language, effective vocabulary use, text organization, and audience adaptation. Conversational presentation received only one comment and was not included in further analysis. The rater comments that reflected problems with scale use, previously set aside, were then discussed by the two raters who sorted eight out of nine of these comments into one of the five categories. One comment, addressing memorization, did not fit into these categories. Based on these instructor and rater comments, revisions to the adapted public speaking rating scale were proposed.

3. Results

3.1. Research question 1: What is the relationship between scores on the adapted public speaking rating scale and grades assigned by a public speaking instructor?

The initial rating scale score obtained for each speech sample was correlated with the percentage grades assigned by the instructor ($r = .77$, $n = 44$, $p < .01$) in order to address research question 1 (see Table 2 for descriptive statistics). This correlation coefficient indicates an agreement between the instrument and the classroom assessment measure (Linn & Gronlund, 2000). The instrument can be considered valid in this context. The results indicate that the two forms of assessment are measuring similar constructs.

3.2. Research question 2: To what extent does the adapted rating scale address the concerns of a public speaking instructor and experienced raters of public speaking?

In order to address research question 2, the researcher compared the written comments provided by the instructor and the feedback from the raters with the rating scale and descriptors, sorting comments that were covered in the rating scale from those that were not (see Table 3).

The feedback from both the instructor and the raters was based on the student samples, and can be considered an empirically based contribution to the revision of the rating scale. There were a

Table 2. Descriptive statistics for ratings and instructor grades

	N	Minimum	Maximum	Mean	SD
Ratings	44	35	54	44.79	5.20
Grades	44	60	98	79.68	7.53

Table 3. Summary table of written comments

	Instructor	Raters	Researcher^a	Total
Speech comments related to rating scale descriptors	80	3	10	93
Speech comments not addressed by rating scale	68	28	8	104
Total comments	148	31	18	197

^aResearcher column refers to transcription notes on instructor-rater workshop.

Table 4. Summary of teacher and rater comments not addressed by rating scale

Category	Definition	Examples	Previous literature	Number of comments
Credibility	Information comes from a reliable source, speaker is well informed	You can enhance your credibility by citing where the information came from; good citing of facts	Burrell and Koper (1998) and Perloff (2003)	37
Distracting language	Use of filler words, nonfluencies, awkward pauses	Watch filler words and nonfluencies; nonfluencies detract from the effectiveness of the speech	Burgoon et al. (1990)	24
Vocabulary use	Effective word choice	Could have worked on the wording to make it more effective; vivid imagery	Chimombo and Roseberry (1998), Gass and Seiter, (1999), and Sopory and Dillard (2002)	19
Text organization	Overall organization of speech	Speech structure needs to be clearer; clear organization pattern	O'Keefe (1997) and Allen (1991)	16
Adaptation	All aspects of speech (topic, language use, organization, delivery) are adjusted for intended audience	Establish common ground with your audience; consider your audience when structuring your arguments	Bell (1984) and Ladegaard (1995)	6

number of comments found in the qualitative evaluations of the speech samples that were inadequately represented in the rating scale. These comments were sorted, eliciting five prominent categories (see Table 4), which encompassed 102/104 (98%) of the total speech comments that were not sufficiently reflected in the rating scale, as indicated by the teacher and rater feedback.

Feedback related to credibility (36%) included authority to speak on a chosen topic (i.e. a bilingual student speaking on bilingual education) and the quality and quantity of sources or evidence to support claims. The use of distracting language (hedges, tag questions, vocalized pauses) was not included anywhere on the rating scale and was mentioned by both raters and the instructor (23.5%) as having a distracting or detrimental influence on the overall effectiveness of a presentation. Feedback from the instructor and raters included reference to vocabulary choice, such as vivid language, rather than appropriate language use (18.6%); these types of language use could not be summarized appropriately in the existing vocabulary or language use scales. Textual organization (15.7%) was discussed as going beyond the description of points and examples to support a thesis (under “Content” in the rating scale descriptors) to “using an effective organizational format,” “presenting the information in a logical way,” and “providing continuity throughout the speech.” Raters found it difficult to evaluate audience adaptation (5.9%) because they viewed it as an independent construct and it was defined as part of several other skills in the rating scale. Non-verbal behavior

was intentionally removed from the rating scale in order to focus on verbal behavior, but one rater noted that memorization (a point under “Non-verbal behavior”) was closely tied to fluency and that lack of memorization was clear in the audio tapes. This remark was supported by instructor comments on non-verbal behavior removed from the analysis. A partial list of the feedback placed under each category is found in Appendix D.

4. Discussion

The insights gained through the instructor and rater feedback were crucial to the understanding of an appropriate measure of public speaking. Although the adapted 11-point public speaking rating scale was positively correlated with instructor grades, and could be considered a valid measure of public speaking in this university context, the final analysis revealed some limitations. Instructor and rater feedback indicated specific expansions to the rating scale and descriptors that would more accurately capture public speaking proficiency. These suggestions included: a measure of credibility, delineation between vocabulary choice and accuracy, a measure of distracting language, textual organization, and the addition of audience adaptation as a separate construct. This brings into focus the question of how characteristics of a rating scale should be identified. Several authors have supported data-based criteria for rating scale development (North & Schneider, 1998; Turner, 2000), and the present analysis also supports these findings. Although there was a great deal of overlap between the adapted public speaking rating scale, originally designed for an EFL context, and the instructor’s evaluation sheet, the comparison revealed a number of areas missing or in need of clarification in the scale. This is particularly relevant because it illuminates a gap between the foreign language assessment measure and mainstream public speaking classroom expectations. The feedback from the raters provided suggestions for improvement of the public speaking rating scale.

Despite their absence in the adapted EFL public speaking rating scale, many of the proposed expansions are supported in the communication studies literature. Personal and source credibility is discussed with respect to speaking in both public and interpersonal contexts (Burrell & Koper, 1998; McCroskey & Mehrley, 1969; Montgomery, 2001; O’Keefe, 2002; Perloff, 2003), and was perceived by the instructor and raters as a discriminating factor of public speaking ability in this study. Personal credibility is vaguely addressed under topic choice, but should be made more explicit in the future, and the inclusion of quality as a descriptor of source information would improve the evaluation of content. The adapted public speaking rating scale utilizes a positive format for evaluation, rating the presence of desirable attributes highly and their absence poorly. This appears to neglect the importance of distracting or negative language behavior, such as nonfluencies. It is interesting to note that several researchers (Burgoon, Birk, & Pfau, 1990; McCroskey & Mehrley, 1969) have connected nonfluencies in delivery to a perceived lack of credibility or trustworthiness in speaking. The instructor and rater comments indicate that a measure of distracting language would be beneficial to the revision of the scale, as the presence of this type of language was believed to be detrimental to public speaking ability. This could be achieved through a reversal of the usual rating scale points so that a 5 would indicate the absence of nonfluencies or other distracting speech, and a 1 would indicate that the presence of distracting speech made the presentation difficult to understand or follow. Distracting speech was a salient factor for raters when listening to the speech samples, as indicated by the frequency of comments.

Teacher and rater comments also suggest that definitions of vocabulary use need to be redefined to include not only appropriate word choice, but also engaging word choice. The present scale asks raters to assess vocabulary in relation to the appropriate word choice and clarity for the audience level. An important aspect of public speaking, as revealed by the instructor and rater comments, is the use of vivid language, and the use of intense language has been proposed as a crucial element in advanced language use by a number of researchers (Chimombo & Roseberry, 1998; Hamilton & Hunter, 1999; Sandell, 1977). This suggests delineation between appropriate vocabulary and optimally effective language choice. Gass and Seiter (1999) introduce specific lexical choices that fit into categories described as *God terms*, *Devil terms*, and *charismatic terms*. *God terms* describe words that a group or culture look up to (i.e. *family values*). This has been used in a number of political

campaigns including George W. Bush's election campaign. *Devil* terms refer to words that the audience almost universally looks down upon. *Nazi* is an example of a term that usually fits this category. Charismatic terms are words that strike a positive affective chord in the audience. The word *freedom* has come to be known as one of these terms, at least in the western world. These terms tend to have a positive effect on the audience and increase the feeling of inclusion or common ground. An improvement to the understanding of vocabulary assessment in the rating scale would include an accessible description of both appropriate and effective lexical choice.

Another suggested expansion to the rating scale is the inclusion of textual organization in the description of content. The structure and order of arguments in public speaking was noted in the raters' feedback and is also discussed in the communication literature (O'Keefe, 1997; Struckman-Johnson & Struckman-Johnson, 1996). The organization of arguments within a speech is vital to its overall effectiveness. For example, Allen (1991) compared the persuasiveness of one and two sided arguments using meta-analysis, and found that two sided arguments that mentioned the opposing positions and refuted them were more persuasive than either one sided arguments or two sided arguments that simply stated the opposing view. These studies show that the perceived success of public speech is highly reliant on the types of arguments provided and their organizational structure, as well as the linguistic choices made during speech. The adapted rating scale describes the presence of points or arguments to support a thesis but makes no mention of their order of presentation. Organization was evaluated explicitly by the instructor and commented on by the raters. Although it was not included in the adapted scale, raters questioned whether or not an evaluation of organization was inherently part of their rating of content. A revision of the scale should include a specific statement assessing effective organization in the speech.

Finally, audience adaptation has been considered a critical aspect of public speaking. Topic choice, language use, and delivery are all elements of public speaking that rely heavily on audience-related factors. The ability to adapt or style shift in a particular context is crucial to successful public speaking (Bell, 1984; Ladegaard, 1995). Audience adaptation was recognized and included in the original rating scale, described within projection, pace, diction, topic choice, and vocabulary, but feedback from the raters indicated that the inclusion of audience adaptation across the descriptors was confusing and overly general. Adaptation as an independent measure would clarify and improve the assessment of public speaking.

Taken together, these five areas describe the gaps found between the public speaking rating scale, originally designed for use in an EFL context, and written comments provided by a public speaking instructor and two trained raters. This process of quantitative and qualitative assessment of a public speaking rating scale was found to be valuable to the understanding of elements that contribute to public speaking success. Although the rating scale was shown to be reliable and relatively valid in both the present study, using primarily native speakers, and an EFL population (Yamashiro, 2002), the insights gained through instructor and rater feedback in the present study were constructive and contribute to an increased understanding of appropriate measures of public speaking in academic contexts.

5. Conclusions

The context specific nature of classroom-based research in the present study does not allow us to make generalized claims regarding foreign language speech pedagogy. The definitive relationship between public speaking instruction in first and second language classrooms remains to be seen. Instead, this research hopes to broaden the understanding of public speaking assessment as it is applied in context, and work toward a more comprehensive evaluation of student public speech. This paper contributes to knowledge of the relationship between a rating scale developed for the foreign language classroom, and evaluations of public speaking in a mainstream university classroom. Investigations of the degree to which foreign language public speaking assessments measure the same constructs as mainstream university evaluations of public speech play an important role in our understanding of how best to prepare international students for credit-bearing university classes, as

many courses across the curriculum require some degree of public speaking. An informal survey of instructors from various departments at the host university showed that 88% of respondents ($n = 25$) describe using oral communication activities, such as presentations in their classes, and the general catalog showed that 27% of undergraduate degree requirements include at least one course which emphasizes speaking. Because the most commonly used assessments of oral language proficiency, such as the OPI and TSE, focus attention on conversation and interview skills, more assessments which emphasize public speaking skills are needed for academic purposes. Public speaking is a fundamental part of academic life in the United States, and international students should be prepared for this type of task. This study takes an important step in understanding the public speaking expectations of students in mainstream university classrooms. Future research should include assessment of revisions to the public speaking rating scale using performance samples. It would be useful to pilot the rating scale in both mainstream and second/foreign language classrooms and study any differences or limitations that arise during implementation. As revisions are made, the scale should be reassessed for reliability and validity in the specified contexts. The development of adequate assessment measures for speaking ability in university contexts continues to be an ongoing process.

Funding

The author received no direct funding for this research.

Author details

Gina Iberri-Shea¹

E-mail: gina.iberri-shea@usafa.edu

¹ Department of English and Fine Arts, United States Air Force Academy, 2354 Fairchild Drive Ste 6D-177, CO 80840, USA.

Citation information

Cite this article as: Adaptation and assessment of a public speaking rating scale, Gina Iberri-Shea, *Cogent Education* (2017), 4: 1287390.

Note

A portion of this research was conducted at Northern Arizona University.

References

- Allen, M. (1991). Meta-analysis comparing the persuasiveness of one-sided and two-sided messages. *Western Journal of Speech Communication*, 55, 390–404.
<http://dx.doi.org/10.1080/105703191109374395>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bailey, K. M. (1998). *Learning about language assessment*. New York, NY: Heinle & Heinle.
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13, 145–204.
<http://dx.doi.org/10.1017/S004740450001037X>
- Burgoon, J. K., Birk, T., & Pfau, M. (1990). Nonverbal behaviors, persuasion, and credibility. *Human Communication Research*, 17, 140–169.
<http://dx.doi.org/10.1111/hcre.1990.17.issue-1>
- Burrell, N. A., & Koper, R. J. (1998). The efficacy of powerful/powerless language on attitudes and source credibility. In M. Allen & R. W. Preiss (Eds.), *Persuasion: Advances through meta-analysis* (pp. 203–215). Cresskill, NJ: Hampton Press.
- Carter, N., Gool, R., & Madeley, C. (1993). The new Mombusho guidelines, part two. *The Language Teacher*, 17, 3–7.
- Chimombo, M. P. F., & Roseberry, R. L. (1998). *The power of discourse*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gass, R., & Seiter, J. (1999). *Persuasion, social influence, and compliance gaining*. Needham Heights, MA: Allyn & Bacon.
- Hamilton, M. A., & Hunter, J. E. (1999). The effect of language intensity on receiver evaluations of message, source, and topic. In M. Allen & R. W. Preiss (Eds.), *Persuasion: Advances through meta-analysis* (pp. 99–138). Cresskill, NJ: Hampton Press.
- Izumi, K. (1995). Teaching debate in Japanese high schools. *The Language Teacher*, 19, 9–11.
- Khattri, N., Reeve, A. L., & Kane, M. B. (1998). *Principles and practices of performance assessment*. Mahwah, NJ: Lawrence Erlbaum.
- Ladegaard, H. J. (1995). Audience design revisited: Persons, roles and power relations in speech interactions. *Language & Communication*, 15, 89–101.
[http://dx.doi.org/10.1016/0271-5309\(94\)00017-7](http://dx.doi.org/10.1016/0271-5309(94)00017-7)
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Prentice-Hall.
- McCroskey, J. C., & Mehrley, R. S. (1969). The effects of disorganization and nonfluency on attitude change and source credibility. *Speech Monographs*, 36, 13–21.
<http://dx.doi.org/10.1080/03637756909375604>
- Montgomery, M. (2001). The uses of authenticity: “Speaking from experience” in a U.K. election broadcast. *The Communication Review*, 4, 447–462.
- Nakamura, Y. (1995). Making speaking tests valid: Practical considerations in a classroom setting. In J. D. Brown & S. O. Yamashita (Eds.), *Language testing in Japan* (pp. 136–148). Tokyo: Japan Association for Language Teaching.
- North, B., & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing*, 15, 217–263. <http://dx.doi.org/10.1191/026553298676177132>
- O’Keefe, D. J. (1997). Standpoint explicitness and persuasive effect: A meta-analytic review of the effects of varying conclusion articulation in persuasive messages. *Argumentation and Advocacy*, 34, 1–12.
- O’Keefe, D. J. (2002). *Persuasion: Theory and research* (2nd ed.). Thousand Oaks, CA: Sage.
- Osborn, M., & Osborn, S. (2002). *Public speaking* (5th ed.). Boston, MA: Houghton Mifflin Company.
- Perloff, R. M. (2003). *The dynamics of persuasion*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Russell, G., & Loschky, L. (1998). The need to teach communication strategies in the foreign language classroom. *JALT Journal*, 20, 100–114.
- Sandell, R. (1977). *Linguistic style and persuasion*. New York, NY: Academic Press.
- Sopory, P., & Dillard, J. P. (2002). Meta-analysis of the empirical research on metaphor and persuasion. In J. P. Dillard & M. Pfau (Eds.), *Figurative language and persuasion* (pp.

407–426). Thousand Oaks, CA: Sage.

Stroud, S. (2001). *Teaching public speaking: Insights from the eastern world*. Paper presented at the Annual Meeting of the National Communication Association, Atlanta, GA.

Struckman-Johnson, D., & Struckman-Johnson, C. (1996). Can you say condom?: It makes a difference in fear-arousing AIDS prevention public service announcements. *Journal of Applied Social Psychology*, 26, 1068–1083.
<http://dx.doi.org/10.1111/jasp.1996.26.issue-12>

Turner, C. E. (2000). Listening to the voices of rating scale developers: Identifying salient features for second language performance assessment. *Canadian Modern Language Review*, 56, 556–586.

Turner, C. E., & Upshur, J. A. (2002). Rating scales derived from student samples: Effects of the scale maker and the student sample on scale content and student scores. *TESOL Quarterly*, 36, 49–70.
<http://dx.doi.org/10.2307/3588360>

Upshur, J. A., & Turner, C. E. (1999). Systematic effects in the rating of second-language speaking ability: Test method and learner discourse. *Language Testing*, 16, 82–111.

Yamashiro, A. D. (1999). Integrating American speech communication research into EFL. *Speech Communication Education*, 12, 111–131.

Yamashiro, A. D. (2002). *Using structural equation modeling for construct validation of an English as a foreign language public speaking rating scale* (Unpublished Doctoral Dissertation). Tokyo: Temple University.

Yamashiro, A. D., & Johnson, J. (1997). Public speaking in EFL: Elements for course design. *The Language Teacher*, 21, 13–17.

Appendix A

Instructor rubric and feedback sheet

Course Title

Average (C: 70–80 points)

Name:

The average speech

Assignment:

conforms to speech type

Speech Grade:

is ready on assigned day

meets time limits

provides a clear purpose/thesis

Comments:

presents an identifiable introduction, body, conclusion

is delivered competently and extemporaneously

is free of major errors in grammar, pronunciation, language

Above Average (B: 80–90)

The above average speech meets the previous criteria and

considers an appropriate topic

fulfills major functions of the introduction/conclusion

provides clear organization of main points/supporting materials

supports main points with adequate evidence-cited sources

is delivered skillfully enough not to detract

Superior (A: 90–100)

The superior speech meets the previous criteria and

gains attention, feeling, commitment from audience

uses vivid language

is delivered in a manner to strengthen message impact

Appendix B

The 11-point public speaking rating scale and descriptors (adapted from Yamashiro, 2002)

Rater: _____ Class _____

Student # _____ Date of recording _____

Topic/Title of Speech _____

Use the following five-point scale with accompanied descriptors.

Voice Control

Projection	5	4	3	2	1
Pace	5	4	3	2	1
Intonation	5	4	3	2	1
Diction	5	4	3	2	1

Content

Introduction	5	4	3	2	1
Body	5	4	3	2	1
Conclusion	5	4	3	2	1

Effectiveness

Topic Choice	5	4	3	2	1
Language Use	5	4	3	2	1
Vocabulary	5	4	3	2	1
Purpose	5	4	3	2	1

Total: ____/55_

Descriptors for 11-Point Public Speaking Rating Scale

Voice Control

Point 1 Projection

Speaking not too loudly or too softly. Should be loud enough for all members in the audience to hear clearly. Should be a little louder than normal conversational voice. Projection should be varied at times of emphasis.

5	Speaker speaks loud enough so that all members of the audience can easily hear. Projection was sometimes louder or softer to emphasize important points.
4	Speaker speaks loud enough, but was too loud or too soft a few times during the presentation. Speaker tried to use variation in projection for emphasis, but was awkward.
3	Speaker generally speaks loud enough, but sometimes may be too loud or too soft for easy listening. No variation in projection to emphasize important points.
2	Speaker generally did not speak loud enough for easy listening. However, it was possible with effort to hear the points being made. No variation in projection for emphasis.
1	Speaker could not be understood without great effort by all members of the audience.

Point 2 Pace

Speaking not too quickly or too slowly. Should be at a good speed for people to hear all points made. Speed should be varied at times of emphasis.

5	Speaker speaks at a good pace, so that all members of the audience can easily hear and comprehend. Pace was sometimes faster or slower to emphasize important points.
4	Speaker uses a good pace, but was too fast or too slow a few times during the presentation. Speaker tried to use variation in pace for emphasis, but was awkward.
3	Speaker generally speaks at a good pace, but sometimes may be too fast or too slow for easy listening. No variation in pace to emphasize important points.
2	Speaker generally did not speak at a good pace for easy listening. However, it was possible with effort to hear the points being made. No variation in pace for emphasis.
1	Speaker could not be understood without great effort by all members of the audience.

Point 3 Intonation

Speaking smoothly in phrases, so that ideas are clear. Intonation is very important for being a good speaker. Pitch patterns help clarify the type of information and its importance. Use pauses to make each point clear.

5	Speaker speaks smoothly so that all members of the audience can easily hear. Pitch patterns and pauses were used effectively to emphasize important points.
4	Speaker speaks smoothly enough, but had a few awkward pauses during the presentation. Speaker tried to use variation in pitch to emphasize important points, but was awkward.
3	Speaker generally speaks smoothly enough but sometimes may have awkward or long pauses. No variation in pitch or use of pauses to emphasize important points.
2	Speaker generally did not speak smoothly enough for easy listening. It was possible with effort to hear the points being made. No variation in pitch for emphasis.
1	Speaker could not be understood without great effort by all members of the audience.

Point 4 Diction

All words/phrases should be clearly spoken so that the audience can easily hear/understand the speaker's points.

5	Speaker speaks clearly so that all members of the audience can easily hear. All key words were clear. Accent and syllable stress were used effectively during the presentation.
4	4 Speaker speaks smoothly enough, but had a few key words that were unclear. Speaker tried to use accent and syllable stress, but was awkward.
3	Speaker generally speaks clearly enough but sometimes may have accent or syllable stress. Overall meaning can be understood, but some key words were unclear.
2	Speaker generally did not speak clearly enough for easy listening. It was possible with effort to hear the points being made. Many key words were not understood.
1	Speaker could not be understood without great effort by all members of the audience.

Content of Oral Presentation

Point 5 Introduction

Speakers should have an attention-getting device, thesis statement, and a sentence of method.

5	Introduction has the following three parts: catches audience's attention, states the main idea, and explains how the information will be presented. All three parts are effective.
4	Introduction has the following three parts: catches audience's attention, states the main idea, and explains how the information will be presented. One part is weak.
3	Introduction may be missing one of the three parts: catches audience's attention, states the main idea, and explains how the information will be presented. Some parts are weak.
2	Introduction has at least one of the three parts: catches audience's attention, states the main idea, and explains how the information will be presented.
1	Introduction is not complete or effective.

Point 6 Body

Speakers should use transition words and have supporting ideas related to the thesis statement.

5	Body has the following three parts: transition words are used, supporting ideas related to the thesis, enough facts or examples are used to develop each idea. All three parts are effective.
4	Body has the following three parts: transition words are used, supporting ideas related to the thesis, enough facts or examples are used to develop each idea. One part is weak.
3	Body may be missing one of the three parts: transition words are used, supporting ideas related to the thesis, enough facts or examples are used to develop each idea. Some parts are weak.
2	Body has at least one of the three parts: transition words are used, supporting ideas related to the thesis, enough facts or examples are used to develop each idea.
1	Body is not complete or effective.

Point 7 Conclusion

Speakers should restate the thesis statement/summarize body and make a closing statement.

5	Conclusion has the following two parts: restatement of the thesis or a summary of the body and a closing statement. Both parts are effective.
4	Conclusion has the following two parts: restatement of the thesis or a summary of the body and a closing statement. One part is weak.
3	Conclusion may have the two parts: restatement of the thesis or a summary of the body and a closing statement. Both parts are weak.
2	Conclusion may be missing one of the two parts: restatement of the thesis or a summary of the body and a closing statement.
1	Conclusion is not complete or effective.

Effectiveness

Point 8 Topic Choice

Speakers should pick topics that will be interesting for the audience. Speakers should also be interested and knowledgeable in the chosen topic. Narrow the topic to fit the assignment (prepare enough examples/support).

5	The topic was very interesting/relevant to the audience, the speaker knows the topic well, and the topic was narrowed to fit the presentation. All three parts are effective.
4	The topic was interesting to the audience, the speaker knows the topic fairly well, and the topic was narrowed to fit the presentation. One part may be weak.
3	The topic seemed interesting to the audience, the speaker knows the topic fairly well, and the topic was narrowed to fit the presentation. Some parts are weak.
2	The topic did not seem very interesting to the audience, the speaker does not know the topic well, and/or the topic was not narrowed to fit the presentation.
1	The topic chosen was not appropriate for the presentation.

Point 9 Language Use

Speaker should proofread for errors before presenting. Sentences should be clear. Various types of sentences to maintain interest. Emphasize important points by using short clear sentences and some repetition.

5	All ideas are clearly expressed, sentence structures are varied, and there is some repetition for important points. There are no serious errors that hinder comprehension.
4	Most ideas are clearly expressed, sentence structures are varied, and there is some repetition for important points. There are few errors that hinder comprehension.
3	In general, ideas are clearly expressed, sentence structures are varied, and there is some repetition/summary for important points. There are some serious errors.
2	Most ideas are not clearly expressed, sentence structures are not often varied, and there is no repetition for important points. There are many serious errors.
1	Speaker was nearly incomprehensible due to structural errors that hinder understanding.

Point 10 Vocabulary

Speaker should be aware of the English level of the audience. Include definitions or examples of difficult/specialized vocabulary. Be careful of word choice and word forms.

5	Vocabulary usage was at a good level for the audience. Speaker included definitions for key words introduced. No serious errors in word choice or word forms.
4	Vocabulary usage was at a good level for the audience. Speaker omitted a few definitions for key words introduced. A few serious errors in word choice or word forms.
3	Vocabulary usage was not always at a good level for the audience. Speaker sometimes omitted definitions for key words. Some serious errors in word choice or word forms.
2	Vocabulary usage was often not at a good level for the audience. Speaker often omitted definitions for key words introduced. Many serious errors in word choice or word forms.
1	Speaker was nearly incomprehensible due to serious errors in word choice or word forms.

Point 11 Purpose

Speakers should be fulfilling the purpose of each speaking task: informing audience or changing opinion (persuading).

5	The speaker fulfilled the purpose of the presentation very well.
4	The speaker fulfilled the purpose of the presentation, but had a few weak sections.
3	The speaker fulfilled part of the purpose of the presentation, but had some weak parts.
2	The speaker did not adequately fulfill the purpose of the presentation, but made some good points during the presentation.
1	The speaker did not fulfill the purpose of the presentation at all.

Appendix C

List of categories identified in the communication studies literature

Personal and source credibility	Burrell and Koper (1998), McCroskey and Mehrley (1969), Montgomery (2001), O'Keefe (2002), Perloff (2003), Osborn and Osborn (2002)
Organization of speech	Allen (1991), O'Keefe (1997), Osborn and Osborn (2002), Struckman-Johnson and Struckman-Johnson (1996).
Effective language	Chimombo and Roseberry (1998), Gass and Seiter (1999), Hamilton and Hunter (1999), Osborn and Osborn (2002), Sandell (1977).
Audience adaptation	Bell (1984), Ladegaard (1995), Osborn and Osborn (2002).
Distracting language	Burgoon et al. (1990) and McCroskey and Mehrley (1969).
Conversational presentation	Osborn and Osborn (2002).

Appendix D

Rater and instructor feedback grouped into five categories

Credibility:

- Nice citing of evidence
- You could have included more sources
- You didn't indicate where any of your information came from
- There was no source for the information
- Cite where your information came from-why should we listen
- Enhance your credibility by citing where your information came from
- Very credible topic
- This topic is ideal for you-very credible

Distracting language:

- Watch filler words and nonfluencies (uh, um)
- Hesitations and nonfluencies detracted from the effectiveness of the message
- Pauses in inappropriate places
- Watch filler words and awkward transition words
- You didn't give the idea that you knew where you were going or what you were going to say next-evidenced with nonfluencies and hesitancy

Vocabulary use:

- You could have worked on the wording to make it sound more effective
- Be careful of your word selection
- Vivid imagery
- Powerful choice of words
- Effective word choice

Text organization:

- Make the speech structure clearer
- Focus more on organization

- Arguments are difficult to follow
- An effective organizational format
- Presenting the information in a logical way
- Provide continuity throughout the speech

Audience adaptation:

- Why is this topic of interest to your audience?
- Try using imagery or a narrative to establish some common ground
- Consider your audience when structuring the arguments



© 2017 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.



Cogent Education (ISSN: 2331-186X) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com

