Successful Lecturing

A Prospective Study to Validate Attributes of the Effective Medical Lecture

H. Liesel Copeland, PhD, David L. Longworth, MD, Mariana G. Hewson, PhD, James K. Stoller, MD

OBJECTIVE: **In a study conducted over 3 large symposia on intensive review of internal medicine, we previously assessed the features that were most important to course participants in evaluating the quality of a lecture. In this study, we attempt to validate these observations by assessing prospectively the extent to which ratings of specific lecture features would predict the overall evaluation of lectures.**

MEASUREMENTS AND MAIN RESULTS: **After each lecture, 143 to 355 course participants rated the overall lecture quality of 69 speakers involved in a large symposium on intensive review of internal medicine. In addition, 7 selected participants and the course directors rated specific lecture features and overall quality for each speaker. The relations among the variables were assessed through Pearson correlation coefficients and cluster analysis. Regression analysis was performed to determine which features would predict the overall lecture quality ratings. The features that most highly correlated with ratings of overall lecture quality were the speaker's abilities to identify key points ($r = .797$) and be engaging ($r = .782$), the lecture clarity ($r = .754$), and the slide comprehensibility ($r = .691$) and format ($r = .660$). The three lecture features of engaging the audience, lecture clarity, and using a case-based format were identified through regression as the strongest predictors of overall lecture quality ratings ($R^2 = 0.67$, $P = 0.0001$).**

CONCLUSIONS: **We have identified core lecture features that positively affect the success of the lecture. We believe our findings are useful for lecturers wanting to improve their effectiveness and for educators who design continuing medical education curricula.**

KEY WORDS: **lecturing; continuing medical education; research; reproducibility of results; medical faculty; evaluation studies.**

J GEN INTERN MED 2000;15:366–371.

The lecture is a staple of medical education that is widely employed in many venues, including continuing medical education. However, despite the widespread use of this educational medium, few studies have examined the features of the effective medical lecture. In a previous study

conducted over 3 years in a large intensive review of internal medicine course, we assessed which features of a lecture were most important to course participants. Several features were deemed important, including the clarity and readability of slides, the relevance of the lecture material to the participants, and the presenter's ability to identify key issues, engage the participants, and present materials clearly and with animation.[1] In addition, we found that course participants appreciated the use of an interactive computerized audience response system (ARS) and felt that such a system facilitated their learning. In that preliminary study, we used a survey design to obtain our results.

In the present study, we sought to validate these observations with a prospective research design using observers and rating tools. Specifically, we assessed in a subsequent offering of this large internal medicine review course the extent to which ratings of these originally identified features as well as several other potentially important features of medical lectures predicted the overall evaluation scores of individual lecturers by participants. This study addresses the question of what combination of lecture features, representing presenters' behaviors, lecture formats, and organizational factors, most efficiently explains the variance in lecture quality ratings.

METHODS

Background Information

The features of lecture quality were originally identified during the Cleveland Clinic Intensive Review of Internal Medicine Symposia in 1993, 1994, and 1995 (i.e., the fifth, sixth, and seventh course offerings), which were attended by a total of 1,221 participants and were conducted over 6 days with 50 sessions and 69 speakers. Ratings for the current validation study were obtained from the Ninth Intensive Review of Internal Medicine Symposium in 1997. All lecturers were advised to use interactive questions and engage the participants using the ARS. With the ARS each participant registers a response on a keypad and the responses are instantly tallied and displayed on a screen for review. Figure 1 presents an example of an interactive question that invites participants' responses.

Design

Because the candidate features of lecture quality were derived from our earlier study, the current research represents a prospective validation study. Data elements con-

Patient Information

Patient profile 1:
   55-year-old man with history of multiple trauma, ARDS, and prolonged intubation.

Patient profile 2:
   65-year-old man with long-standing rheumatoid arthritis and cricoarytenoid involvement.

Patient profile 3:
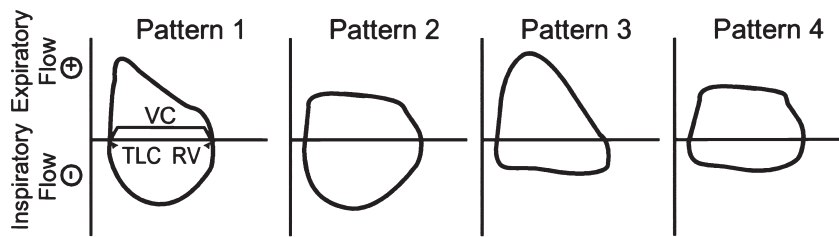   40-year-old woman with painful ears, saddle nose deformity, and arthraligias.

Patient profile 4:
   45-year-old woman with "factitious asthma" presenting as stridor.

Patient profile 5:
   30-year-old man with relapsing polychondritis and expiratory wheezing.

Patterns of the flow-volume loop

FIGURE 1. This example of an interactive question for engaging participants poses a series of 5 clinical vignettes with which each participant is invited to match the most appropriate flow-volume loop.

sisted of ratings of course quality and specific lecture features by course participants in June 1997. The dependent variables included the overall quality of the lecture ratings from all course participants and from a subset sample of 7 selected participants. The independent variables were the following lecture features: presenters' behaviors, i.e., 5 original features including engaging the participants, identifying important points, presenting material that is relevant to the participants, being clear, and using understandable and clear slides, and 1 additional feature for using humor; lecture formats, i.e., case-based, simulated board sessions, and use of the ARS for a pretest, during the talk, or for a posttest; and organizational factors, i.e., time of day and day in weeklong course.

Data Collection

Ratings for the analyses were obtained from 3 different groups: (1) All course participants rated the overall quality of each lecture on a 4-point interval scale (where 1 = poor and 4 = excellent) utilizing the ARS after each lecture; (2) after each lecture a convenience sample of 7 randomly selected course participants rated the overall lecture quality and the quality of the presenter's behavior features (original features) using a 5-point visual analogue scale with anchors of 1 and 5 (where 1 = awful and 5 = great); and (3) the 2 course codirectors (DLL and JKS) rated each lecture at its completion regarding the presenter's behavior features and coded the lecturer's use of humor and the lecture format features.

Statistical Analysis

Data were described by calculating means and standard deviations. To test whether the data were sufficiently good to merit analysis, reliability coefficients were estimated for ratings of lecture features by computing γ-coefficients and κ coefficients. Before addressing the study question of which variables would best predict lecture quality ratings, we assessed the relation between the variables by computing Pearson correlation coefficients. Because the independent variables were related, we used principal component cluster analysis as a selection process to avoid multicollinearity of variables in the regression. Variables explaining the greatest degree of variance in each cluster were then chosen for entry into the regression analysis. With lecture quality ratings by the entire group of participants as the dependent variable, modified stepwise multiple regression analysis was performed by sys-

tematically deleting individual lecture features (as rated by the 7 randomly selected participants and 2 codirectors) from the model until the most efficient prediction equation was obtained. A variable was left in the model if it had a squared semipartial correlation of at least ≥.01 and if deleting it caused a statistically significant change in the amount of variance explained by the regression equation. Statistical analyses were computed using SAS (SAS Institute Inc., Cary, NC) and Genova.[2]

RESULTS

The Course

The Ninth Cleveland Clinic Intensive Review of Internal Medicine Symposium was a 6-day course featuring 69 speakers and 50 teaching sessions. All but 7 speakers utilized the ARS. Of the 50 sessions, 16 were board simulation sessions, which were usually case-driven and employed the ARS.

Four hundred and twenty-four individuals attended the 1997 course from Puerto Rico, Saudi Arabia, and 34 states in the United States. Fifty-eight percent of the participants were currently trainees in internal medicine. Overall, 90.6% practiced internal medicine, 2.8% practiced an internal medicine subspecialty, 2.8% were family or general practitioners, and 1.0% represented other medical specialties. Participants' degrees were as follows: 85% MD, 11% DO, 2% MBBS, MB, MBCH, and 0.5% certified physician assistants.

Correlations of Specific Lecture Features with Overall Lecture Quality

Using a 4-point rating scale in which the maximal rating was 4 (excellent), the mean lecture rating by course participants was very good (3.48; [SD, 0.339]; range, 2.47–3.93). The median number of respondent participants was 280 (66%). Ratings of lecture features and techniques by the 7 randomly selected participants and the 2 course codirectors were highly consistent and judged to have adequate reliability for use in all analyses (feature γ-coefficients ranged from 0.48 to 0.75 and κ coefficients from 0.52 to 0.96). Overall consistency in rating the 6 lecture features was 0.75, and the average κ coefficient for the descriptive format features rated by the course codirectors was 0.82, indicating excellent concordance.

Table 1 presents Pearson correlation coefficients between the dependent variables and the independent variables. The correlation between lecture quality ratings by the subset of 7 participants and all course participants is .768 ($P < .001$), indicating that both groups agree on lecture quality. High correlation coefficients between overall lecture ratings and 6 specific lecture characteristics confirmed earlier results. Specifically, the participants' ratings of overall lecture effectiveness most closely correlated with the assessment of several lecture features: the lecturer's ability to identify important points and engage the participants, the overall clarity of the lecture, the understandability and clarity of the slides, and the relevance of the lecture to the participants.

Table 1. Correlation Matrix Relating Overall Lecture Ratings to Ratings of Specific Lecture Features*

	E	I	R	C	SU	SC	Course Participants' Overall Rating	Subset of Raters' Overall Rating
Lecturer								
Is engaging (E)	—	.91[†]	.76[†]	.91[†]	.78[†]	.71[†]	.78[†]	.95[†]
Identifies key points (I)		—	.77[†]	.88[†]	.69[†]	.80[†]	.80[†]	.91[†]
Is relevant to audience (R)			—	.70[†]	.55[†]	.47[†]	.64[†]	.76[†]
Is clear (C)				—	.82[†]	.74[†]	.75[†]	.92[†]
Uses humor	.24	.26	.13	.24	.09	.05	.24	.17
Uses understandable slides (SU)					—	.98[†]	.69[†]	.87[†]
Uses clearly formated slides (SC)						—	.66[†]	.80[†]
Lecture format								
Case-based approach	.34[‡]	.33[‡]	.30[‡]	.28	.35[‡]	.33[‡]	.35[‡]	.33[‡]
Board simulation	.35[‡]	.32[‡]	.18	.32[‡]	.38[‡]	.34[‡]	.19	.34[‡]
Uses ARS before test	−.01	−.04	.01	−.11	−.03	−.04	−.09	−.03
Uses ARS during test	.19	.23	.23	.16	.23	.23	.33	.31[‡]
Uses ARS after test	−.24	−.18	−.23	−.17	−.25	−.23	−.09	−.17
Outside factors								
Time of talk	.00	.04	−.00	−.04	−.02	−.02	−.04	−.02
Day of talk	−.15	−.06	−.02	−.07	−.24	−.24	.01	−.19

*Correlations between ratings of specific lecture features and overall lecture ratings were based on overall ratings from 2 sources: the whole group of course attendees and the subset of 7 raters (who also rated individual lecture features). E indicates that the lecturer "is engaging"; I, "identifies key points"; R, "is relevant to audience"; C, "is clear"; SU, "uses understandable slides"; and SC, "uses clearly formatted slides." ARS indicates audience response system.
[†]$P < .001$; [‡]$P < .01$.

In ratings by the 7 randomly selected participants, the 3 features that most highly correlated with the course participants' ratings of overall lecture quality (coefficients ≥ .754) were the presenter's abilities to identify important points and to engage the participants, and the lecture clarity. Notably, for neither group were overall lecture quality ratings strongly correlated with the use of humor, use of a board simulation format, use of the ARS before or after the presentation, or time of day and date of talk. Interestingly, the use of the ARS to pose a posttest was consistently associated with poorer ratings. There were modest correlations that achieved statistical significance relating the overall rating of lecture quality with the use of a case-based format ($r = .35$, $P = .003$) and with the use of the ARS to introduce new components of a talk throughout the lecture ($r = .33$, $P = .006$).

Explanatory Power of Lecture Features in Predicting Lecture Quality Ratings

Through the selection process with principal component cluster analysis (Table 2, it was found that 2 features demonstrated the greatest explanatory power: lecture clarity ($R^2 = .9010$) and the presenter's ability to engage the participants ($R^2 = .9013$). Because these variables explained the same amount of variance, they were both chosen for entry into the regression coefficient to determine which of them was more important in explaining the amount of variance in lecture ratings. Because of the multicollinearity imposed by using these 2 variables, we also computed a regression analysis with only 1 of the 2. The cluster analysis supported the centrality of the original lecture features in that the 6 previously identified lecture features made up 1 core component of our independent variables and explained the largest proportion of variance. The 3 other features selected through this analysis were the use of the ARS for testing after the material was presented, the use of a case-based format, and the use of humor.

Regression analysis using these 5 lecture features selected by the cluster analysis as independent variables showed that 3 features were the strongest predictors of overall lecture quality ratings by all course participants: the presenter's ability to engage the participants, the lecture clarity, and use of a case-based format (Table 3). Taken together, these three features explained approximately 66% of the variance in lecture quality ratings $F = 45.01$, $P = .0001$, Table 3). This explanation was chosen over using only the presenter's ability to engage the participant and use of a case-based format ($R^2 = .66$), or only the lecture clarity and use of a case-based format ($R^2 = .62$), because of the similarity in R^2 values and the inability to judge the relative importance of these two features.

DISCUSSION

Our findings in this prospective validation study appear robust and confirm the results of our previous survey study. Because this course was attended by multiple levels of learners (residents, practicing physicians) from numerous different settings (institutions and types of practice) and covered a comprehensive range of internal medicine topics, we feel these findings should generalize to lectures in other institutions, for participants with different training levels, and to various medical topics. We believe that our findings provide important empirically based guidelines for lecturers. Specifically, lecturers should be encouraged to engage the audience's attention, provide a clear and organized lecture, and use a case-based format. The lecture can be further enhanced by identifying important points and presenting relevant material with readable slides. The use of an ARS to present questions throughout a talk may supply one method of being organized, engaging, and summarizing or identifying important points.

Although these results may appear self-evident, our collective personal experience of witnessing poorly crafted, ineffective medical lectures suggests that these points warrant emphasis. Indeed, there is a long-standing discussion in the literature regarding the features that contribute to an effective lecture,[3-6] most of which is theoretical and lacking in empirical support. For example, the

Table 2. Cluster Analysis of Lecture Features*

Cluster 1	R^2 with Own Cluster	Cluster 2	R^2 with Own Cluster	Cluster 3	R^2 with Own Cluster	Cluster 4	R^2 with Own Cluster
Lecture clarity	.901	ARS after	.745	Case-based	.712	Uses humor	1.00
Engaging	.901	ARS during	.673	Board simulation	.583		
Identifies points	.838	ARS before	.267	Date of lecture	.398		
Slides, understandable	.817			Time of lecture	.095		
Slides, clear format	.776						
Relevance of material	.629						
Proportion of the variance explained by all clusters = .667							

*The lecture features are sorted into groups in which they are most closely associated. Overall the features represent 4 areas: lecturer's presentation techniques, method of using technology, specific lecture structure/format, and humor (which does not closely associate with any of the 3 previous categories, so it is categorized separately). ARS indicates audience response system.

Table 3. Results of the Regression Analysis: Predicting Lecture Quality Ratings from Features of a Lecture

Feature	Correlation with Overall Lecture Rating	B/β*	sr^2*
Engages/interests me	.78	.4859/.5214	.6476
Clarity of the lecture	.75	.3207/.2751	.0128
Case-based format	.35	.0877/.1120	.0112
Mean overall lecture rating (SD)	3.48 (0.339)	R^2* = .6717[†]	
		Adjusted R^2 = .6567[‡]	
		[‡]Intercept: B = −.1529	

The β weights (unstandardized and standardized: B/β) and the sr^2 are useful for interpretation of the importance of each variable in the regression equation. The unstandardized regression coefficient (B) is used with raw scores in the linear equation (regression line). The standardized regression coefficient is easier to interpret and allows one to see which variables (higher β) are more important to the solution. The squared semipartial correlation (sr^2) is the amount of variance added to R^2 by each variable at the point that it enters the equation.
[†]*R2 is the proportion of variance in lecture ratings explained by the linear combination of the 3 variables.*
[‡]*Adjusted R2 takes into account the positive change fluctuations that tend to make R2 an overestimate.*

literature on effective teaching and lecturing suggests that teachers should be clear, use effective communication, use explicit organization, be interactive, be aware of and responsive to class progress, develop group or individual rapport, demonstrate enthusiasm, stimulate learning, be flexible in their approachs, and tailor their approach to particular learners.[3,5,7,8] Similarly, in studies of lecture styles, findings suggest preference for a more pedagogically oriented style that incorporates such activities as articulating objectives, responding to students, using emphasis, summarization, and repetition frequently, using media effectively, supplying a supplemental handout in order to limit the content or detailed information of the lecture, and intending to produce a conceptual change.[3,9] The similarity between our findings and the views expressed in the literature both provide validation for our findings and suggest they generalize to other higher education settings outside medicine. Furthermore, the concepts of engaging the participants and identifying important points may well generalize to other teaching settings such as small group discussions, though the techniques for implementing these concepts may vary. Our study brings data-driven support to current educational views and theories that to be effective as a lecturer, one should *engage the participants, identify important points, and use clear, understandable visual aids.*

The ability of a lecturer to engage and interest learners is related to the contextual (usually clinical) relevance of the lecture content. Lecturers can gain participants' attention with provocative questions or problems. They should clearly demonstrate interest in and enthusiasm for the topic. Case-based lecturing greatly increases the clinical relevance of new material and enhances the participants' interest in the topic. The ARS allows the lecturer to introduce a topic by asking focused questions that may be provocative and intriguing, both at the beginning of and during the lecture. Lecturers can maintain participants' interest during the lecture through tailoring the content (the answers provided to the provocative questions) to the learners' perceived needs as well as to

their prior knowledge of a given topic. Humor should be used with caution and is not a necessary component of an effective lecture. Once the participants are engaged with the topic under discussion, the lecturer can proceed with the body of the lecture.

Lecturers should introduce the topic, clearly outline the scope of the lecture, present the main points in an organized fashion, and provide clear supporting materials and examples. It is seldom necessary for lecturers to reveal everything they know on a topic. Rather, they should identify the important ideas, show the connections between these ideas, and demonstrate a clear line of organized expert reasoning that integrates the ideas with the original focus question. At the end, the lecturer should apply the main points to the problem or case posed during the introduction, extrapolate to other similar situations, and close with an energized summary of the main points. The use of a "pop quiz" at the end of the lecture may not substitute for an adequate summary and was not shown to add to the perceived quality of a lecture in our results.

Slides and overhead transparencies enhance the clarity of a lecture and help the learner to understand material that is complex and difficult to describe. Poor slides may undermine an otherwise effective lecture. Good slides should be completely visible to every member of the audience. They should be simpler than the diagrams, graphs, and tables published in articles or textbooks, which often contain more information than is required to make a point. Visual representations of the main ideas being discussed are enhanced with the judicious use of color and background. Slides are useful to participants, but should not substitute as notes for the lecturer. Rather, the lecturer should highlight only the main ideas and explain, elucidate, substantiate, defend, or extrapolate the ideas being presented. Handouts should provide the participants with a summary of useful points to take home and can also serve to present further detailed information and facts. Printouts of the lecturers' spoken discourse may not be appropriate as handouts.

Several limitations of the current research warrant comment. Although the current study confirms the survey results obtained from our earlier study, they probably cannot be generalized to other medical education contexts such as precepting, one-on-one instruction, and bedside teaching. The relation of our findings to student achievement remains unstudied. Another possible concern with our study is the restricted range (possible ceiling and floor effects) in scores of the overall lecture quality ratings that may have limited the amount of variance to be explained, which could impact the regression. Despite the restricted range, our large sample size limits the impact that individual outliers would have on the regression equation.

Another limitation may result from our methods of dealing with the multicollinearity of variables through cluster analysis and a modified regression analysis. Specifically, by using clustering to drive our independent variable selection, we chose one set of variables to enter into the regression equation. A different method of selecting variables may have resulted in a different starting point and end equation. Also, though we modified the stepwise regression method to more closely reflect a hierarchical strategy rather than the stepwise exploratory method, the modifications were study-designed statistical criteria. Notwithstanding this potential objection, we believe that the logical basis for our cluster choices and the hierarchical approach used in generating the multivariate model supports our results.

Our results suggest that those who design continuing medical education courses could improve them through faculty development or by providing guidelines to their lecturers. In addition, they could collect data and give feedback on these specific, behaviorally based, and important lecture features.

In summary, this study validates the importance of previously identified features of effective lecturers. It allows us to predict success in continuing medical education lectures and provide practical recommendations for those who provide continuing medical education and for medical educators who want to improve their effectiveness as lecturers. Finally, our findings provide insight into physicians' preferences concerning lectures.

REFERENCES

1. Copeland HL, Stoller JK, Hewson MG, Longworth DL. Making the continuing medical education lecture effective. J Contin Educ Health Prof. 1998;18:227–34.
2. Brennan RL. Elements of Generalizability Theory. Iowa City, Iowa: ACT Publications; 1992.
3. Saroyan A, Snell LS. Variations in lecturing styles. Higher Educ.1997;33:85–104.
4. Irby D, DeMers J, Scher M, Matthews D. A model for the improvement of medical faculty lecturing. J Med Educ. 1976;51:403–9.
5. Pritchard RD, Watson MD, Kelly K, Paquin AR. Searching for the ideal: a review of teaching evaluation literature. In: Helping Teachers Teach Well. San Francisco, Calif: The New Lexington Press; 1998:3–42.
6. Gagnon RJ, Thivierge R. Evaluating touch pad technology. J Contin Educ Health Prof. 1997;17:20–6.
7. Centra JA. Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness. San Francisco, Calif: Jossey-Bass; 1993.
8. Hewson MG. Clinical teaching in the ambulatory setting. J Gen Intern Med. 1992;7:76–82.
9. Brown G, Bakhtar M, Youngman M. Toward a typology of lecturing styles. Br J Psychol. 1984;54:93–100.