Evaluation of physiology lectures conducted by students: comparison between evaluation by staff and students

Mahinda Kommalage and Sampath Gunawardena
Department of Physiology, Faculty of Medicine, University of Ruhuna, Galle, Sri Lanka

Submitted 15 August 2010; accepted in final form 8 November 2010

Kommalage M, Gunawardena S. Evaluation of physiology lectures conducted by students: comparison between evaluation by staff and students. Adv Physiol Educ 35: 48–52, 2011; doi:10.1152/advan.00091.2010.—As a peer-assisted learning process, minilectures on physiology were conducted by students. During this process, students lecture to their colleagues in the presence of faculty staff members. These lectures were evaluated by faculty staff and students simultaneously. The aim of this study was to compare feedback from faculty members and students on 66 minilectures conducted by students. Their perception of different qualities of lecture was assessed using a questionnaire. There were significant correlations between students and faculty members for many qualities of the lecture, including the speed of the lecture, retaining attention, clear introduction, and the overall quality of the lecture. However, ratings for gesture, eye contact, language usage, illustration usage, audiovisuals, voice usage, and important points stressed were significantly different between students and faculty members. Multiple regression analysis was performed to assess the degree of effect of different aspects of a lecture on its overall quality. Aspects such as gesture, eye contact, and language usage showed very low β-values, suggesting a poor contribution of these factors to the overall quality of the lecture for both students and faculty members. The speed of the lecture, retaining attention, and clear introduction were qualities that faculty members and students rated equally, and these were the main contributors to the overall quality of the lecture. Awareness about the possible discrepancy between ratings given by faculty members and students may be important when interpreting the evaluation results of formal lectures by these two groups.

student feedback; faculty feedback; peer evaluation; peer-assisted learning

PEER-ASSISTED LEARNING (PAL) is a process of acquiring of knowledge and skills through active help and support from status-equal or matched companions (16). Although teachers in PAL are not professional teachers or “experts” in the subject, previous studies (4, 8, 17) have shown that the effectiveness of PAL is equal to the main teaching process.

PAL has several advantages, such as providing an informal environment for better learning, improving communication skills, increasing intrinsic motivation, facilitating cooperation among students, learning to respect colleagues’ views, facilitating social interactions, and reducing the cost compared with professional teacher-assisted conventional learning (3, 6, 11, 15, 18).

The curriculum in the medical school of the University of Ruhuna (Galle, Sri Lanka) is of the traditional type. During the first and second years, students study physiology together with anatomy and biochemistry. The teaching activity for physiology consists of formal lectures, practical sessions, and tutorial classes. As a PAL process, minilectures conducted by students (popularly known as student seminars) have been used in this medical school for years. During this PAL process, students lecture to their colleagues in the presence of faculty staff members. The faculty staff supervises the preparation of lectures. Lecture topics are mainly on physiology with minor contribution from biochemistry, anatomy, and clinical medicine when relevant. These lectures were evaluated by faculty staff and students simultaneously.

Feedback from the recipient is considered one of the main methods of evaluating the effectiveness of any service. In terms of teaching, feedback from students is considered an important and essential evaluation tool (12). Previous authors (14) have emphasized the use of different sources of information in assessing a teaching process while questioning the reliability of student feedback. Combinations of students and peers (10), students and administrators (7), student self-evaluation (13), and students and residents (19) are commonly used evaluator combinations.

Students have given higher average ratings compared with faculty members (9), residents (19), department chairmen (7), and self-rating (20) in previous studies. On the other hand, students have given similar ratings to faculty peers (10) and self-evaluation (13) in some studies. Students have rated lectures less discriminatingly compared with trained observers (1). Residents have given higher ratings for clinical teaching than peer physicians (2).

Evaluation of formal lectures by peer faculty staff has been recently introduced in many places to further improve the quality of teaching. Some aspects of quality of teaching, such as the quality of the teaching materials and teaching skills, can be assessed by students as well as faculty members (12). Faculty staff may be better at assessing the quality and depth of contents and achievement of the expected outcome of the lecture. Expectations, subject knowledge, experience, and evaluation skills may influence the feedback. Therefore, when interpreting the feedback, it is important to consider the differences in these qualities among different groups who provided the feedback.

The aim of this study was to compare the ratings given by students and faculty members on minilectures conducted by students. We compared the mean rating on different qualities of the lecture within each group and between the two groups as well as the pattern of rating. We also investigated the contribution of each individual aspect of teaching to the overall quality of the lecture and correlation of ratings of the two groups.

Address for reprint requests and other correspondence: M. Kommalage, Dept. of Physiology, Faculty of Medicine, Univ. of Ruhuna, PO Box 70, Galle, Sri Lanka (e-mail: mahinda1@gmail.com).
The questionnaire consisted of 12 statements that addressed different qualities of the lecture. Each statement was rated on a six-point Likert scale, where 1 = “strongly disagree” and 6 = “strongly agree.” All evaluations were done during the lecture, and the questionnaires were collected afterward and analyzed.

Of the 12 statements, 10 statements were analyzed individually (use of gesture, use of eye contact, use of language, use of illustration, use of voice, important points stressed, speed of the lecture, retaining attention, clear introduction, and overall quality of the lecture), and 2 statements were averaged (clear audiovisuals and handling of audiovisuals) to produce one quality on the use of audiovisuals. In addition, a total composite mark was calculated by averaging the ratings for the 12 individual qualities.

RESULTS

Reliability within the questionnaire was calculated using internal consistency. Cronbach’s α-value was 0.82. Correlation of student and faculty ratings. Pearson correlations between the student and faculty ratings were calculated. Statistically significant correlations were found for use of gesture, eye contact, use of illustrations, use of audiovisuals, use of voice, important points stressed, retaining attention, clear introduction, and overall quality of the lecture. Correlations between faculty members and students were not statistically significant for language usage and speed of the lecture (Table 1). The intercorrelations among all qualities within faculty and student ratings are shown in Tables 2 and 3. Higher correlation was found between all qualities for both students and faculty members.

Comparison of student and faculty ratings. Mean ratings of students and faculty members for the use of gesture, use of eye contact, use of language, use of illustration, use of audiovisuals, use of voice, and important points stressed were significantly different (Table 4). Faculty members gave lower ratings than students for all of these qualities. Mean ratings of student and faculty members for the speed of the lecture, retaining attention, and clear introduction were not significantly different, nor were mean ratings for the overall quality of the lecture.

Determination of overall quality. The contribution of each individual aspect of competence to the overall quality of the lecture by both students and faculty members was assessed. Multiple regression analysis was performed, and the results are shown in Table 5. Qualities such as gesture, eye contact, and language usage (type 1 qualities) showed lower β-values (standardized regression coefficients), suggesting a poor contribu-
tion to the rating of the overall quality of the lecture by both students and faculty members. The speed of the lecture, retaining attention, and clear introduction of the lecture (type 3 qualities) showed comparatively high β-values, suggesting a greater contribution to the rating of the overall quality of the lecture by both students and faculty members. There was also a group of qualities whose contribution was intermediate (type 2 qualities).

There was a high correlation between rating on the overall quality of the lecture and total score (total composite). Pearson correlations were 0.76 and 0.75 for faculty members and students, respectively.

**DISCUSSION**

This study demonstrates that there are differences between student and faculty ratings for many individual qualities of a lecture. Some qualities showed a wider difference than others. Furthermore, it is clear that faculty ratings are lower than that of students on many individual qualities. Faculty staff expect higher standards than students; thus, they give lower ratings. Unlike students, faculty staff are experts in assessing the content of the lecture as well as the achievement of the expected outcome of the lecture; students are probably less able to do so. Faculty members are also more experienced in lecture evaluation since they have been evaluating them for a longer period of time than students. Therefore, it is reasonable for faculty members to expect higher standards. In agreement with this finding, previous studies reported that students gave higher average ratings for assessed qualities compared with faculty members (9), residents (19), department chairmen (7), and self-ratings (20).

Qualities such as gesture, eye contact, and language usage did not contribute significantly to the decision on the overall quality of the lecture by both students and teachers. Nonetheless, these qualities had a significant difference in ratings between students and faculty members. Among all qualities, use of gesture showed the widest difference. We categorized gesture, eye contact, and language as type 1 qualities. Compared with other qualities, evaluators should pay more attention and be more experienced to assess these qualities accurately. Faculty members might have paid specific attention to these aspects since they are more experienced. In a comparison of evaluations done by people from different disciplines, Kane et al. (9) reported that eye contact and gesture were less-evaluated qualities of a lecture. Being a group with less experience, students probably face difficulty in assessing qualities such as gesture and eye contact. When students were asked to assess these qualities, they may assess them superficially and rate them higher than faculty members. A previous study (9) showed that students wanted to assess only a few qualities in a lecture compared with faculty members. Therefore, we can consider these as “difficult to evaluate” qualities.

Table 3. Correlation between all qualities (intercorrelations) of student ratings

<table>
<thead>
<tr>
<th>Use of Gesture</th>
<th>Eye Contact</th>
<th>Use of Language</th>
<th>Use of Illustration</th>
<th>Audiovisuals</th>
<th>Use of Voice</th>
<th>Important Points Stressed</th>
<th>Speed of Lecture</th>
<th>Retaining Attention</th>
<th>Clear Introduction</th>
<th>Overall Quality of the Lecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye contact</td>
<td>0.678</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of language</td>
<td>0.301</td>
<td>0.261</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Illustration</td>
<td>0.291</td>
<td>0.295</td>
<td>0.334</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audiovisuals</td>
<td>0.291</td>
<td>0.334</td>
<td>0.455</td>
<td>0.416</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of voice</td>
<td>0.263</td>
<td>0.302</td>
<td>0.516</td>
<td>0.294</td>
<td>0.359</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important points stressed</td>
<td>0.419</td>
<td>0.537</td>
<td>0.328</td>
<td>0.332</td>
<td>0.37</td>
<td>0.318</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed of the lecture</td>
<td>0.403</td>
<td>0.326</td>
<td>0.471</td>
<td>0.24</td>
<td>0.402</td>
<td>0.4</td>
<td>0.318</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retaining attention</td>
<td>0.524</td>
<td>0.574</td>
<td>0.426</td>
<td>0.4</td>
<td>0.431</td>
<td>0.46</td>
<td>0.559</td>
<td>0.599</td>
<td>0.503</td>
<td></td>
</tr>
<tr>
<td>Clear introduction</td>
<td>0.433</td>
<td>0.423</td>
<td>0.414</td>
<td>0.302</td>
<td>0.478</td>
<td>0.478</td>
<td>0.472</td>
<td>0.451</td>
<td>0.507</td>
<td>0.639</td>
</tr>
<tr>
<td>Overall quality of the lecture</td>
<td>0.446</td>
<td>0.506</td>
<td>0.461</td>
<td>0.431</td>
<td>0.556</td>
<td>0.489</td>
<td>0.544</td>
<td>0.537</td>
<td>0.639</td>
<td>0.563</td>
</tr>
</tbody>
</table>

Each quality was compared with all the other qualities. Pearson’s correlation coefficient values are shown. All were significant with \( P < 0.001 \).

Table 5. Dependency of the overall quality of the lecture

<table>
<thead>
<tr>
<th>Use of Gesture</th>
<th>Eye Contact</th>
<th>Use of Language</th>
<th>Use of Illustration</th>
<th>Audiovisuals</th>
<th>Use of Voice</th>
<th>Important Points Stressed</th>
<th>Speed of Lecture</th>
<th>Retaining Attention</th>
<th>Clear Introduction</th>
<th>Overall Quality of the Lecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye contact</td>
<td>0.54</td>
<td>0.97</td>
<td>0.097</td>
<td>0.017</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of language</td>
<td>0.020</td>
<td>0.517</td>
<td>-0.010</td>
<td>0.794</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of gesture</td>
<td>0.037</td>
<td>0.173</td>
<td>0.020</td>
<td>0.560</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of illustration</td>
<td>0.089</td>
<td>0.000</td>
<td>0.096</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audiovisuals</td>
<td>0.155</td>
<td>0.000</td>
<td>0.155</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of voice</td>
<td>0.102</td>
<td>0.000</td>
<td>0.107</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important points stressed</td>
<td>0.119</td>
<td>0.000</td>
<td>0.145</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed of the lecture</td>
<td>0.191</td>
<td>0.000</td>
<td>0.177</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retaining attention</td>
<td>0.193</td>
<td>0.000</td>
<td>0.192</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clear introduction</td>
<td>0.142</td>
<td>0.000</td>
<td>0.119</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Shown are β-values of multiple regression analysis used to determine which features would predict the overall quality of the lecture.

Values are means (SD). *Significant difference from the respective rating by faculty members (\( P < 0.05 \) by Student’s \( t \)-test).
Qualities such as the speed of the lecture, retaining attention, and clear introduction can be considered as main determinants of the overall quality of the lecture. These qualities had higher β-values compared with others. Ratings of students and faculty members on these qualities did not show a significant difference, suggesting that both groups assessed these qualities equally. We categorized these three qualities as type 3 qualities. They seem to be evaluated accurately irrespective of experience or expertise. Therefore, we can consider them as “easy to evaluate” qualities. Flournoy et al. (5), in a study of medical, dental, and physician assistant students, showed that characteristics related to retaining student attention such as interest, excitement, and enthusiasm were frequently mentioned lecture qualities, whereas characteristics such as visual aids were not.

Illustration usage, audiovisuals, voice usage, and important points stressed were categorized as type 2 qualities. These qualities significantly contributed to the overall quality of the lecture and showed a significant difference between faculty and student ratings. These qualities are “intermediate in strength” and fall between type 1 and type 3 qualities.

Students and faculty members had similar perceptions about the overall quality of the lecture. Even though students are less experienced in assessing the individual qualities of a lecture, both groups seemed to be equally capable of assessing the overall quality of a lecture. The fact that there were similar β-values for individual characteristics suggests that the contribution of each individual characteristic to the decision on the overall quality of the lecture was similar for both students and faculty members.

Evaluation of formal lectures by students is an established practice in this medical school. Faculty staff improve the quality of their teaching based on student feedback. More recently, peer evaluation by faculty staff was introduced to further improve the quality of teaching in many institutions. Some faculty members in this medical school perceived differences between student and peer evaluation results of the lectures. However, a comparison of peer evaluation with student evaluations of the formal lecture was not feasible. In many evaluation sessions of formal lectures, the number of peer evaluators was few (2 or 3 in this medical school), whereas the number of student evaluators was many. Previously, Leamon et al. (10) performed such a comparison using video-recorded lectures. Other than that study, we did not find medical education-related studies comparing the feedback of students and faculty members.

Evaluation of formal lectures by students is an established practice in this medical school. Faculty staff improve the quality of their teaching based on student feedback. More recently, peer evaluation by faculty staff was introduced to further improve the quality of teaching in many institutions. Some faculty members in this medical school perceived differences between student and peer evaluation results of the lectures. However, a comparison of peer evaluation with student evaluations of the formal lecture was not feasible. In many evaluation sessions of formal lectures, the number of peer evaluators was few (2 or 3 in this medical school), whereas the number of student evaluators was many. Previously, Leamon et al. (10) performed such a comparison using video-recorded lectures. Other than that study, we did not find medical education-related studies comparing the feedback of students and faculty members.

In this PAL process, faculty members and students were together and evaluated the process. Many PAL processes described in the literature were evaluated only by students, but here we had a unique opportunity to directly compare evaluations of both faculty members and students. The high correlation of the evaluation by the two groups suggests the success of PAL process.

Even formal lecturers should be aware of the existence of these differences when interpreting results of student and peer faculty evaluations of lectures. They also need to consider, when preparing and delivering a lecture, that students concentrate less on certain individual qualities compared with faculty staff. It is apparent that if lecturers want to assess the individual qualities of the lecture, it is better to use feedback from peer faculty than students. Faculty members are more critical of many individual qualities than students.

We can highlight a few limitations of this study. In the evaluation process using statements in the questionnaire, we need to consider the subjective nature of some responses, especially from students. Different individuals understand statements differently, which may influence our results. Since students were rating their colleague’s lectures, they may be biased and could have given higher ratings than faculty members. Since lectures were conducted by students, faculty members may have been more critical and wanted to improve certain qualities of lectures. That may have lead to lower ratings from faculty members. However, in assessing the overall quality of the lectures, student and faculty ratings were similar, and the high correlations indicate that they ranked the lectures the same way. One advantage of PAL is that it provides an informal environment for better learning. In this PAL process, the presence of faculty staff might have altered the informality of the process. That may have some influence on student learning and their evaluation of the lectures.

In conclusion, there is a clear discrepancy in ratings given by faculty members and students on lectures. Different qualities of the lecture are rated differently by these two groups. Interestingly, the speed of the lecture, retaining attention, and clear introduction are qualities that faculty members and students rated equally, and these are the main contributors to the overall quality of the lecture. Awareness about the possible differences between student and faculty evaluations may be important for faculty staff when interpreting evaluation results of formal lectures by these two groups.

ACKNOWLEDGMENTS

The authors express gratitude to John J. Norcini for suggestions on improving the manuscript.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES


