The Correlation between Faculty Evaluation Ratings by Undergraduate students, Postgraduate students and Authorities in Shahid Beheshti Dental School

1Soodeh Tahmasbi *2Azam Valian 3Maryam Moezzyzadeh

1Assistant Professor, Dept. of Orthodontics, Dental School, Shahid Beheshti University of Medical Sciences, Tehran- Iran.
*2Corresponding Author: Assistant Professor, Dept. of Operative Dentistry, Dental School, Shahid Beheshti University of Medical Sciences, Tehran- Iran. E-mail: azamvalian@dent.sbu.ac.ir
3Associate Professor, Dept. of Operative Dentistry, Dental School, Shahid Beheshti University of Medical Sciences, Tehran- Iran.

Abstract

Objective: Instructor evaluation is the most complex type of evaluation due to the low reliability and inaccuracy of the means and methods of assessment. In order to solve this issue, it has been suggested that a combination of evaluation ratings be utilized to cast a final judgment. In Shahid Beheshti Dental School (SBDS), instructor evaluation is carried out by five groups of undergraduate students, post-graduate students, chief of department, deputy of education, and the school dean. The aim of this study was to assess the correlation between instructor evaluation ratings offered by undergrad and post-grad students and the authorities in SBDU during 2009-2010.

Methods: In this cross-sectional study, instructor evaluation scores given by the undergraduate students, post-graduate students, chief of department, deputy of education, and the university dean were separately retrieved and analyzed by SPSS version 11.5 software. Pearson’s correlation coefficient was used to assess the correlation between scores given by the abovementioned five groups for instructor evaluation. In order to determine the contribution of each group’s rating to the final evaluation score of an instructor, Variance Component Ratio was calculated.

Results: Assessment of the correlation coefficient between the opinions of undergraduate students, post-graduate students, chief of department, deputy of education, and the school dean revealed that a moderate significant correlation only existed between the perspectives of the dean and the deputy of education (r=0.612) and no other correlation was found between the remaining groups. Variance Component Ratio Analysis showed that given the score weights were similar in all groups, post-grad students had the highest (50.89%) and the deputy of education had the lowest contribution (1.11%) to the change in instructor evaluation score.

Conclusion: Our study results demonstrated that a moderate significant correlation (0.6) existed only between the ratings offered by the deputy of education and the school dean.

Key words: Evaluation, Scientific faculty, Correlation

Please cite this article as follows:

Received: 11.06.2011     Final Revision: 20.03.2012     Accepted: 28.03.2012

Introduction:

Assessment of the instructional skills of mentors has a significant application for enhancement of the quality of educational systems. Therefore, instructional assessment is routinely performed for each educational activity to ensure its correct path and optimal quality (1). Evaluation is defined as collecting information about the activities, characteristics and outcome of a program in order to cast a judgment, improve the efficacy of the program or increase awareness for decision making regarding future programs (2). Data obtained from evaluation provides feedbacks for educational programmers and instructors. By utilizing these feedbacks, they can make informed decisions to improve their method of teaching and become aware of their own success rate. Furthermore, the majority of faculty members have learned teaching through
watching their own instructors and up until recently, there was no official education regarding the methods of instruction for the majority of educational programs. Due to the presence of such shortcomings, authorities in charge of medical education provided some opportunities for education of mentors and instructors in this respect during the past two decades (3). Thus, the results of instructor evaluations can provide the authorities and programmers with valuable information regarding the necessity of holding such educational courses. Instructor evaluation is the most complex form of assessment due to the low reliability and precision of means and methods of assessment. The applied assessment methods and current information sources cannot offer accurate and non-objective information. Therefore, designing an efficient standard means for evaluation seems necessary (4, 5).

Student evaluation of teaching is a method commonly used in most countries and especially Iran for determining the quality of instruction by a mentor or an instructor (4, 6, 7). According to some authors, student evaluation is the best type of assessment because students are directly taught by the instructors and are in best position to judge their instructional skills (8, 9).

However, these evaluations have not always been well accepted. Some chiefs, instructors and students believe that student evaluations are valid and reliable while some authors have claimed them to be invalid and unreliable. Some studies have demonstrated that the perspectives of students on their instructors may be biased and influenced by factors unrelated to the instructor’s performance. Numerous studies have been conducted on this matter during the past couple of years (10, 11). However, despite such controversies, student evaluations of teaching are increasingly used to decide on the efficacy of instruction (9).

In order to solve this problem, it has been suggested that a combination of evaluation ratings be used to cast a final judgment (4). Therefore, in SBDS instructor evaluation is done by five different groups of undergrad students, post-grad students, chief of the department, deputy of education and school dean. However, in order to be able to use the results of these evaluations in the school's macro-decisions, the reliability and validity of the ratings offered by these groups need to be confirmed. To date, only the reliability of ratings by one of these groups namely post-grad students has been assessed which was not acceptable (12). The validity of ratings offered by each of the five mentioned groups has not been studied either. Thus, in the first step, we aimed to assess the correlation and agreement among the ratings offered by these five groups. If the mentioned ratings have a highly significant correlation with one another, it is indicated that the mentioned five groups all assess a common issue; which, is the success rate of the respective instructor in achieving instructional goals. Otherwise, complementary studies will be required to separately assess the factors affecting the instructor evaluation ratings offered by each one of the mentioned five groups.

The aim of this study was to determine the correlation between the instructor evaluation ratings given by students and authorities in SBDS during 2009-2010.

**Methods:**

This cross sectional study was conducted on SBDU instructor evaluation ratings during 2009-2010. Each instructor was allocated a code in order for the process of data collection to be confidential. Scores given to each instructor in 2009-2010 educational year by the 5 understudy groups including undergrad students, post-grad students, chief of the department, deputy of education and the school dean were separately collected. Evaluation questionnaire was different.
for the 5 understudy groups covering different aspects. The questionnaire designed for undergrad and post-grad students contained questions about order, method of teaching, manner, scientific level and overall evaluation of the mentor. The questionnaire designed for the school authorities included questions on a Likert scale of 1-5 (very poor to excellent) regarding the manner, scientific and behavioral qualifications, teaching capabilities, interaction with the educational environment and colleagues and physical attendance. Pearson’s correlation coefficient was applied to determine the correlation between each mentor’s ratings from the viewpoints of the mentioned 5 groups. In order to determine the contribution of different ratings, Variance Component Ratio Analysis was calculated for each group’s rating. For determining the contribution of different ratings, given the score weights were similar in all groups, first the ratings were standardized and then the Variance Component Ratio was calculated based on the new values. For assessing the contribution of each group’s rating to the final score of a mentor, first the percentage of each rating in the final score was calculated and the obtained mean was considered as the contribution of that specific opinion. Data were analyzed using SPSS version 11.5 software.

**Results:**

Two-way ANOVA showed that of a total of 110 faculty members, 68 had evaluation scores from all 5 groups and entered the study. The correlation coefficients between the opinions of undergrad students, post-grad students, chiefs of the departments, deputy of education and school dean are demonstrated in Table 1. As observed, a significant correlation was only detected between the perspectives of the dean and deputy of education and no significant correlation was found between other groups. In order to determine the effect (contribution) of each group’s opinion on a mentor’s final score, Variance Component Ratio was calculated for each group.

<table>
<thead>
<tr>
<th>Table 1- Correlation coefficients between the instructor evaluation ratings offered by undergrad students, post-grad students, chiefs of departments, deputy of education and university dean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Undergraduate students</strong></td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td><strong>Undergraduate students</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Postgraduate students</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Chiefs of departments</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Deputy of education</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>University dean</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Table 2 shows the results of this test. The right column of Table 2 reveals the contribution of each group to the final score given the score weights are similar in all groups. However, in this study, these rates were not similar and the scores weights given by the undergrad students, post-grad students, chiefs of the departments, deputy of education and school dean were 36, 18, 15, 16 and 15, respectively. Thus, the next column in Table 2 shows the contribution of
each group’s rating with the current situation of score weights. In order for all groups to have equal contribution to the final score, the left column in Table 2 shows the required score weights considering the present variance. Table 3 shows the percentage of score obtained by the instructors from the total possible score in each group. As observed, instructors obtained the highest scores from the chiefs of departments.

Table 2- Variance component ratio analysis for determining the contribution of each group’s rating to the final evaluation score

<table>
<thead>
<tr>
<th></th>
<th>Weight percentage in case of similar score weights</th>
<th>Contribution with actual current score weights</th>
<th>Contribution with similar score weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate students</td>
<td>40.56</td>
<td>49.16</td>
<td>0.95</td>
</tr>
<tr>
<td>Postgraduate students</td>
<td>50.89</td>
<td>44.06</td>
<td>1.06</td>
</tr>
<tr>
<td>Chiefs of departments</td>
<td>4.89</td>
<td>4.24</td>
<td>10.99</td>
</tr>
<tr>
<td>Deputy of education</td>
<td>1.11</td>
<td>0.77</td>
<td>60.47</td>
</tr>
<tr>
<td>University dean</td>
<td>2.53</td>
<td>1.26</td>
<td>26.53</td>
</tr>
</tbody>
</table>

Table 3- The mean scores obtained by instructors from each group

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate students</td>
<td>84.4</td>
<td>15.4</td>
<td>46.8</td>
<td>100</td>
</tr>
<tr>
<td>Postgraduate students</td>
<td>71.8</td>
<td>17.3</td>
<td>30.6</td>
<td>99.3</td>
</tr>
<tr>
<td>Chiefs of departments</td>
<td>95.6</td>
<td>4.4</td>
<td>75</td>
<td>100</td>
</tr>
<tr>
<td>Deputy of education</td>
<td>88.4</td>
<td>4.8</td>
<td>81.3</td>
<td>93.8</td>
</tr>
<tr>
<td>University dean</td>
<td>88.4</td>
<td>3.2</td>
<td>81.3</td>
<td>93.8</td>
</tr>
</tbody>
</table>

Discussion:

Undoubtedly, “instructor evaluation” due to its nature is among the most challenging subjects in educational systems (13). Revision of educational systems in the universities and designing research-based educational activities are both strongly linked to instructor evaluation. Therefore, assessment of the eligibility, performance, knowledge and capabilities of the faculty members is in the focus of attention of higher education councils. All educational authorities are now searching for an ideal method to achieve this goal (4). Several models are available for faculty evaluation such as the colleague assessment, student assessment and self-assessment (14, 15). The model adapted and routinely used by the SBDS includes undergrad student assessment, post-grad student assessment, chief of department assessment, deputy of education assessment and the school dean assessment.
In a pilot study conducted in 1998 in Stanford University School of Medicine, it was concluded that faculty development with a combination of assessment methods can significantly improve the instructional skills of basic science instructors (6).

Also, Motlagh and Jahanmardi in their review study in 2002 on 100 review and original articles found that student ratings of instruction do not have the required validity and cannot be reliably and solely used for instructor evaluation. Other complementary methods are required for this purpose. These methods should include a combination of measures for the assessment of the mentor’s instructional skills and only reflect factors controllable by the instructor (17).

In another study by Fatahi et al, in 2003, the opinion of academic members of Kerman University M.S was sought about the educational evaluation of lecturers in 2002-2003. In their study, 113 faculty members were questioned and it was found that the majority of instructors agreed with the nature of evaluation and believed that this evaluation should be done with the use of multi-dimensional methods, proper questions, appropriate interaction and confidential feedback of results since instructor evaluation is influenced by several factors. In order to encourage the mentors and authorize the findings, these ratings should better be considered for the annual promotion and academic rankings of instructors and selection of the instructor of the year (18).

Evaluation of perspectives of 375 instructors selected from major Canadian universities on the utility of student ratings of instruction in 2009 revealed that mentors consider the “instructor evaluation” as an acceptable means for the assessment of the institutional integrity. They believed it to be beneficial for making summative decisions by the authorities. However, they stated that this type of assessment has small value in enhancement of their own performance (20).

The results of the present study showed a significant correlation only between the perspectives of the dean and deputy of education and no other significant correlation was detected. This finding indicates that the perspectives of the undergrad students, post-grad students, chiefs of departments, deputy of education and the school dean on the instructor evaluation are not in accord with one another. Such difference in opinions may be attributed to the use of different questionnaires for the 5 groups. The questionnaires designed for undergrad and post-grad students mainly focused on the quality of theoretical and clinical instruction by the mentors; while, the questionnaires designed for the dean and deputy of education mainly focused on the interaction between faculty members in the university; the chiefs of departments were questioned about both of the mentioned domains. No agreement between the perspectives of the undergrad and post-grad students, despite the relative equality of the questioned domains, necessitates further scrutiny. Different interaction of faculty members with the undergrad and post-grad students may be responsible in this regard. A group of faculty members are assigned to work with the under-grad students while some others have the responsibility to interact with postgraduate students.

Similar results were obtained in a study by Moezi et al, (21). They studied the effect of difference in ratings offered by the students and authorities on the final faculty assessment score in Shahrekord University M.S in 2004. They evaluated the instructor assessment scores in three semesters and three different groups of
basic science, clinical medicine and nursing and midwifery. Study results demonstrated significant differences between student and authority ratings. One-way ANOVA failed to find a significant difference in student ratings in the three groups but significant differences were noted in authority ratings between three groups. AghaMolaei and Abedini in 2007 compared student evaluation and self-evaluation of faculty members in School of Health of Hormozgan University M.S (22). The mean self-assessment and student-assessment scores were 89.7± 6.3 and 85.1 ± 6.5; which, were significantly different. The correlation coefficient between the mean self-assessment and student-assessment scores was 0.26. The self-assessment score of the majority of instructors (75% of cases) was higher than their student assessment score.

In a similar study by Miron in 1988 on the correlation of student evaluation and self-assessment of instructors, 52 instructors and their students filled out a 20-question questionnaire. A low correlation (mean: 0.28) existed between the student evaluation and self-evaluation of faculty members. Furthermore, self-assessment scores of less experienced instructors were closer to student ratings (23). Goharian et al, in 2005 reported similar results in their study on comparison of resident evaluation and self-evaluation of attending surgeons in Isfahan University M.S (24).

In addition to determining the level of correlation, the present study assessed the contribution of each group’s perspectives to the final score using variance component ratio analysis in two situations. In the first situation, the score weights were considered similar in all groups. In this situation, post-grad students with 50.89% had the greatest contribution to the change in final instructor evaluation scores followed by undergrad students (40.56%). School dean (2.53%) and deputy of education (1.11%) had the least contributions in this regard. In the second situation, percentage of contribution was calculated in the current situation of ratings. In this situation, undergrad students’ contribution increased to 46.16% considering the score weight of 36 and ranked first. Post-grad students ranked second with 44.06% contribution and score weight of 18. Deputy of education and school dean had the lowest contributions. Therefore, we may state that 49.16% of the change applied to instructor evaluation scores is based on the perspectives of undergrad students.

In this situation, although the score weights of post-grad students, chiefs of departments, deputy of education and the dean were not significantly different (18, 15, 16 and 15, respectively), the contribution of post-grad students to the change in final evaluation score was significant (44.06 versus 4.24, 0.77 and 1.26, respectively). Another point worthy of noting is the fact that changing the score weight did not change the contribution of chiefs of departments to the final score (4.89 versus 4.24). Evaluation of the mean percentage of scores gained by the instructors from each of the 5 groups indicated that faculty members gained the highest acceptance from the chiefs of departments and gained 95.6% of the maximum possible score from this group. School dean (mean 88.4%), deputy of education (mean 88.4%), undergrad students (mean 84.4%) and post-grad students (71.8%) ranked next in this regard. It means that post-grad students evaluate their instructors with greater scrutiny.

Considering the results of this study and differences among the perspectives of the 5 groups, it appears that combination of different evaluation scores for one instructor and summing them up as one final score cannot indicate the success rate of an instructor in achieving instructional goals. Therefore, complementary studies are recommended to evaluate the factors affecting the ratings in each group. Also, it would be better if the instructors were provided with separate evaluation ratings from different groups so that they could benefit
the most from this process and enhance their educational performance and instructional techniques.

**Conclusion:**

The present study results revealed a significant correlation only between the instructor evaluation ratings offered by the deputy of education and school dean and no other correlation was detected between the other groups. Furthermore, pot-grad and under-grad students had the greatest contribution to the change in final evaluation score of instructors.

**Acknowledgment:**

This study was approved by the Research Council and Ethics Committee of SBDS. The authors would like to express their gratitude to the Research Deputy of SBDS for funding this research project.

**References:**