Validity of Peer Evaluation for Team-Based Learning in a Dental School in Japan

Keisuke Nishigawa, DDS, PhD; Rika Hayama, DDS; Katsuhiro Omoto, DDS; Kazuo Okura, DDS, PhD; Toyoko Tajima, DDS, PhD; Yoshitaka Suzuki, DDS, PhD; Maki Hosoki, DDS, PhD; Mayu Ueda, DDS, PhD; Miho Inoue, DDS, PhD; Omar Marianito Maningo Rodis, DDS, PhD; Yoshizo Matsuka, DDS, PhD

Abstract: The aim of this study was to determine the validity of peer evaluation for team-based learning (TBL) classes in dental education in comparison with the term-end examination records and TBL class scores. Examination and TBL class records of 256 third- and fourth-year dental students in six fixed prosthodontics courses from 2013 to 2015 in one dental school in Japan were investigated. Results of the term-end examination during those courses, individual readiness assurance test (IRAT), group readiness assurance test (GRAT), group assignment projects (GAP), and peer evaluation of group members in TBL classes were collected. Significant positive correlations were found between all combinations of peer evaluation, IRAT, and term-end examination. Individual scores also showed a positive correlation with group score (total of GRAT and GAP). From the investigation of the correlations in the six courses, significant positive correlations between peer evaluation and individual score were found in four of the six courses. In this study, peer evaluation seemed to be a valid index for learning performance in TBL classes. To verify the effectiveness of peer evaluation, all students have to realize the significance of scoring the team member's performance. Clear criteria and detailed instruction for appropriate evaluation are also required.

Dr. Nishigawa is Professor, Department of Oral Health Sciences, Faculty of Health and Welfare, Tokushima Bunri University, Tokushima, Japan; Dr. Hayama is a Postgraduate Student, Department of Stomatognathic Function and Occlusal Reconstruction, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan; Dr. Omoto is a Postgraduate Student, Department of Stomatognathic Function and Occlusal Reconstruction, Institute of Biomedical Sciences, Tokushima University Graduate School, Tokushima, Japan; Dr. Okura is Associate Professor, Department of Stomatognathic Function and Occlusal Reconstruction, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan; Dr. Tajima is Assistant Professor, Oral Implant Center, Tokushima University Hospital, Tokushima, Japan; Dr. Suzuki is Assistant Professor, Department of Stomatognathic Function and Occlusal Reconstruction, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan; Dr. Hosoki is Assistant Professor, Department of Stomatognathic Function and Occlusal Reconstruction, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan; Dr. Ueda is Assistant Professor, Department of Stomatognathic Function and Occlusal Reconstruction, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan; Dr. Inoue is Assistant Professor, Department of Stomatognathic Function and Occlusal Reconstruction, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan; Dr. Rodis is Associate Professor, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan; and Dr. Matsuka is Professor, Department of Stomatognathic Function and Occlusal Reconstruction, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan. Direct correspondence to Dr. Yoshizo Matsuka, Department of Stomatognathic Function and Occlusal Reconstruction, Graduate School of Biomedical Sciences, Tokushima University, 3-18-15 Kuramoto-cho, Tokushima 770-8504, Japan; +81-886-33-7350; matsuka@tokushima-u.ac.jp.

Keywords: dental education, assessment, peer evaluation, team-based learning, prosthodontics, teaching effectiveness, educational methodology

Submitted for publication 12/2/16; accepted 7/5/17
doi: 10.21815/JDE.017.106

Team-based learning (TBL) is an instructional strategy of active learning. This learning method uses group activities, helps students perceive for themselves their responsibility in the group, and encourages active discussion about the study subjects.[1,2] The united effort of all students builds confidence in each other and develops mature learning processes. The effects of TBL classes in clinical dental education have been reported at many institutions.[3-9] We adopted TBL classes for a fixed prosthodontics course for undergraduate dental students at Tokushima University beginning in 2013 and reported the effectiveness of those classes.[7-9] In those studies, we reported that the TBL classes promoted study preparation and an active student attitude and were potentially a more effective method of teaching than the usual style of classes.

Evaluation of the learning performance in the TBL class was generally done with an individual test conducted at the beginning of class, group test, group activity, and peer evaluation. Peer evaluation was performed by mutual evaluation of the students

who belonged to the same group, and the contribution level of each student for group work was scored. Since most of the students were untrained in scoring their colleagues, they often experienced psychological stress. For this reason, some studies reported difficulty in peer evaluation for TBL classes.[10,11]

The aim of this study was to determine the validity of peer evaluation TBL classes in dental education in comparison with the term-end examination records and TBL class scores. To assess the validity of peer evaluation for TBL classes, we conducted a retrospective cohort study investigating the term-end examination record and TBL class scores of students who took TBL classes for the fixed prosthodontic courses. The term-end examination records, individual test score, group score during TBL classes, and peer evaluation were also compared.

Materials and Methods

This study was approved by the Research Ethics Committee of Tokushima University Hospital (No. 2720). All 256 third- and fourth-year dental students at the School of Dentistry of Tokushima University participated.

These students took the fixed prosthodontics course from the first semester of 2013 to the second semester of 2015. This course was for third- and fourth-year students, and each course involved 15 classes that included six to eight TBL classes. Each class was 60 minutes long, and an orientation about TBL was held at the beginning of the TBL classes. Six courses concerning fixed prosthodontics were held during this investigation. Since the fixed prosthodontics courses were held from the second semester of the third year to the first semester of the fourth year, 77 students (out of 154) took these courses in both semesters. Inclusion criteria of the participants were to maintain a 100% attendance of the classes during the study period. Since there were no dropouts from those courses, no student was excluded from the investigation.

One week before the TBL class, study materials for preparation and adequate homework were given to the students. At the beginning of the TBL class, all students took an individual readiness assurance test (IRAT) with multiple-choice questions that covered preparation material. Then, the class was divided into small groups of six or seven members who were instructed to participate in active group discussion to answer the group readiness assurance test (GRAT), which has the same questions as the IRAT. The difficulty of those questions was adjusted for a basic level but enough to stimulate group discussion. If a student asked an insightful question during the GRAT feedback, it was rewarded by the addition of group appeal points to all group members. After groups received feedback from the instructor about these questions, practical questions testing their ability to use knowledge in clinical situations were solicited for group assignment projects (GAPs).

The group score of the TBL class was counted with the sum of the GRAT, GAP, and group appeal point, so that all members in the same group had the same group score. Finally, all students were asked to perform peer evaluation by grading the class activity of other members in their group.

Final results of students in the fixed prosthodontics course were scored by adding 50% of term-end examination results and 50% of TBL class scores. This grading manner was announced to all students at the beginning of the course. More detailed description of the TBL class was presented in our previous reports.[7-9]

Term-end examination of the course was performed with multiple-choice questions. The examination record covered by the subjects of the TBL classes was selected, and the correct answer ratio of those questions for each student was determined. TBL class score was assessed with the individual score of IRAT, group score, and peer evaluation of the student. Those scores were standardized by converting to percentage the full score except for peer evaluation. Since peer evaluation was performed by mutual marking of the group members, if the number of members was different among groups, the total score of the group members would not be the same for all groups. To avoid such discrepancy between groups, the average score of the peer evaluation in each group in each class was adjusted to 60%.

To verify the validity of the peer evaluation, the following statistical analyses were conducted. The O'Brien test was performed to examine the homogeneity of variance in all the class scores (peer evaluation, individual score, group score, and term-end examination score) during the total investigation period. Correlations among all the class scores were examined with Spearman's rank correlation coefficients to evaluate interrelationships. Multiple regression analysis that used total score for peer evaluation as the dependent variable and other examination and class scores as independent variables was also performed. This analysis was adopted to evaluate associations between the peer evaluation score and other class and examination scores.

Since individual score exhibited higher effect of peer evaluation in the analysis, Spearman's rank correlation coefficients between those two scores in each of the six courses were calculated to evaluate the steadiness of the effect. Multiple regression analysis for peer evaluation for the six courses was also performed to evaluate the association of each of those scores in each course. Longitudinal data analysis with repeated measure analysis of variance (ANOVA) and correlation analysis were also performed for the scores of 77 students who took the fixed prosthodontics courses two times in different semesters to investigate the relationship of the class scores during two courses. JMP8 (SAS Institute Japan Ltd., Tokyo, Japan) was used for statistical analysis.

Results

The total average scores of the students through the whole investigation period were 59.6±8.9% for peer evaluation, 63.1±13.7% for individual score, 77.8±9.9% for group score, and 69.8±15.0% for term-end examination. Since the results of the O'Brien test denied homoscedasticity among these scores ($p<0.0001$), the Spearman's rank correlation coefficients, which is a non-parametric test, was adopted for correlation analysis.

Correlation of the total scores through the whole investigation period is shown in Figure 1 and Table 1. Significant positive correlations were found for peer evaluation vs. term-end exam, peer evaluation vs. individual score, individual score vs. term-end exam, and individual score vs. group score (n=256, Spearman's rank correlation coefficients). However, group score was not correlated with peer evaluation and term-end exam. Table 2 shows the results of the multiple regression analysis for total peer evaluation score. This model predicts the dependent variable (peer evaluation score) from individual, group, and term-end exam scores. These results showed that the peer evaluation score was significantly associated with individual score ($p<0.0001$) and term-end exam score ($p=0.0032$).

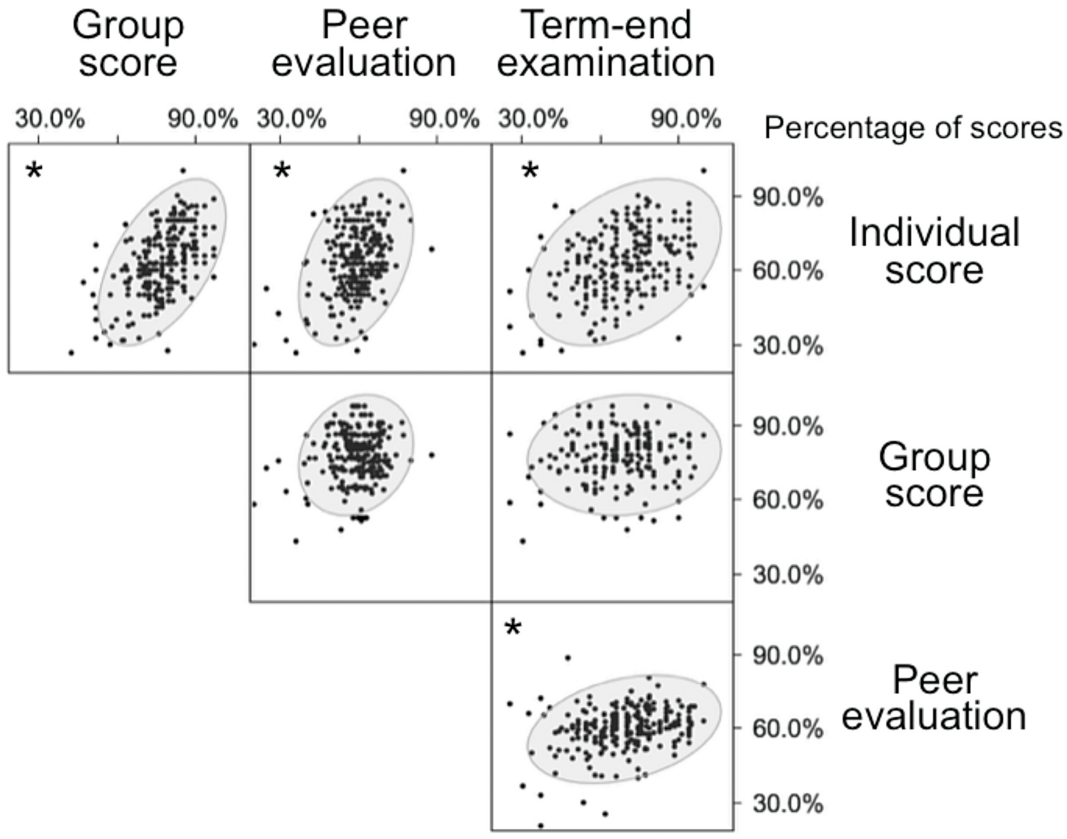

Figure 1. Correlation of total scores of team-based learning (TBL) classes and term-end examination (n=256)

*$p<0.0001$, Spearman's rank correlation

Table 1. Correlation coefficients between total scores of team-based learning (TBL) classes and term-end examination (n=256)

Variable	Individual Score	Group Score	Peer Evaluation	Term-End Exam
Individual score	1.0000	0.5257	0.2900	0.3516
Group score	*	1.0000	0.0449	0.0365
Peer evaluation	*		1.0000	0.2698
Term-end exam	*		*	1.0000

*p<0.0001, Spearman's rank correlation

Table 2. Multiple regression analysis for total peer evaluation score

Independent Variable	Regression Coefficients	Standard Error	p-value
Individual score	21.59	0.0504	<0.0001
Group score	-0.0045	0.0631	0.9427
Term-end exam	0.1127	0.0379	0.0032

n=256; multiple R2=0.187, p<0.0001

Figure 2 shows semester correlations between peer evaluation and individual score in six courses during the study. Significant positive correlations were found except for the second semester of 2014 and the second semester of 2015. Results of multiple regression analyses for peer evaluation of the six courses are shown in Table 3. Association between peer evaluation, other classes, and examination scores were found for all courses; but the association of each independent variable varied widely depending on the course.

Table 4 shows the result of the longitudinal data analysis for the 77 students who took the fixed prosthodontics course two times in different semesters.

Table 3. Multiple regression analysis for total peer evaluation scores through six courses

Independent Variable	Regression Coefficients	Standard Error	p-value
2013 1st; n=36; multiple R^2=0.4995; p<0.0001			
Individual score	0.2313	0.0982	0.0249
Group score	0.2468	0.1534	0.1175
Term-end exam	0.2402	0.0697	0.0016
2013 2nd; n=41; multiple R^2=0.2476; p=0.0156			
Individual score	0.1993	0.0748	0.0114
Group score	-0.0335	0.1066	0.7551
Term-end exam	0.0780	0.0648	0.2362
2014 1st; n=45; multiple R^2=0.6274; p<0.0001			
Individual score	0.5529	0.1053	<0.0001
Group score	0.4914	0.1902	0.0134
Term-end exam	0.0619	0.0967	0.5259
2014 2nd; n=41; multiple R^2=0.5749; p<0.0001			
Individual score	0.0228	0.0637	0.7221
Group score	0.8559	0.1408	<0.0001
Term-end exam	0.0106	0.0603	0.8612
2015 1st; n=43; multiple R^2=0.4612; p<0.0001			
Individual score	0.1272	0.1383	0.3631
Group score	0.6482	0.1869	0.0013
Term-end exam	0.1323	0.0985	0.1869
2015 2nd; n=40; multiple R^2=0.3022; p=0.0044			
Individual score	0.5641	0.1747	0.0027
Group score	-0.2533	0.2135	0.2430
Term-end exam	0.1721	0.1798	0.3448

Repeated ANOVA showed a significant main effect between the class scores but not between the semesters. Significant interaction effect was seen between the class scores and the semesters. Correlation analysis between two semesters' class scores showed significant correlation for all class scores (individual score p<0.0001; group score p<0.0001; peer evaluation p=0.0133; and term-end exam p=0.0200).

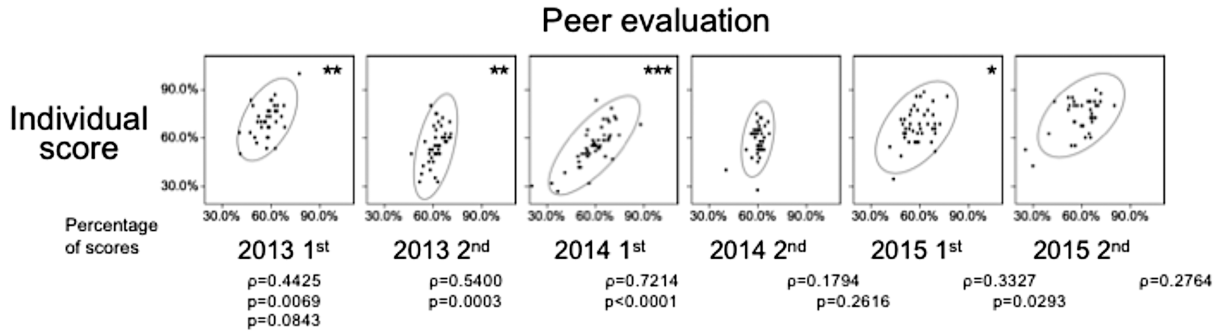

Figure 2. Correlation of peer evaluation and individual scores through six cours

*p<0.05, **p<0.01, ***p<0.0001, Spearman's rank correlation

Table 4. Repeated measure analysis of variance for class scores of students who took fixed prosthodontics courses two times in the semester (N=77)

Factor	df	Mean Square	Standard Error	F value	p-value
Class score	3	3.0758	0.00728	94.2214	<0.0001
Semester	1	0.0156	0.00420	1.4337	0.2316
Class score x semester	3	1.0029	0.00728	30.7222	<0.0001

Note: Class scores were TBL scores and term-end exam records in each semester.

Discussion

In our university, the fixed prosthodontics course is offered for third- and fourth-year students. This course is one of the first occasions for students to learn clinical knowledge of dental technologies. To encourage an active attitude for learning at an early stage of clinical education, we adopted TBL for these courses. The average correct answer ratio of the term-end exam in all courses was about 70%. Since the qualifying score of the exam in those courses was 60%, the total learning performance of the students in TBL classes could be assessed as satisfactory. More detailed information on the TBL classes in those courses and its learning effect can be seen in our previous reports.[7,9]

Term-End Exam and TBL Class Scores

Although most of the exam and class scores exhibited significant correlations, group score did not correlate with peer evaluation and term-end exam (Figure 1, Table 1). Since those scores were adopted to evaluate the learning performance of each student, these scores naturally have a positive correlation. However, group members who belonged to the same group had the same group score; and since peer evaluation was based on the ranking of those members, these two scores did not exhibit correlation. Likewise, because group score was not associated with individual learning performance, this score was not correlated with the result of the term-end exam.

Multiple regression analysis exhibited higher association of the individual score for peer evaluation (Table 2). This finding indicated that the students who marked higher individual scores obtained higher evaluations from group members. The term-end exam score also showed association with peer evaluation score since this score should reflect the general learning ability of the students.

During the investigation period, about half of the students took the fixed prosthodontics course two times in different semesters. Since those scores were graded from same students, the class scores between those semesters exhibited significant correlation. Results of repeated ANOVA exhibited the main effect of the class scores for those students, indicating essential differences between class scores. Since the group score was marked from the same questions preceding the IRAT, this score was naturally higher than the individual score. On the other hand, the average score of the peer evaluation in each group was adjusted to 60% to avoid group discrepancy, so those TBL class scores could have a substantial difference. Such main effect was not observed between the course semesters in those scores, which indicated a reproducing tendency of the scores in sequential semesters. Significant interaction effect with the class score and the semesters indicated that transitions of the class scores along those two semesters were not similar among the class scores. The average class score in those semesters increased with individual scores (57.4±10.6% to 62.1±12.6%) and group scores (74.2±8.1% to 80.6±11.1%), did not change with peer evaluation (61.2±4.9% to 60.1±9.4%), and decreased with term-end exam (74.6±10.7% to 60.5±13.6%). Therefore, we think that the practice for TBL classes for the students could increase the individual and group scores in the latter semester.

Availability of Peer Evaluation

Since peer evaluation was performed with mutual evaluation of group members who belonged to the same group, this score should exhibit positive correlation with individual scores in the same classroom. Even if total scores during the whole investigation period exhibited a significant positive correlation, no significant correlations were observed with the courses that were held in the second semester of 2014 and 2015 (Figure 2). Results of multiple regression analysis in each course also exhibited inconsequential association of individual score for peer evaluation at the second semester of 2014 (Table 3).

These findings indicated that preparatory study and accompanied group contributions of the student were not associated with evaluation from group members with those courses. Some educators have reported the difficulty of incorporating peer evaluation for their courses.[10,11] Those studies indicated that students may have dissatisfaction with peer evaluation because it interferes with the friendship between group members when scoring each other's performance. In our study, we found a slight standard deviation with the peer evaluation score in the second semester course of 2014 (Figure 2). In this semester, there was a group in which every member of the group had exactly the same peer evaluation score eventually. It could be suspected that some kind of arrangement was made among group members to equalize peer scores in this semester. Such an arrangement undermines the validity of peer evaluation.

Even if it was the first experience for all of these students to participate in TBL classes, the average scores of the term-end exam were relatively high. We therefore conclude that the adaptation of the students for TBL class was satisfactory. Because the total score of peer evaluation showed a positive correlation with other exam and class scores, it could be said that peer evaluation was a valid indicator of the learning performance of the student. However, peer evaluation for two of the six courses did not show positive correlation with individual score. Thus, proper understanding of the student for peer system might not be enough in those semesters. To confirm the validity of peer evaluation in all courses, we should have taught students the significance of scoring team members' performance.[12] At the beginning of the TBL classes, the instruction of the peer evaluation was given to the student. During the instruction, students were asked to grade group members based on preparatory study level, contribution for group activity, consideration for group members, and flexible thinking. However, the students may not have had the skill to provide appropriate evaluation, so improvement of the scoring criteria and more detailed instruction will be required to confirm the effectiveness of the peer evaluation in all TBL classes.

Since this study took place in a single institution, its results may not be generalizable to students in other dental schools. In the future, we hope to demonstrate validity and efficiency of TBL classes for dental education with improved peer system.

Conclusion

TBL classes were performed in six fixed prosthodontics courses with 256 dental students over three years. Since the total score of peer evaluation in those courses showed positive correlation with term-end exam and other TBL class scores, we think that the mutual evaluation score of the students was a valid indicator for students' learning performance. To verify the effectiveness of the peer evaluation, students' understanding of the significance of mutual evaluation, clear criteria, and detailed instruction for appropriate evaluation are required.

REFERENCES

1. Michaelsen LK, Knight BA, Fink LD. Team-based learning: a transformative use of small groups in college teaching. Sterling, VA: Stylus Publishing, 2004.
2. Michaelsen LK, Parmelee DX, Mcmahon KK, Levine RE. Team-based learning for health professions education: a guide to using small groups for improving learning. Sterling, VA: Stylus Publishing, 2007.
3. Parmelee S, Michaelsen LK, Cook S, Hudes PD. Team-based leaning: a practical guide. AMEE guide no. 65. Med Teach 2012;34:e275-87.
4. Pileggi R, O'Neill PN. Team-based learning using an audience response system: an innovative method of teaching diagnosis to undergraduate dental students. J Dent Educ 2008;72(10):1182-8.
5. Kumar V, Gadbury-Amyot CC. A case-based and team-based learning model in oral and maxillofacial radiology. J Dent Educ 2012;76(3):330-7.
6. Haj-Ali R, Al Quran F. Team-based learning in a preclinical removable denture prosthesis module in a United Arab Emirates dental school. J Dent Educ 2013;77(3):351-7.
7. Takeuchi H, Omoto K, Okura K, et al. Effects of team-based learning on fixed prosthodontic education in a Japanese school of dentistry. J Dent Educ 2015;79(4):417-23.
8. Omoto K, Okura K, Tajima T, et al. Introduction and effects of team-based learning on prosthodontics education. J Oral Health Biosci 2015;27:81-7.
9. Nishigawa K, Omoto K, Hayama R, et al. Comparison between flipped classroom and team-based learning in fixed prosthodontic education. J Prosthodont Res 2017;61:217-22.
10. Van Rosendaal GM, Jennett PA. Resistance to peer evaluations in an internal medicine residency. Acad Med 1992;67:63.
11. Farland MZ, Sicat BL, Franks AS, et al. Best practices for implementing team-based learning in pharmacy education. Am J Pharm Educ 2013;77(8):Article 177.
12. Levine RE, Kelly A, Karakoc T, Haidet P. Peer evaluations in a clinical clerkship: students' attitudes, experiences, and correlations with traditional assessments. Acad Psychiatry 2007;31:19-24.