

Reliability and validity of peer assessment at tutorless PBL classes

Aung, H.L., Tashiro, Y., Morio, K., Hori, H.

Abstract

Introduction: Learning behaviors in PBL and team-based learning (TBL) at Mie University Faculty of Medicine have been measured with peer assessments in addition to tutor assessments. The change in class components posed a question on reliability of peer assessments at tutorless sessions in TBL.

Methods: Correlations between peer and tutor assessments in the PBL program comprised of tutor and tutorless sessions were analyzed with data from 377 students. The three domains for student's performance, "self-learning in preparation", "active participation in group discussion" and "cooperative attitude at group work", were collected at 4 UNITS with the formal PBL and at 4 UNITS with the mixture of PBL and TBL.

Results: No statistically significant difference between peer and tutor assessments was found in all three domains in units with the mixture of PBL and TBL. The correlation analysis between peer and tutor assessments in units with the mixture of PBL and TBL demonstrated strong positive correlations in self-learning in preparation ($r=0.62$) and in active participation in group discussion ($r= 0.63$), and a weak positive correlation in cooperative attitude at group works ($r=0.35$).

Discussion: This study revealed that students evaluated their group-peers at the units comprised of both tutor and tutorless sessions similarly to those comprised of only tutor sessions. Peer assessments were well correlated with tutors' assessment at the units comprised of tutor and tutorless session. Our study suggests that TBL might be expanded in the situation where teaching resource is limited.

Keywords: problem-based learning, tutorials, team-based learning, medical education, peer assessment, tutor assessment.

Introduction

Evaluation of learning outcome in medical education should be multifaceted, including knowledge, technique and attitude. Students' learning behaviours in a group activity reflect their diligence, independence and cooperativeness, but the evaluation is more difficult than measuring knowledge by a paper test. A classical method of assessment is based on the observation by a teacher of which the objectivity is not secured sufficiently.

After Barrows HS pioneered a problem-based learning (PBL) program at McMaster University Medical School in the late 1960s (Barrows and Tamblyn, 1980), the innovative learning strategy has been implemented at medical schools all over the world. In Japan, small group collaborative learning based on the concept of PBL (PBL-tutorials) were introduced in a few medical schools in the 1990s (Kozu, 2006) and then have extended to other schools. While the educational effects of PBL have widely been recognized, the strategy has faced some challenges in practice. Those include the quality assurance of teachers' tutoring, securing teaching resources and improving facilities to assist students' self-directed and

Mie University Graduate School of Medicine, Japan.

Corresponding author: Prof Yasura Tashiro
ytashiro@doc.medic.mie-u.ac.jp

DOI: <http://doi.org/10.4038/seajme.v15i2.351>



© SEAJME. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

small group collaborative learning (Wood, 2003). To solve the issues, modified PBL classes have been tried at many medical schools. The examples are a combination of PBL-tutorials and ordinary lectures in a large classroom, and team-based learning in a single classroom (TBL). TBL is an instructional strategy developed at a business school in the early 1990s by Michaelsen L (Michaelsen & Sweet, 2008). Recently, PBL tutorials at some medical schools are replaced or combined with TBL. TBL is largely different from PBL tutorials and other small group approaches. TBL does not need a tutor in respective groups like PBL tutorials.

Mie university, a middle-sized university located in a local city in Japan, introduced PBL-tutorials into pre-clinical curriculum for 3rd to 4th year students in 1995. The classes were composed of self-learning in preparation for the subsequent group session, small group collaborative learning facilitated by a tutor and case-related lectures. The students proceeded their learning with a clinical case-scenario provided every week. The program had been conducted with 13 UNITS during the period of 18 months for the first 20 years. But the shortage of teaching resources became evident thereafter because the number of students participating in this program increased from 100 to 125 along with the expansion of admission quota. Then, the term of the PBL tutorial program was shortened from 18 to 14 months. Accordingly, the number of units was also reduced from 15 to 10 UNITS. Furthermore, half of PBL-tutorials in 4 out of 8 UNITS were replaced with the mixture of PBL tutorials and TBL. In TBL sessions, one to two teachers managed the progress of discussion at all 16 groups in a single classroom. In another 2 UNITS, TBL was introduced throughout a UNIT, mainly focusing on specialized fields. Thus, the PBL program at Mie University Faculty of Medicine is partially comprised of TBL sessions, where student groups have discussion without tutor's facilitation.

Self-reflection and self-evaluation are the integral objectives of the PBL curriculum. Many previous studies showed that a self-assessment alone is not reliable, but a peer

assessment is more reliable and valid (Sullivan et al., 1999; Papinczak et al., 2007; Kamp et al., 2011; Das et al., 1998). In the PBL tutorial program at Mie University Faculty of Medicine, a student's performance is evaluated at each UNIT with the score of paper tests, the rating of reports on self-learning, and an attendance rate at PBL/TBL sessions and learning behaviors in PBL/TBL sessions. The learning behaviors are evaluated with peer and tutors' assessments throughout the PBL/TBL sessions.

Our recent studies indicated that peer assessments well correlate with tutor assessment (Tashiro et al., 2014; Tashiro et al., 2015). However, the studies did not separately deal with the two different type of group sessions, exclusive PBL tutorials and combinatory sessions of PBL and TBL. Therefore, we planned this study to assess the validity of peer assessments at UNITS with tutorless TBL sessions. In this study, we analyzed the data collected from students and tutors who joined these programs in recent three years and discussed the correlations between peer and tutor assessments in UNITS with and without TBL.

Methods

Subjects

A total of 377 students at Mie University Faculty of Medicine participated in pre-clinical classes operated with the concept of PBL in their 3rd to 4th year during the period between 2014 and 2017. The number of students by the academic year were 124 in 2014-2015, 127 in 2015-2016 and 126 in 2016-2017. The data of these students were applied to the analysis in this study.

This study was approved by the institutional review board of Mie University Faculty of Medicine (the reference number of U2018-017). Because this study was a retrospective observational study with data collected for educational activities, the students were notified of this study through the university official website and freely allowed to notice their refusal to participate in the study unless their data were desirable to be used.

Contents of PBL tutorial program

The PBL tutorial program contains 10 UNITS constructed according to the organ systems (Table 1). A single UNIT is carried out for 4 to 6 weeks. The PBL tutorials are operated along a case-scenario. The UNITS are categorized into

three types by the combination of PBL tutorials and TBL (Fig. 1): 4 units with the formal PBL tutorials (UNIT 2, 3, 6, 8), 4 unit with a mixture of PBL tutorials and TBL (UNIT 1, 4, 5, 7), and 2 units with exclusive TBL (UNIT 9, 10). A formal PBL tutorial has 2 sessions per week while TBL has 1 session a week.

Table 1: UNITS composition of pre-clinical classes for 3rd to 4th year students

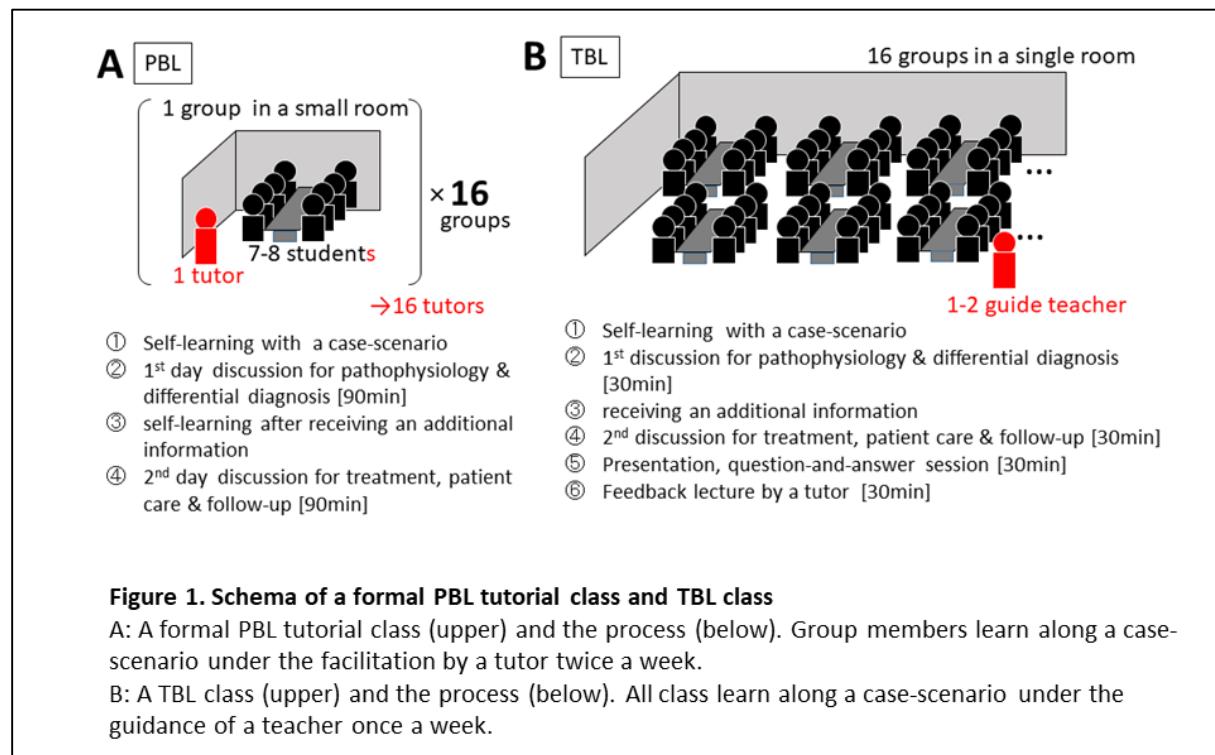
UNIT	Systems	Total weeks	No. of case-scenario	PBL tutorials	TBL
1	Infection & Immunology	5	3	6	1
2	Cardiovascular System/Hematology	6	4	8	0
3	Respiratory System/Endocrine & Metabolism	6	4	8	0
4	Kidney & Urinary Tract	4	2	4	1
5	Neurology	6	3	6	1
6	Digestive System	6	3	6	0
7	Breast & Female Genital System	4	2	4	2
8	Pediatric & Geriatric Medicine	5	3	6	0
9	Emergency, ENT, Skin, Eye	5	4	0	4
10	Orthopedic, Maxillofacial, Anesthesiology, Psychiatric Medicine, Family Medicine	6	4	0	4

About 125 students in 3rd or 4th year are allocated into 16 small groups every 2 units and engaged in collaborative learning in a group consisting of 7 to 8 students (Fig. 1). The group members are not replaced throughout two consecutive units. From the UNIT 1 to 8, the same group members participated in the consecutive two units, one with the formal PBL tutorials and another with a mixture of PBL tutorials and TBL. In the UNIT 9 and 10, the same group members participated in the two units with exclusive TBL.

Seven to 8 students and a tutor meet in a small room and have a PBL tutorial session for approximately 90 min (Fig. 1A). In the sessions, students discuss patient's signs and symptoms, pick up learning topics and share their knowledge obtained from self-learning in preparation. Through the group discussion, they learn problem distillation, problem-solving, clinical reasoning. Then, they tried to understand pathophysiology, treatment, and patient care along the case-scenarios

containing the records of medical interview, physical examination, laboratory tests, treatment/surgical intervention and follow-up. The clinical information is provided dividedly according to the progress of learners.

A TBL session typically takes 2 hours long. The whole class, a guide teacher and a few supporting staff members meet at a single classroom (Fig. 1B), but students learn along a case-scenario similarly to the formal PBL tutorials. For the first 30 minutes, students discuss the case and share the products of their self-learning with group peers. For the following 30 minutes, students have further discussion on information additionally supplied on the spot. Then members of two groups appointed by the guide teacher present the outcome of group discussions to share the learning achievements in the process of PBL. Finally, the guide teacher give feedback comments to the presenters and then delivered a short lecture on the case-scenario.



Evaluation for student in PBL tutorial program

Students were evaluated every single UNIT with the score of a paper test (50%), the rating of reports on self-learning (20%), the attendance rate at PBL/TBL sessions (15%) and learning behaviours at PBL/TBL sessions (15%). The learning behaviours were scored as a sum of a tutor assessment and an average of peer assessments at the PBL/TBL sessions. At UNITS with the mixture of PBL tutorials and TBL, the tutor assessment was performed only for PBL tutorial sessions and the peer assessment was done for both PBL tutorial and TBL sessions. At UNITS with TBL only, a student is evaluated only by all group peers at each single UNIT. The assessment is carried out with a 10-point Likert scale in three domains of learning behaviours in group activity: 1) self-learning in preparation, 2) active participation in group discussion, 3) cooperative attitude at group works. The peer assessment includes self-assessment of the evaluator him/herself. Students submit the completed evaluation forms onto the website “Mie University Moodle”, a kind of learning management system. At the guidance of the PBL tutorials, the program director explains the significance and criteria of peer assessments, requesting fair and impartial evaluation without any collusion among the

students. A tutor attending PBL tutorials also evaluates all students in a group with the similar evaluation format.

Data collection and analysis

Students' peer evaluation forms and tutor evaluation forms were centrally collected, aggregated, analysed and archived at the Centre for Medical Education. The data from 16 groups at the UNITS 1 to 8 were collected for 3 years and applied to this study after anonymization. Data from 2 units with exclusive TBL were excluded from the analysis because only peer assessments were collected in those UNITS. Therefore, in this study, the data from students and tutors were analyzed in 4 UNITS with the formal PBL tutorials only and 4 UNITS with the mixture of PBL tutorials and TBL.

The mean values were analyzed between peer and tutor assessments for students by the paired-sample *t*-test. The correlation between the two types of assessments was analyzed with the Pearson's correlation coefficient for the normally distributed data and the Spearman's correlation coefficient for the non-normally distributed data. The data analysis was computed with a statistical software, SPSS Version 24.0 for macOS.

Results

Data Validation

The normality test with data from 377 subjects revealed that there were non-normal distributions in all three variables for peer and tutor assessments. The frequency histogram for each variable displayed a deviation from normal distribution curve. The *P*-values of Shapiro-Wilk proved to be <0.001 in all variables for peer and tutor assessments. The boxplot uncovered 4 extreme outliers and the other 6 minimum outliers. Because 4 extreme outliers earned zero points automatically in some units with the result of disqualification such as non-attendance at classes or failure to carry out peer evaluation in time, we omitted those students from the analysis.

The removal of the 4 extreme outliers made the distribution pattern of 2 out of 3 domains to be normal (self-learning in preparation and active participation in group discussion). The other 6 minimum outliers were decided not to be withdrawn because we could not find any proper reason to omit from this correlation study. In addition, the exclusion proved to exhibit a minimal effect for the distributions of normality test. Finally, we analyzed the data from 124 students in 2014-2015, 125 in 2015-2016, and 124 in 2016-2017. The total number of students included were 373.

Comparison of scores in three domains of assessment between peer and tutor assessments

The mean value of assessment scores for student's achievement was compared between peer and tutor assessments at three domains, self-learning in preparation, active participation in group discussion, and cooperative attitude at group works. The analysis with data from all 8 units including two different types of classes found that peer assessment showed the significantly higher mean values than tutor assessment in all three domains ($t_{372}=3.58$, $p<0.001$ in self-learning in preparation; $t_{372}=3.14$, $p=0.002$ in active participation in group discussion; $t_{372}=2.14$, $p=0.033$ in cooperative attitude at group works) (Table 2).

The analysis in 4 units with the formal PBL tutorials demonstrated the similar trends to that in all 8 units ($t_{372}=4.82$, $p<0.001$ in self-learning in preparation; $t_{372}=4.04$, $p<0.001$ in active participation in group discussion; $t_{372}=3.16$, $p=0.002$ in cooperative attitude at group works) (Table 2). However, no statistically significant difference between peer and tutor assessments was found in all three domains in 4 units with the $p=0.108$ in self-learning in preparation; $t_{372}=1.26$, $p=0.21$ in active participation in group discussion; $t_{372}=0.61$, $p=0.541$ in cooperative attitude at group works) (Table 2).

Table 2: Mean scores of peer and tutor assessments for three assessment domains

		Assessment domain	Peer evaluation	Tutor evaluation	<i>P</i> -value
			Mean (SD)		
4 UNITS with the formal PBL tutorials		1	8.43 (0.56)	8.29 (0.71)	<0.001
		2	8.39 (0.56)	8.28 (0.70)	<0.001
		3	8.49 (0.47)	8.41 (0.55)	0.002
4 UNITS with PBL tutorials and TBL		1	8.37 (0.56)	8.32 (0.74)	0.108
		2	8.34 (0.55)	8.29 (0.76)	0.21
		3	8.43 (0.45)	8.41 (0.61)	0.541
All 8 UNIT		1	8.40 (0.56)	8.31 (0.66)	<0.001
		2	8.36 (0.55)	8.28 (0.66)	0.002
		3	8.46 (0.45)	8.41 (0.50)	0.033

Correlation between student and tutor assessments

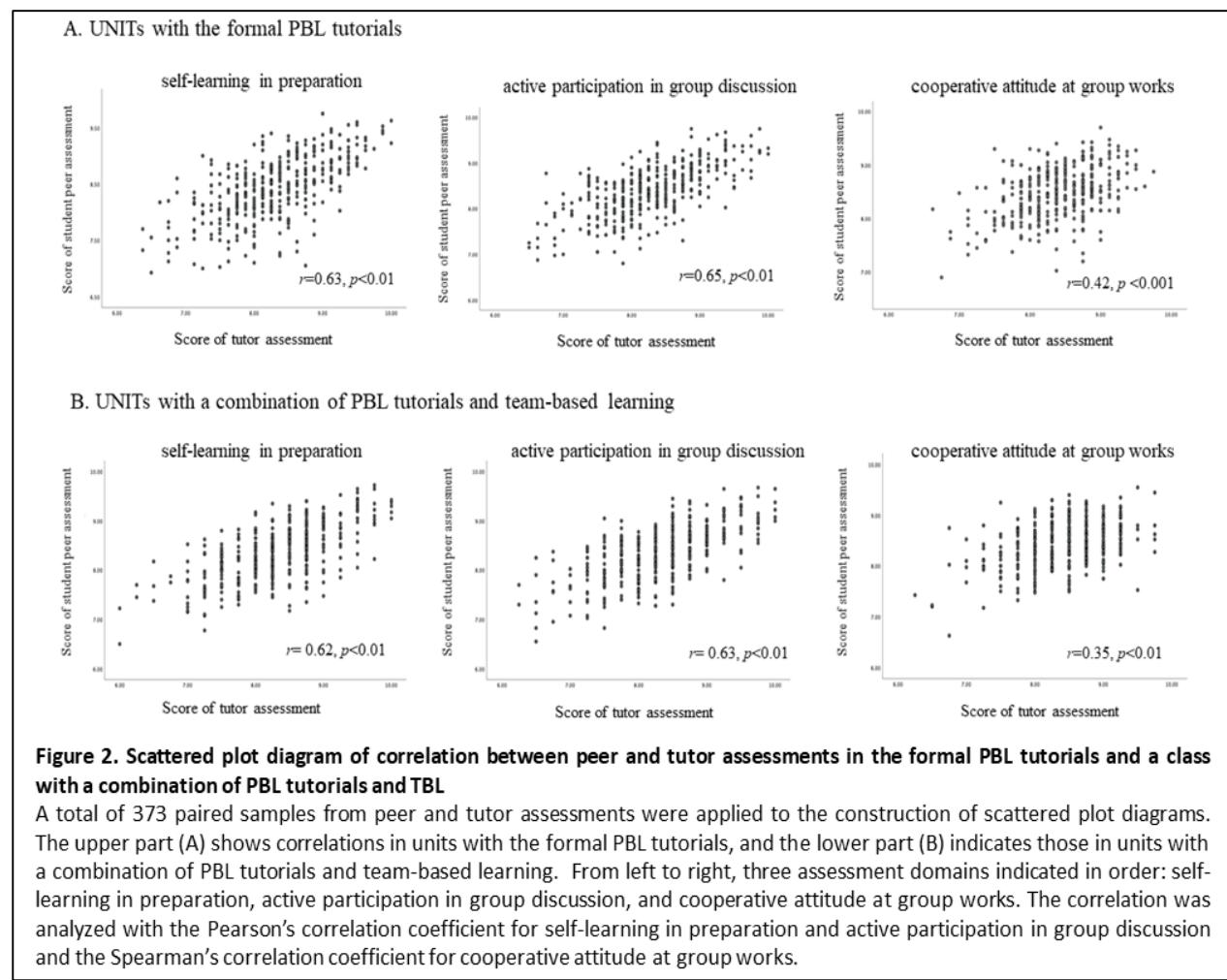
First, the correlation between peer and tutor assessments was analyzed in all 8 units. Strong positive correlations between the two assessments were noted in self-learning in preparation ($r= 0.71, p <0.01$), and active participation in group discussion ($r=0.73, p<0.01$) (Supplementary table-Table S1). A moderate positive correlation was found in cooperative attitude at group works ($r=0.45, p<0.01$).

The subgroup analysis by the academic year revealed strong positive correlations in all three domains in 2015-2016 academic year ($r=0.80, p<0.01$ in self-learning in preparation; $r=0.85, p<0.01$, active participation in group discussion;

$r=0.75, p<0.01$ in cooperative attitude at group works).

Second, we analyzed the correlation in 4 units with the formal PBL only. Strong positive correlations were observed in self-learning in preparation ($r=0.63, p<0.01$) and active participation in group discussion ($r=0.65, p<0.01$) while a moderate positive correlation was noted in cooperative attitude at group works ($r=0.42, p<0.01$) (Fig. 2).

Third, we analyzed the correlation in 4 units with the mixture of PBL tutorials and TBL. Strong positive correlations in self-learning in preparation ($r=0.62, p<0.01$) and in active participation in group discussion ($r= 0.63, p<0.01$) and a weak positive correlation in cooperative attitude at group works ($r=0.35, p<0.01$) were found.



Discussion

This study showed that, in general, scores in peer assessments were higher than those in tutor assessments, while a positive correlation was observed between the two types. The relatively higher scoring in peer assessments is in accord with previous studies which warned the validity of self and peer assessment in medical education (Machado *et al.*, 2008; Reiter *et al.*, 2002; Evans *et al.*, 2007; Basheti *et al.*, 2010). Such trends in medical education could result from student's self-confidence, self-efficacy, appreciation to peer, positive feedback to peer and less understanding of learning-objectives (Gukas *et al.*, 2008). However, we note the positive correlation between peer and tutor assessments, especially for two variables, "self-learning in preparation" and "active participation in group discussion". This suggests that students might evaluate themselves fairly aligned with the tutor's assessment. Topping (2009) reported that peer assessors with less assessment skill come to produce an assessment with equal reliability and validity to that of teachers with accumulating experiences. The practice of peer assessment could improve their skills at evaluation in a year of the educational program. In addition, 1st year students of all faculties at the university are required to take an introductory course of PBL in the liberal arts education, in which students participate in group works and practice peer assessments. The experience of peer assessments in early school year might contribute to improvement in their assessment skills.

For an assessment variable, "cooperative attitude at group works", the correlation was moderate. We postulate that this may be due to the difference in an evaluation criterion for "cooperative attitude" between students and teachers. A tutor can observe who is good at cooperation in the tutorials whereas students usually know who is acting cooperatively in daily campus life. We need to investigate the background of this result with another study.

Although the results discussed above were found in all 8 UNITS and 4 UNITS with the formal PBL tutorials, no significant difference between

tutor and peer assessments was found at the units with the mixture of PBL tutorials and TBL for all three variables. In those units, the score of peer assessments got closer to that of tutor assessments. One can hypothesize that group works without tutor's facilitation might be not active, and consequently students might have negative impressions for "active participation in group discussion" and "cooperative attitude at group works" of the group members. We expect that another variable, "self-learning in preparation", might not be affected by the attendance/absence of a tutor at TBL, but the result at the units with the mixture of PBL tutorials and TBL was not similar to that at the UNITS with the formal PBL tutorials. The tutor's attendance might motivate their self-learning at PBL tutorials.

A multilateral assessment is ideal to assess the performance of students in medical education. Peer assessment could be one of the useful methods in the multilateral assessment strategy. To assess the reliability and validity of peer evaluation in the units with tutorless sessions, we conducted this study. In the units including both tutor and tutorless sessions, a tutor evaluated students only at PBL tutorial sessions, and each student assessed his/her peers at both PBL tutorial and TBL sessions. The results demonstrated that there was no significant difference in assessment scores between peer and tutor assessments in the units with the combination, and the scores were well correlated between the two different types of assessors. This result suggests that peer assessments for the units comprised of both tutor and tutorless sessions can evaluate student's learning behaviours similarly to those comprised of only tutor sessions. Furthermore, we postulate that, in TBL environments, students may have a chance to evaluate their group-peers by contrasting with students in other groups who learn together in a single classroom. The inter-group activity might contribute to the fairer assessment for their group-peers.

The present study has some limitations. Because this study was planned in classes routinely conducted in the regular curriculum, the data available for the analysis were limited.

The peer evaluation was not done solely for tutorless sessions and no evaluation was performed by a teacher at TBL. Additionally, the correlation of peer assessments was analyzed only with tutor assessments, not with others such as the rating of reports or the score of a paper test. However, the rating of reports and the score of a paper test partially reflect the learning outcome of student's self-learning as well as that of their group works. Student's performance in clinical rotations following this program should be traced carefully and the results should be reflected to improvements of this program.

In conclusion, Mie University Faculty of Medicine introduced TBL into a pre-clinical PBL program to reduce the workload of teachers. The classes were organized in combination of PBL tutorials and TBL. Although teachers had some concerns about validity and reliability of peer assessments at tutorless classes of TBL, students performed peer assessments well correlated with tutor assessments. Our study suggests that TBL might be an option of PBL in the situation where teaching resource is limited.

Acknowledgments: This study was supported by Mie University Faculty of Medicine

Authors' contribution: HLA, HH, and YT conceived and designed the study. HLA contributed to statistical analysis, its interpretation and drafted the manuscript. HH supervised the research team and the major contributor for receiving ethical approval. YT helped for study design, raw data collection, and data interpretation. KM participated in statistical analyses and data interpretation and validation. All the authors read and approved the final version of the manuscript.

Conflict of interest: The authors declare that they have no competing interests.

References

- Barrows, H.S. & Tamblyn, R. (1980) Problem-Based Learning: An Approach to Medical Education. Springer, New York USA.
- Basheti, I.A., Ryan, G., Woulfe, J. & Bartimote-Aufflick, K. (2010) Anonymous peer assessment of medication management reviews. American Journal of Pharmaceutical Education, 74, pp. 77.
- Das, M., Mpofu, D., Dunn, E. & Lanphear, J.H. (1998) Self and tutor evaluations in problem-based learning tutorials: is there a relationship? Medical Education, 32, pp. 411-418.
- Evans, A.W., Leeson, R.M.A. & Petrie, A. (2007) Reliability of peer self-assessment scores compared with trainers' scores following third molar surgery. Medical Education, 41, pp. 866-872.
- Gukas, I.D., Miles, S., Heylings, D.J. & Leinster, S.J. (2008) Medical students' perceptions of peer feedback on an anatomy student-selected study module. Medical Teacher, 30, pp. 812-814.
- Kamp, R.J.A., Dolmans, D.H.J.M., Van Berkel, H.J.M. & Schmidt, H.G. (2011) Can students adequately evaluate the activities of their peers in PBL? Medical Teacher, 33, pp. 145-150.
- Kozu, T. (2006) Medical education in Japan. Academic Medicine, 81, pp. 1069-1075.
- Machado, J.L., Machado, V.M., Grec, W., Bollela, V.R. & Vieira, J.E. (2008) Self- and peer assessment may not be an accurate measure of PBL tutorial process. BMC Medical Education, 8, pp. 55.
- Michaelsen, L.K. & Sweet, M. (2008) The essential elements of team-based learning. New Directions for teaching and learning, 116, pp. 7-27.
- Papinczak, T., Young, L., Groves, M. & Haynes, M. (2007) An analysis of peer, self, and tutor assessment in problem-based learning tutorials. Medical Teacher, 29, pp. e122-132.
- Reiter, I.L., Eva, K.W., Hatala, R.A. & Norman, G.R. (2002) Self and peer assessment in tutorials: Application of a relative-ranking model. Academic Medicine, 77, pp. 1134-1139.
- Sullivan, M.E., Hitchcock, M.A. & Dunnington, G.L. (1999) Peer and self assessment during problem-based tutorials. American Journal of Surgery, 177, pp. 266-269.
- Tashiro, Y., Mochiki, I., Morio, K., Semiya, S., Shiraishi, T. & Hori, H. (2014) Improvement of attitude evaluation in PBL by peer-evaluation. Igakukyouiku, 45(suppl.), pp. 91.
- Tashiro, Y., Morio, K., Mochiki, I., Shiraishi, T. & Hori, H. (2015) Improvement of attitude evaluation in PBL by peer-evaluation on a scale of one to ten. Igakukyouiku, 46(suppl.), pp. 156.
- Topping, K.J. (2009) Peer Assessment. Theory Into Practice, 48, pp. 20-27.
- Wood, D. (2003) ABC Of Learning and Teaching in Medicine: Chapter 3 (Problem-Based Learning). BMJ Publishing Group, 8, pp. 10-11.

Supplementary table - Table S1 Correlation between peer and tutor assessments in units with the formal PBL tutorials and those with a combination of PBL tutorials and TBL by the academic year

Academic year	2014-2015 (n=124)			2015-2016 (n=125)			2016-2017 (n=124)			2014-2017 (n=373)		
Assessment domain	1	2	3	1	2	3	1	2	3	1	2	3
4 units with the formal PBL tutorials	.74**			.72**			.74**			.63**		
		.80**			.78**			.80**			.65**	
			.60**				.61**		.60**			.42**
4 units with PBL tutorials and TBL	.68**			.71**			.56**			.62**		
		.65**			.78**			.58**			.63**	
			.27**				.68**		.40**			.35**
All 8 units	.79**			.80**			.77**			.71**		
		.78**			.85**			.79**			.73**	
			.48**				.75**		.60**			.45**

PBL, Problem-based learning; TBL, Team-based learning

1, self-learning in preparation; 2, active participation in group discussion; 3, cooperative attitude at group works.

The correlation was analyzed with the Pearson's correlation coefficient for self-learning in preparation and active participation in group discussion and the Spearman's correlation coefficient for cooperative attitude at group works.

**, $p<0.01$