RESEARCH ARTICLE

The validity and reliability of a problem-based learning implementation questionnaire

Bhina Patria[1,2]*

[1]Faculty of Psychology, Universitas Gadjah Mada, Yogyakarta, Indonesia; [2]International Centre for Higher Education Research Kassel, University of Kassel, Kassel, Germany

Abstract

Purpose: The aim of this paper is to provide evidence for the validity and reliability of a questionnaire for assessing the implementation of problem-based learning (PBL). This questionnaire was developed to assess the quality of PBL implementation from the perspective of medical school graduates. **Methods:** A confirmatory factor analysis was conducted to assess the validity of the questionnaire. The analysis was based on a survey of 225 graduates of a problem-based medical school in Indonesia. **Results:** The results showed that the confirmatory factor analysis model had a good fit to the data. Further, the values of the standardized loading estimates, the squared inter-construct correlations, the average variances extracted, and the composite reliabilities all provided evidence of construct validity. **Conclusion:** The PBL implementation questionnaire was found to be valid and reliable, making it suitable for evaluation purposes.

Key Words: Multivariate analysis; Problem-based learning; Psychometrics; Questionnaires; Statistical factor analysis

INTRODUCTION

In recent years, there has been an increased interest in comparing problem-based learning (PBL) with learning in conventional classrooms (i.e. non-PBL) [1,2]. However, there are some concerns about the research methods used in studies comparing PBL and non-PBL. In PBL research papers, the learning environment is labelled PBL, but no further information about the learning environment is provided [3]. This has led to concerns that a learning environment might be labelled PBL while in reality it is not, or, conversely, labelled non-PBL even when the principles of PBL (e.g. student-centred and self-directed learning) are being observed. Prior studies have also noted that identifying PBL is difficult [4]. Although some characteristics have been proposed by Barrows [5], institutions of higher education have different interpretations of these characteristics [6]. As a result, two universities may declare that they use PBL in their curriculum, but their implementation of PBL may be completely different. Currently, as the identification of PBL is considerably difficult, it is improper to compare the outcomes of PBL among different institutions without evaluating how well the PBL approach was implemented. The present study proposes that a PBL implementation be evaluated before a comparison of PBL and other approaches. This evaluation used the perspective of medical school graduates. The use of graduate survey data in PBL studies is not new [7]. However, prior studies have mostly used graduate survey data to investigate the long-term effects of PBL on educational outcomes. In the present study, graduate survey data were used for evaluating how well the PBL approach was implemented. As part of this goal, a questionnaire assessing the PBL implementation was developed considering the theoretical bases of PBL [5]. This study aims to demonstrate the validity and reliability of the PBL implementation questionnaire.

METHODS

Subjects

The subjects of this study were recent graduates from the Faculty of Medicine, Gadjah Mada University (UGM), Indonesia. Although the teaching faculty members began to introduce PBL into the curriculum in 1992, it was not until 2002 that they started to implement PBL as a full system. A medical degree requires five years at UGM and consists of three phases: a thorough grounding in medical knowledge, the transition from theory to practice, and a clinical rotation phase. Participants in the survey were graduates awarded their doctorates in medicine between February 5, 2009 and July 8, 2011. The graduate survey was done in May 2012, meaning that graduates were surveyed between eight months and three years after graduation. Out of 719 graduates, 225 participated in this study. The gender proportion of the dataset (54.2% female and 45.8% male) was precisely equal to the proportion in the population of interest. Graduates' average age when the survey was conducted was 26.3 years old ($SD = 2.27$, $M = 26$).

Instrument

The development of the PBL implementation questionnaire began with the construction of the indicators of the factors. These factors were based on six characteristics of PBL: student-centred learning; learning in small groups; teacher as facilitator; problems as the stimulus for learning; problems that reflect the real world; and acquiring new information through self-directed learning [5]. The graduates responded to the indicators based on the question: "To what extent did the following statements match the conditions in your study course?" The Likert scale responses ranged from 1 ("Not at all") to 5 ("To a very large extent"). The indicators of each factor were developed particularly for this study except for *Teacher as facilitator*. The indicators of *Teacher as facilitator* were adapted from Dolmans and Ginn's tutor effectiveness questionnaire [8]. The following sections elaborate on the development of indicators for each factor:

Student-centred learning: The indicators of the student-centred learning (SCL) factor were based on several definitions of PBL [9].

Small groups: The indicators of the small group factor were developed on the basis of the notion that small student groups should consist of five to nine students [5]. The indicators of the small group factor were also based on a study that argued that small groups should have several characteristics: a positive, nonthreatening atmosphere; active student participation and group interaction; adherence to group goals; clinical relevance and integration; and the effective use of certain pedagogical materials (e.g. cases) that promote thinking and problem solving [10].

Problem as stimulus: The indicators of problem as stimulus were created on the basis of four criteria for a problem. A PBL problem should match the students' level of knowledge, motivate students for further study activities, be suitable for the process of the analysis to be applied, and direct the students to conform with the faculty's educational objectives [11].

Real-world problems: The real-world problems factor was developed from the criteria for constructing problems in PBL [12]. An additional item was constructed on the basis of Barrows' suggestion that such problems must be presented as poorly structured problems [13], as poorly structured problems stimulate learners to generate multiple hypotheses about the problem's cause and possible solutions.

Teacher as facilitator: The indicators of teacher as facilitator were adapted from a tutor-effectiveness questionnaire developed by Dolmans and Ginns [8]. The questionnaire consists of 11 items representing five underlying factors: constructive or active learning, self-directed learning, contextual learning, collaborative learning, and intra-personal behaviour. The self-directed learning factor was excluded because the present study had the same factor, which is described in the next section.

Self-directed learning: The indicators of the self-directed learning (SDL) factor were developed on the basis of the definition of SDL. SDL is a process in which individuals take the initiative to diagnose their learning needs, formulate learning goals, identify resources for learning, choose and implement learning strategies, and evaluate learning outcomes [14]. Additional indicators were created on the basis of the notion that SDL includes self-monitoring and self-assessment components [15].

Statistical analysis

This study used confirmatory factor analysis (CFA), a type of structural modelling, to assess the relationship between the indicators of PBL and their factors. After the indicators of the PBL factors were developed on the basis of the theoretical foundation, the structural models were tested. The structural equation modelling was conducted using AMOS 20 (IBM, Armonk, NY, USA) and R statistical software using two packages, *lavaan* and *sem*. Each individual construct of the PBL implementation questionnaire was tested separately before testing the structural model. To increase the individual construct model fitness, re-specification was conducted by reducing items with a small standardized loading estimate. As a consequence, the total number of items between constructs was different.

Table 1 shows that all individual construct models have good fit statistics, given that the threshold value of the comparative fit index (CFI) is equal to or more than 0.97 and the

Table 1. Fit statistics of problem-based learning factors

	χ^2	df	P	RMSEA [a]	CFI
Student-centred learning	0.000	0	0.000	0	1
Small group	0.000	0	0.000	0	1
Problem as stimulus	19.381	9	0.022	0.075	0.986
Real-world problems	1.827	2	0.401	0	1
Teacher as facilitator	32.013	16	0.010	0.070	0.984
Self-directed learning	14.483	7	0.043	0.072	0.991

[a]RMSEA: root mean square error of approximation.

Table 2. Final indicators of problem-based learning questionnaire

Indicators	Variable
Student-centred learning	-
Students are responsible for their own learning	B14_A1
Students are actively involved in the process of learning	B14_A2
Students have autonomy in the process of learning	B14_A3
Small group	-
The group size is appropriate to stimulate group discussion	B14_B2
The learning groups have a positive atmosphere (i.e. non-threatening)	B14_B3
The group size is appropriate to encourage active student participation	B14_B4
Problem as stimulus	-
The problems in the tutorial process	-
…match students' level of knowledge	B14_C1
…stimulate thinking, analysis, and reasoning	B14_C2
…ensure self-directed learning	B14_C3
…activate students' prior knowledge	B14_C4
…lead to the discovery of the learning objectives	B14_C5
…arouse students' curiosity	B14_C6
Real-world problems	-
The problems in the tutorial process	-
…are realistic	B14_D1
…are clinically relevant	B14_D2
…are related to a public health topic	B14_D3
…generate multiple hypotheses about their cause and solution	B14_D4
Teacher as facilitator	-
The tutors have a clear picture about their strengths or weaknesses as a tutor	B14_E1
The tutors are clearly motivated to fulfil their role as a tutor	B14_E2
The tutors stimulate the students	-
…to summarize what they had learnt in their own words	B14_E3
…to search for links between issues discussed in tutorial groups	B14_E4
…to understand underlying mechanisms or theories	B14_E5
…to apply knowledge to other situations or problems	B14_E7
…to give constructive feedback about group work	B14_E8
…to evaluate group co-operation regularly	B14_E9
Self-directed learning	-
Students take initiative in diagnosing their learning needs	B14_F1
Students formulate their own learning goals	B14_F2
Students decide the resources for learning	B14_F3
Students choose appropriate learning strategies	B14_F4
Students evaluate the accuracy and value of the resources	B14_F5
Students self-monitor their learning progress	B14_F6
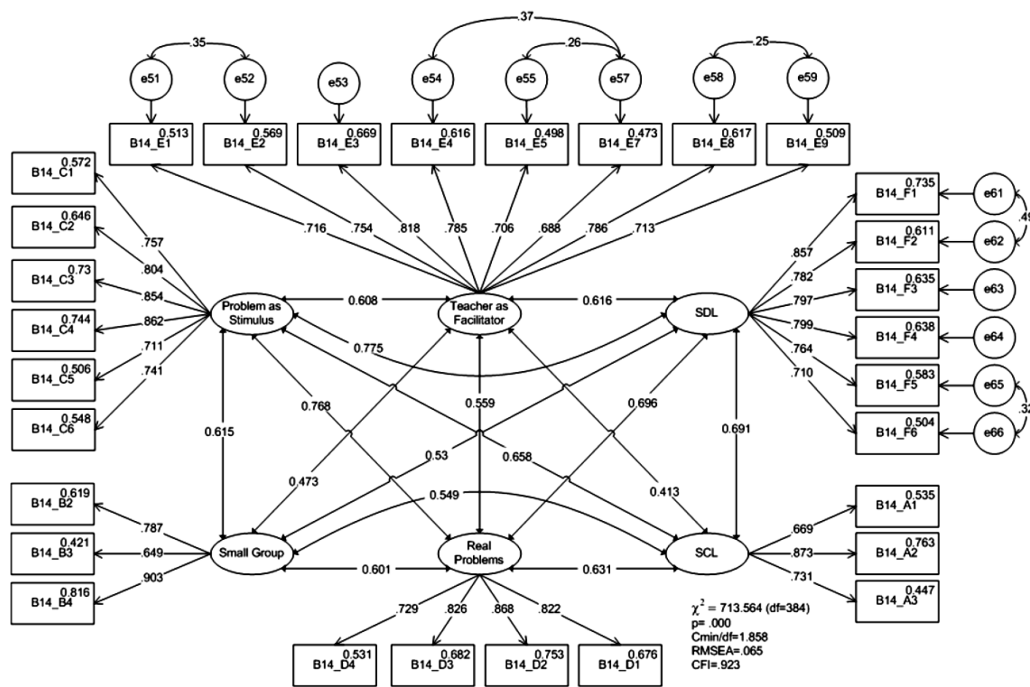
Fig. 1. The measurement model of a problem-based learning implementation. There are no correlated measurement errors for the factors of *problem as stimulus*, *small group*, *real problems*, or *student-centred learning*. Therefore, the measurement errors of these factors are omitted to reduce the size of the model.

Table 3. Construct validity statistics

	(1)	(2)	(3)	(4)	(5)	(6)	AVE	CR	ASV
(1) Student-centred learning	**0.762**	0.433	0.398	0.171	0.477	0.301	0.581	0.804	0.356
(2) Problem as stimulus	0.658	**0.790**	0.590	0.370	0.601	0.378	0.624	0.908	0.474
(3) Real-world problem	0.631	0.768	**0.813**	0.312	0.484	0.361	0.661	0.886	0.429
(4) Teacher as facilitator	0.413	0.608	0.559	**0.747**	0.379	0.224	0.558	0.910	0.291
(5) Self-directed learning	0.691	0.775	0.696	0.616	**0.786**	0.281	0.618	0.906	0.445
(6) Small group	0.549	0.615	0.601	0.473	0.530	**0.787**	0.619	0.827	0.309

Values below the diagonal are correlations among constructs. Diagonal values are construct variances. Values above the diagonal are squared correlations. AVE, average variance extracted; CR, composite reliability; ASV, average shared squared variance.

root mean square error of approximation (RMSEA) is less than 0.08 [16]. Table 2 depicts the final indicators of each PBL factor. After the specification of the indicators for each construct, the next stage was to develop the measurement model (i.e. confirmatory model). As mentioned above, the data in this study were based on the survey responses of 225 graduates of the Faculty of Medicine at the Universitas Gadjah Mada, Indonesia. After removing the non-responses (i.e. graduates who skipped all items of the PBL questionnaire), the data consisted of the responses of 207 graduates. Imputation (i.e. mean substitution) was applied to the missing values, as this procedure allowed the author to create modification indices, which are necessary in the re-specification of a model.

RESULTS

When the CFA model (Fig. 1) was fitted to the data, the following fit indices resulted: χ^2 (384, N = 207) = 713.564, P = 0.000, root mean square error of approximation (RMSEA) = 0.065, CFI = 0.923. The model consisted of 30 observed variables (N = 207). For a model with 30 or more observed variables and N < 250, the suggested fit statistics are as follows: CFI ≥ 0.92 and RMSEA < 0.08 [15]. Therefore, this model was used as the final measurement model of the PBL questionnaire without re-specification. The validity of a measurement model depends on both establishing acceptable levels of the goodness-of-fit for the measurement model and finding specific evidence

of construct validity [16]. Therefore, the main objective of this study was not only to assure the goodness-of-fit of the PBL implementation questionnaire but also to assess its construct validity.

Additional statistics were calculated from the result of the model testing to provide evidence of construct validity. These statistics are squared inter-construct correlation, average variance extracted (AVE), composite reliability (CR), and average shared squared variance (ASV). Table 3 presents the complete statistics. A classical reliability analysis was conducted to check the internal consistency of the questionnaire. All alpha coefficients were above the suggested value of 0.70, ranging from 0.787 (*small group*) to 0.921 (*teacher as facilitator* and *self-directed learning*). The alpha coefficient for the total items was 0.963, indicating that the questionnaire was internally consistent in measuring the target construct. The omega hierarchical coefficient (ωh) for the PBL implementation scale was 0.97, confirming that the indicators of the PBL implementation scale measure a common latent variable [17] (i.e. the implementation of PBL at the institution). The omega hierarchical coefficient was calculated using the *psych* package in the R statistical software.

DISCUSSION

The CFA results showed that the PBL implementation questionnaire had acceptable goodness-of-fit, which indicates that the measurement model fit the data. In addition to acceptable goodness-of-fit, it is necessary to provide evidence of construct validity. Construct validity was confirmed in the present study by establishing face validity, convergent validity, and discriminant validity. In this study, convergent validity and discriminant validity refer to the concepts proposed by Hair et al. [16]. The former is a condition where the indicators of a construct share a high proportion of variance in common. The latter refers to the extent to which a construct is truly distinct from other constructs. In addition to establishing construct validity, this section will also address two issues related to the limitations of the present study: dealing with factors with three indicators, and the correlated measurement errors.

Face validity

Face validity assesses the extent to which an instrument appears to measure what it is intended to measure. This type of validity can be achieved by approval from experts regarding whether the indicators measure the construct of interest or not. Accordingly, the draft of the PBL implementation questionnaire was checked by experts from the field of PBL research and methodology. The draft was also assessed by an expert on quantitative research methods, doctoral students in

medicine, and medical doctors who had graduated from Gadjah Mada University.

Convergent validity

Convergent validity requires the following: minimum standardized loading estimates of 0.5 (ideally, 0.7 or higher), an AVE of 0.5 or higher, and a CR of 0.7 or higher [15]. The final measurement model showed that the PBL questionnaire had close to ideal standardized loading estimates (Fig. 1). There were only three factor loadings below the ideal cut-off of 0.7 (B14_B3, B14_E7, and B14_A1), and all three remained above the minimum value of 0.5. Table 3 shows that all AVEs were above 0.5, with values ranging from 0.581 (*student-centred learning*) to 0.661 (*real-world problems*). Both factor loading and AVE indicate that the variance for each item in the PBL implementation questionnaire is explained more by a specific latent construct than by the measurement error. The CR values were above the suggested level of 0.7, ranging from 0.804 (*student-centred learning*) to 0.906 (*self-directed learning*). This indicates that the indicators of each construct are strongly interrelated [18].

Discriminant validity

To assure discriminant validity, the AVE should be greater than the squared inter-construct correlation. Table 3 indicates that the AVE of SCL (0.581) is higher than the squared correlation between SCL and *problem as stimulus* (0.433), SCL and *real-world problem* (0.398), SCL and *teacher as facilitator* (0.171), SCL and SDL (0.477), and SCL and *small group* (0.301). This means that the indicators of the SCL factor measure a specific construct that was not measured by the other factors. The other factors also showed a similar result, with all AVE values higher than the value of the inter-construct squared correlations. Another way to show discriminant validity is to use the average shared squared variance (ASV); discriminant validity is achieved when the AVE is greater than the ASV. The ASV was computed by averaging the inter-construct squared correlations. For example, the ASV of SCL = (0.433 + 0.398 + 0.171 + 0.477 + 0.301)/5 = 0.358. Table 3 shows that the AVE values of all factors are higher than the ASV, which indicates discriminant validity. Finally, the absence of factor cross-loading in the PBL measurement model also supports the discriminant validity of the PBL measurement model. Cross-loading is a condition where an indicator loads to more than one construct. Fig. 1 shows that all indicators load to only one factor.

Factors with three indicators

In the present study, there were two factors that had three indicators each: the *student-centred learning* and *small group* factors. A three-indicator model, or *just-identified* model, by

nature will lead to a perfect fit, as there are just enough degrees of freedom to estimate all the parameters (degree of freedom = 0). Just-identified models do not test theories because their fit is determined by their specific circumstances. However, a model with three-indicator factors is acceptable, particularly when other factors have more than three indicators [16]. In the present study, these three-indicator factors are acceptable because the measurement model includes other factors that each consist of more than three indicators: *problem as stimulus* (six indicators), *real-world problems* (four indicators), *teacher as facilitator* (eight indicators), and *self-directed learning* (six indicators). Although goodness-of-fit does not apply to a just-identified model, the model can still be evaluated in terms of the interpretability and strength of its parameter estimates (e.g. magnitude of factor loading) [18]. In the present study, the questionnaire was reviewed by experts on PBL and the chosen methodology. These experts' agreement on the validity of the questionnaire provides sufficient evidence of good interpretability. Finally, the factor loadings of *student-centred learning* (0.71, 0.92, and 0.63) and *small-group* factor (0.75, 0.60, and 0.97) were all satisfactory.

Correlated measurement errors

In a structural model, it is common to establish correlation paths among the measurement errors to improve the fit statistics. This method was used in the measurement model of *teacher as facilitator* and self-directed learning. In cross-sectional studies, there should be no correlated measurement errors; that is, the indicators should measure nothing other than the construct that the indicator is intended to measure. However, a correlated measurement error is acceptable in panel studies because the shared variance between the indicators might come from a prior measurement effect. The correlated measurement error can, however, be justified in a cross-sectional study when there is evidence of *source* or *method effects*. Method effects exist when the measurement approach, rather than the substantive latent factors, causes differential covariance among items [18]. Some possible method effects related to the present study are the *scale format* and *scale anchor, similar item wording*, and *social desirability* [19].

Scale format and anchor are related to the use of a standardized rating system. Most of the questions in the present study used a similar scale format, that is a semantically differential style with similar scale anchors or values. The anchor ran from 1 = 'Not at all' to 5 = 'To a very large extent'. A possible concern is that the consistency in the scales used, rather than the actual content of the item, might have affected the covariance in the construct. This means that the graduates might have fallen into a repetitive response pattern and disregarded the real content of the questionnaire. Respondents perceived items with high social desirability to correlate with each other because of the similar level of social desirability, rather than because of the items' content [19]. This could explain the correlated measurement error that was found in the present study, for example, for the *teacher as facilitator* factor in the PBL questionnaire. Item B14_E1 ('The tutors have a clear picture about their strengths/weaknesses as a tutor') and B14_E2 ('The tutors are clearly motivated to fulfil their role as a tutor') were suspected to have similar levels of social desirability as compared to the other items, one possible cause of their correlated errors. The correlated measurement errors in the present study are also acceptable because the variance of most items came from the latent construct rather than from the measurement error, with the two strong pieces of evidence being that the factor loadings of most items were higher than 0.70 and that the AVE values of most items were higher than 0.5. Therefore, although the measurement errors were correlated—indicating the existence of an unknown construct—most of the items' variance still came from the latent factor, not from the unknown construct. Finally, the correlated errors existed within single factors, with no inter-factor correlated errors. Thus, the correlated error did not violate the model's underlying theory.

The results of CFA demonstrated that all individual constructs of PBL had acceptable goodness-of-fit and that the goodness-of-fit of the PBL measurement model was sufficient for a good model. All the structural models proposed in this study fulfilled the requirements of good model fit, indicating that the theoretical models fit the data well. Measurement model validity does not only depend on establishing goodness-of-fit but also on providing evidence of construct validity [16]. In this study, the evidence for construct validity was supported by acceptable values for factor loadings, AVE, squared inter-construct correlation, and composite reliability. Further, there was no cross-loading factor.

In conclusion, on the basis of the above findings, we concluded that the PBL implementation questionnaire under study is valid and reliable, making it suitable for evaluation purposes. The questionnaire can thus be used to evaluate graduates' perspectives regarding the implementation of PBL. The results of such an evaluation could then help to improve PBL within institutions. This finding has important implications for the development of a better method for creating studies that compare the PBL and non-PBL approaches, as well as for future studies investigating the effects of individual PBL components on specific educational outcomes.

ORCID: Bhina Patria: http://orcid.org/0000-0003-1419-4240

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

SUPPLEMENTARY MATERIAL

Audio recording of the abstract.

REFERENCES

1. Schmidt HG, Cohen-Schotanus J, Arends LR. Impact of problem-based, active learning on graduation rates for 10 generations of Dutch medical students. Med Educ. 2009;43:211-218. http://dx.doi.org/10.1111/j.1365-2923.2008.03287.x
2. Watmough S, Cherry MG, O'Sullivan H. A comparison of self-perceived competencies of traditional and reformed curriculum graduates 6 years after graduation. Med Teach. 2012;34:562-568.
3. Dolmans D, Gijbels D. Research on problem-based learning: future challenges. Med Educ. 2013;47:214-218. http://dx.doi.org/10.1111/medu.12105
4. Newman M. A pilot systematic review and meta-analysis on the effectiveness of problem-based learning [Internet]. Newcastle (UK): Catherine Cookson Centre for Medical Education and Health Informatics; 2003 [cited 2014 Sep 16]. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.133.6561&rep=rep1&type=pdf
5. Barrows HS. Problem-based learning in medicine and beyond: a brief overview. In: Wilkerson L, Gijselaers W, editors. Bringing problem-based learning to higher education: theory and practice. San Francisco (CA): Jossey-Bass; 1996, p.3-11.
6. Neville AJ. Problem-based learning and medical education forty years on. Med Princ Pract. 2009;18:1-9. http://dx.doi.org/10.1159/000163038
7. Patria B. Problem-based learning, graduates' competencies and career success. Yogyakarta, Indonesia: Inparametric Press; 2011, p.15-21.
8. Dolmans DHJM, Ginns P. A short questionnaire to evaluate the effectiveness of tutors in PBL: validity and reliability. Med Teach. 2005;27:534-538. http://dx.doi.org/10.1080/01421590500136477
9. Brandes D, Ginnis P. A guide to student-centred learning. London (UK): Basil Blackwell Ltd.; 1986.
10. Steinert Y. Student perceptions of effective small group teaching. Med Educ. 2004;38:286-293. http://dx.doi.org/10.1046/j.1365-2923.2004.01772.x
11. Majoor GD, Schmidt HG, Snellen-Balendong HAM, Moust JHC, Stalenhoef-Halling B. Construction of problems for problem-based learning. In: Nooman Z, Schmidt HG, Ezzat ES, editors. Innovation in medical education. New York (NY): Springer; 1990, p.114-122.
12. Marchais JED. A Delphi technique to identify and evaluate criteria for construction of PBL problems. Med Educ. 1999;33:504-508. http://dx.doi.org/10.1046/j.1365-2923.1999.00377.x
13. Savery JR. Overview of problem-based learning: definitions and distinctions. Interdiscip J Probl based Learn. 2006;1:9-20. http://dx.doi.org/10.7771/1541-5015.1002.
14. Brockett RG. Self-directed learning readiness and life satisfaction among older adults. Ann Arbor (MI): University Microfilms International; 1982, p.237.
15. Barrows HS. Problem-based learning applied to medical education. Springfield (IL): Southern Illinois University School of Medicine; 2000.
16. Hair J, Black W, Babin B, Anderson R. Multivariate data analysis. 7th ed. Harlow (UKJ): Pearson Education Limited.; 2013, p.752.
17. Zinbarg RE, Yovel I, Revelle W, McDonald RP. Estimating generalizability to a latent variable common to all of a scale's indicators: a comparison of estimators for ωh. Appl Psychol Meas. 2006;30:121-144. http://dx.doi.org/10.1177/0146621605278814
18. Brown TA. Confirmatory factor analysis for applied research. New York (NY): The Guilford Press; 2006, p.475.
19. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol. 2003;88:879-903. http://dx.doi.org/10.1037/0021-9010.88.5.879