Educational Measurement and Evaluation Review (2016), Vol. 7


# Multigroup Invariance of Course Experience Questionnaire Across Gender Using Malaysian Sample

**Lei Mee Thien**
*SEAMEO RECSAM*

## Abstract

The purpose of this study is to examine the measurement invariance of CEQ across gender of Malaysian undergraduate students in two public universities. A survey method was employed for data collection. A total of 315 undergraduate students in two Malaysian public universities responded to 23 items measuring five scales of Good Teaching, Generic Skills, Clear Goals and Standards, Appropriate Workload, and Appropriate Assessment. Covariance Based Structural Equation Modelling was used as the analysis method. A multigroup analysis of invariance was performed to examine measurement invariance across male and female students using AMOS 20.0 computer software. Findings showed that configural invariance was fully supported whereas metric invariance was partially supported across female and male groups. This study has shown the satisfactory psychometric properties of CEQ 23 with only one item that was found not convey the same meaning across gender. Findings could be used as the empirical evidences to compare with the future empirical-based CEQ studies.

**Keywords:** multigroup invariance, Course Experience Questionnaire, teaching quality, configural invariance, metric invariance

## Introduction

Surveys of student perceptions of their program or course experience in higher education are found widely used over the world such as Australia and United Kingdom (Yorke, 2009). The results from

the survey are particular importance as it serves as the key performance indicators of the higher education institution (Yorke, 2009). In relation to this, Course Experience Questionnaire (CEQ) was developed and adopted as a domain-neutral indicator of university course quality. CEQ is an Australian-based instrument that has been used to probe university students' perceptions of their programme of study experience in higher education. CEQ has been widely used as a performance indicator and benchmarking of students' learning experience in Asia Pacific and Western countries particularly Australia, UK, and Canada universities context (Ginns, Prosser, & Barrie, 2007). Deriving from the theory of learning and teaching, CEQ was originated from the item pool of Course Perception Questionnaire (Ramsden & Entwistle, 1981), School Experience Questionnaire (Ramsden, Martin, & Bowden, 1989), as well as Experiences Studying and Higher Education Questionnaire (Entwistle & Tait, 1990). CEQ 30 which was advanced by Trigwell and Prosser (1991) and Ramsden (1991) is the first version that evolved from the item pools mentioned. CEQ 30 comprised five scales, such as: Good Teaching (8 items), Clear Goals and Standards (5 items), Appropriate Workload (5 items), Appropriate Assessment (6 items) and Emphasis on Independence (6 items). The psychometric properties of CEQ 30 was warranted and has undergone series of improvement from the statistical perspective over decades (e. g., Ainley & Long, 1994; Richardson, 1994; Wilson, Lizzio, & Ramsden, 1997).

The improved version refers to CEQ 23. CEQ 23 was developed in consultation with the Department of Employment, Education, and Training (DEET) and has been used in national survey of Australian graduates since 1993. Differs from former version, CEQ 23 consisted of the scales of Good Teaching (6 items), Generic Skill Scale (6 items), Clear Goals and Standards (4 items), Appropriate Workload (4 items), and Appropriate Assessment (3 items) with the exclusion of scales of Emphasis on Independence (6 items). Interestingly, CEQ 23 has been extensively piloted in different contexts using a variety of data analysis. However, the main concern of the existing studies is limited to the levels of internal consistency and satisfactory of its factor structure (Wilson et al., 1997). For instances, in British university context, Broomfield and Bligh (1998) confirmed the five factor structures of CEQ 23 and further split Good Teaching scale into two subscales: teacher interaction and presentation; and the quality

of feedback given to students. Distinctly, Eley (2001) examined three parallel versions of CEQ 23 to examine whether altering question format and phrasing can improve the effectiveness of CEQ. The first version refers to Graduate Careers Council of Australia (GCCA) with 5-point Likert scale from agree (1) to disagree (6). The second and third versions modified the regular CEQ 23 to devise Behavioural Observation Scale (BOS) and Dimensional Rating Scale (DRS), respectively. The internal consistency appeared to be satisfactory for the three versions of regular CEQ, BOS, and DRS version. However, the five factor solutions for each version warrant comments with the presence of several items which loaded on unintended scales. Similarly, Byrne and Flood (2003) confirmed the satisfactory levels of reliability and construct validity for use in the accounting discipline in an Irish university. Meanwhile, in Hong Kong universities context, Ho (1998) as well as Law and Meyer (2011) have provided the satisfactory evidence of reliability of the CEQ 23. On the other hand, Thien and Ong (2016) found that only two dimensions of CEQ, namely: Good Teaching and Generic Skills were applicable in Malaysian higher education context.

Notably, most of previous studies have typically assumed CEQ 23 was operating in exactly the same way across the groups of interest. Such assumption sounds less convincing as extant literature informed gender differences in many areas of higher education research (Grebennikov & Skaines, 2009). For instance, female students now outnumber male students (Bradley, 2000). Female students were also found over-represented in most of the Malaysian public universities (Ismail, 2015). In relation to this, multigroup invariance which refers to the extent to which the content of each item is being perceived and interpreted in the same way across groups should be initially emphasised (Byrne & Watkins, 2003). The ignorance of multigroup analysis of invariance across groups often lead to contradictory findings that subsequently misleading the direction of future studies (Byrne, 2008, 2010). As such, the multigroup analysis of invariance need to be conducted using a rigorous statistical techniques suggested by Byrne (2004). The significance of this empirical study hinges upon the fact that it contributes methodological knowledge in higher education literature using a Malaysian undergraduate student sample.

## Conceptualization and Operationalization

Course Experience Questionnaire (CEQ) advanced by Ramsden (1991) has five underlying scales: (1) Good Teaching, (2) Generic Skills, (3) Clear Goals and Standards, (4) Appropriate Workload, and (5) Appropriate Assessment.

According to the report of Graduate Course Experience Sydney 2009, Good Teaching is conceptualized as the nature of teaching experienced during a course. Good Teaching is operationalized as the degree to which the graduates feel that the teaching staff or lectures of their course provided a high level of teaching quality. Meanwhile, Generic Skills is conceptualized as the enhancement of selected generic skills with its operationalization as the extent to which the course adds to the generic skills that graduates might be expected to possess.

The third scale, namely, Clear Goals and Standards is conceptualized as the clarity and meaningfulness of course structure. Clear Goals and Standards is thereby operationalized as the extent to which graduates feel that they were provided with enough information regarding the learning objectives of their course and the standards of work expected from them.

On the other hand, Appropriate Workload is conceptualized as the level of workload that hindered deeper forms of learnings and it is operationalized as the degree to which graduates felt the workload involved in their course were excessive. The fifth scale, namely, Appropriate Assessment is conceptualized as the level that assessment that promoted deeper forms of learning. It is operationalized as the extent to which courses depend on the recollection of factual knowledge for assessment purpose. Table 1 shows the number of items for each scale and its description. Each scale was capitalized throughout the paper to avoid interpretation confusion.

Table 1
*Item Descriptions*

| Scale | Item | Description |
|---|---|---|
| Good Teaching (GTS) | GTS1 | The lecturer put a lot of time into commenting on my work. |
| | GTS2 | The lecturer of this course motivated me to do my best work. |
| | GTS3 | The lecturer made a real effort to understand difficulties I might be having with my homework. |

| Scale | Item | Description |
|---|---|---|
| | GTS4 | The lecturer normally gave me helpful feedback on how I was doing. |
| | GTS5 | The lecturers were extremely good at explaining things. |
| | GTS6 | The lecturers work hard to make their subject interesting. |
| Generic Skills (GSS) | GSS1 | The program developed my problem-solving skills. |
| | GSS2 | The program sharpened my analytic skills. |
| | GSS3 | The program helped me develop my ability to work as a team member. |
| | GSS4 | As a result of my program, I felt confident about tackling unfamiliar problems. |
| | GSS5 | The program improved my skills in written communication. |
| | GSS6 | The program helped me to develop the ability to plan my own work. |
| Clear Goals and Standards (CGS) | CGS1 | It was easy to know the standard work expected. |
| | CGS2 | I usually had a clear idea of where I was going and what was expected of me in this program. |
| | CGS3* | It was often hard to discover what was expected of me in this program. |
| | CGS4 | The lecturers made it clear right from the start what they expected from students. |
| Appropriate Workload (AWS) | AWS1* | The workload was too heavy. |
| | AWS2 | I was generally given enough time to understand the things I had to learn. |
| | AWS3* | There was a lot of pressure on me as a student in this program. |
| | AWS4* | The sheer volume of work to be done though in this program it means it couldn't all be thoroughly comprehended. |
| Appropriate Assessment (AAS) | AAS1* | To do well in this program all you really needed was a good memory |
| | AAS2* | The lecturers seemed more interested in testing what I had memorized than I had understood. |
| | AAS3* | Too many lecturers asked me questions about facts. |

*Note*: * represents recoded items.

## The Present Study

The purpose of the present study is to examine the item measures for each dimension of CEQ, namely (1) Good Teaching, (2) Generic Skills, (3) Clear Goals and Standards, (4) Appropriate Workload, and (5) Appropriate Assessment across gender of Malaysian undergraduate students in two public universities.

## Method

### Participants

The CEQ 23 was administered on 350 students in second, third, and fourth year undergraduate programs in academic year 2013 in two Malaysian public universities. The cohort of second, third, and fourth year students were selected because they have the experience with university learning culture. Therefore, their evaluation on teaching quality can be considered reliable. The participants were selected with convenience basis due to the time and cost constraints. The survey was administered by the current researchers with the assistance from the researcher officers who worked in the university. Questionnaires were distributed to the randomly selected social science and science-based faculties, including the faculties of education, social science, humanities, engineering, biological science, and computer sciences. A total of 315 of the completed questionnaires were returned with the response rate of 90%. According to Rasoolimanesh, Jaafar, Kock, and Ramayah (2015), only 166 sample for model testing is needed to generate a power value of 0.95. Therefore, in this study, the power value of sample size of 315 was exceeded 0.95 and thus considered sufficient.

The student sample consisted of 109 male and 206 female students. The dominant ethnic group was Malay (224) followed by Chinese (66), Indian (9) and the other minority ethnic groups (16).The composition of academic year of study for the participant was second year (222), third year (91) and fourth year (2) undergraduate students with the average age of 22 years old.

## Measures and Data Collection Procedures

Table 1 shows the defining items of the CEQ scales according to the report of Graduate Course Experience Sydney 2009. The data collected were entered into SPSS data file. Items that had an opposite meaning to that of the relevant scale were then recoded for the ease of finding interpretation. The recoded items were referred as AAS1, AAS2, AAS3, AWS3, AWS4 and CGS3. Each item was scored on a 5-point Likert scale from strongly disagree (1) to strongly agree (5). For convenience, the terms of 'staff' in original version of CEQ 23 has been modified as 'lecturer' in this study due to the contextual nature in local public universities. Similarly, the terms of 'course' has been modified to 'program' as the evaluation is based on the entire program instead of a specific course offered in each faculty.

## The Hypothesized Model

Figure 1 shows the hypothesized model schematically. There are five scales of CEQ: Generic Skills scale (6 items), General Teaching Scale (6 items), Clear Goals and Standards scale (4 items), Appropriate Workload scale (4 items), and Appropriate Assessment scale (3 items). Figure 1 shows the five scales or factors are intercorrelated as indicated by the double-headed arrows, and the measurement error terms associated with the observed variables ($e$'s) are uncorrelated.

## Data Analysis Procedures

This study employed Byrne's (2010) steps for testing measurement invariance using AMOS 20.0 computer software. In Step 1, the overall model, male- and female model are initially tested to ensure the model fit to provide an overview of how consistent the models results are. If consistency of factor structure is found, then the analysis proceeds to Step 2. Otherwise, the models need to re-specify rigorously for consistency. Step 2 involves the establishment of configural and metric invariance when conducting two or more different groups of sample (Hair, Black, Babin, & Anderson, 2010).

Configural invariance is satisfied when the basic model structure, indicating the pattern of fix and non-fixed parameters is invariant across group. The initial baseline model has no between-

group invariance constraints on estimated parameters. The configural invariance is therefore critically important because it provides the basis for comparison with all subsequent models. On the other hand, conducting metric invariance is to assure whether the items underlying each scale are used similarly across groups (Vandenberg, 2002; Hair et al., 2010). The test of metric invariance is conducted by constraining the factor loadings to be equal across groups because the pattern coefficients carry the information about the relationship between latent scores and observed scores. When metric invariance is established, the different scores on the item can be meaningfully compared across groups. Vandenberg and Lance (2000) proposed that configural and metric invariance should be established before comparisons across groups can be meaningful.

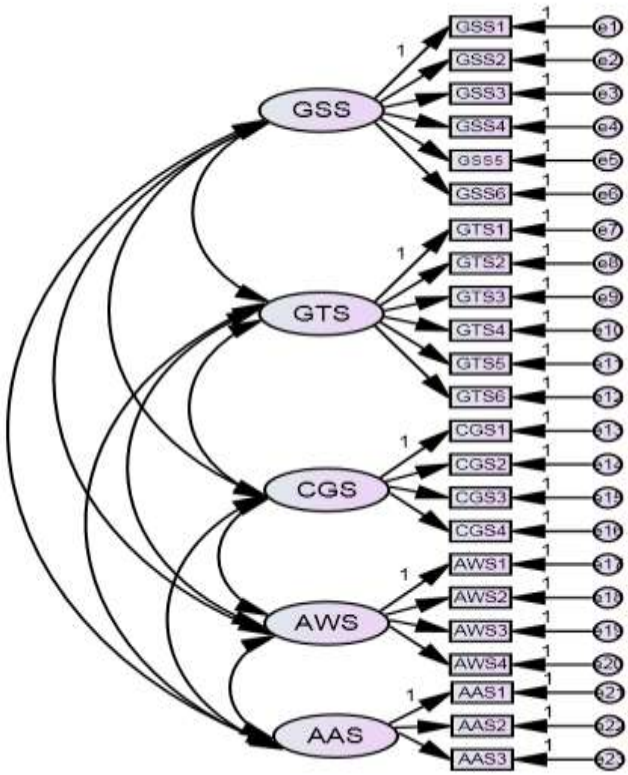


***Figure 1*** Hypothesized Model

The analysis procedures of Step 2 begins with a baseline model with chi-square value which derived by computing model fit for the pooled sample of all groups. Evidence of non-invariance is claimed if the chi-square difference value is statistically significant. The analysis was then followed by a hierarchical series of additional test that aimed at targeting which parameters are accounting for the non-invariant findings.

## Results

Prior to multigroup analysis, a preliminarily Confirmatory Factor Analysis (CFA) for the overall data, female, and male groups were conducted. Table 2 shows the hypothesized model for overall data, female and male groups yield only a marginally good fit to the data. The initial poor CFA results showed the needs for model re-specification.

The overall data model was re-specified with the exclusion of Item CGS3 and inclusion of error covariance between GTS1 and GTS3 yielded some improvement in goodness-of-fit, including Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Root Mean-Square Error of Approximation (RMSEA). This results in a modestly adequate fit for both female and male groups. Accordingly, the re-specified female and male models were deemed the most appropriate baseline models for both groups.

The first step of testing was configural invariance where no equality constraints were imposed. This configural model simply incorporates the baseline models for female and male groups and allows for their simultaneous analyses. Table 3 shows the configural model which also termed as unconstrained model fits reasonably well with $\chi^2(396, 206) = 1002.286$, TLI = 0.81, CFI = 0.84 and RMSEA = 0.07 and 95% C.I. between 0.064 and 0.075. The results concluded that the number of factors and pattern of their item loadings were similar across female and male groups.

For measurement invariance, we compared the factor loadings equivalence model to the unconstrained model. This initial test was termed full metric invariance. The results indicated the factor loadings were significantly different or not invariant with the difference in $\chi^2$ difference was 39.90 for female and male group and significant at the 0.05 level.

Table 2
*Summary Goodness-of-Fit Statistics in Determination of Baseline Models*

| Model description | $\chi^2$ | df | CFI | TLI | RMSEA | 90% CI |
|---|---|---|---|---|---|---|
| *Overall data (N=315)* | | | | | | |
| Hypothesized five-factor model (Model 1) | 854.73 | 220 | 0.83 | 0.80 | 0.090 | 0.089, 0.103 |
| Model 1 with the exclusion of Item CGS3 (Model 1a) | 735.82 | 199 | 0.84 | 0.82 | 0.089 | 0.087, 0.101 |
| Model 1a with one error covariance specified (Item GTS1 and GTS3) | 703.05 | 198 | 0.86 | 0.84 | 0.087 | 0.083, 0.097 |
| *Female undergraduate students (n=206)* | | | | | | |
| Hypothesized five-factor model (Model 2) | 735.43 | 220 | 0.79 | 0.76 | 0.107 | 0.098, 0.115 |
| Model 2 with the exclusion of Item CGS3 (Model 2a) | 663.39 | 199 | 0.81 | 0.77 | 0.106 | 0.098, 0.115 |
| Model 2a with one error covariance specified (Item GTS1 and GTS3) | 626.67 | 198 | 0.82 | 0.80 | 0.103 | 0.094, 0.112 |
| *Male undergraduate students (n=109)* | | | | | | |
| Hypothesized five-factor model (Model 3) | 442.95 | 220 | 0.84 | 0.81 | 0.097 | 0.084, 0.110 |
| Model 3 with the exclusion of Item CGS3 (Model 3a) | 393.22 | 199 | 0.85 | 0.83 | 0.095 | 0.081, 0.109 |
| Model 3a with one error covariance specified (Item GTS1 and GTS3) | 375.62 | 198 | 0.87 | 0.84 | 0.091 | 0.077, 0.105 |

*Note. df* represents degree of freedom, CFI represents Comparative Fit Index, TLI represents Tucker-Lewis Index, RMSEA = Root Mean-Square Error of Approximation, and CI represents confidence interval. Threshold of CFI, TLI, and RMSEA is 0.90 (Hu & Bentler, 1999). 0.90 (Hu & Bentler, 1999), and 0.03 (Browne & Cudeck, 1993), respectively.

Table 3
*Invariance Tests for Loadings across Female and Male Groups*

| Model description | Comparative model | $\chi^2$ | *df* | $\Delta\chi^2$ | $\Delta df$ | p |
|---|---|---|---|---|---|---|
| Unconstrained model (model1) | | 1002.286 | 396 | n.a. | n.a. | n.a. |
| Metric invariance (model 1A) | 1A Vs 1 | 1042.104 | 413 | 39.818 | 17 | $p <$ 0.001 |
| Constrained all loadings of GSS to be equal (model 1B) | 1B Vs 1 | 1007.756 | 401 | 5.469 | 5 | n.s. |
| Constrained all loadings of GSS and GST to be equal (model 1C) | 1C Vs 1 | 1015.175 | 406 | 12.889 | 10 | n.s. |
| Constrained all loadings of GSS,GST and AWS to be equal (model 1D) | 1D Vs 1 | 1037.063 | 409 | 34.777 | 13 | $p <$ 0.001 |
| Constrained all loadings of GSS, GST & AWS2 to be equal (model 1E) | 1E Vs 1 | 1036.110 | 407 | 33.824 | 11 | $p <$ 0.001 |
| Constrained all loadings of GSS, GST & AWS3 to be equal (model 1F) | 1F Vs 1 | 1015.274 | 407 | 12.988 | 11 | n.s. |
| Constrained all loadings of GSS, GST & AWS3 & AWS4 & CGS to be equal (model 1G) | 1G Vs 1 | 1019.192 | 410 | 16.906 | 14 | n.s. |
| Constrained all loadings of GSS, GST & AWS3 & AWS4, CGS & AAS to be equal (model 1H) | 1H Vs 1 | 1022.976 | 412 | 20.69 | 16 | n.s. |

*Note.* n.s. represent nonsignificant at .05 level and n.a. represents not applicable.

Since we do not have full metric invariance, the analysis was further proceeded to partial metric invariance. Partial metric invariance is used to identify at least two equal factor loadings between all constructs (Vandenberg & Lance, 2000). Thus, given a finding that the test rejected the null of equality, we further investigated the scales following the multisteps recommended by Byrne (2010). We investigated scale with problematic loading(s) to detect problem scale in advance in female and male groups by imposing constraint on all loadings within each scale in order. A chi-square different test fails to reject the null of equality of all loadings in a certain scale, then all the items in the scale were invariant between female and male group. If any chi-square different test rejected the null of equality of all loadings in a scale between female and male group, we then used a series of analyses by placing constraints on individual loadings in sequence in the scale. Finally, we compare chi-square values of constrained models with the baseline model (Byrne, 2010). Through the multistep process, all tests for invariance of the 22 loadings across female and male were completed. Table 3 shows only one item, namely, Item AWS2 is found non-equivalent between female and male groups.

## Discussion

In reality, it is possible that female and male respondents could perceive the contents of the items in CEQ 23 differently. In line with this point of view, Bentler (2004) reiterates that there is no guarantee that the instrument particularly in the form of questionnaire operates equivalently across different groups such as gender and ethnic groups. Considering this, there is a need to establish the consistency with respect to the relationship between a latent variable and its corresponding subscales or items across different groups. In other words, the items should be equally valid across different groups before conducting further analysis. Therefore, it is worthy for this present study to examine the mutligroup analysis of the measures of CEQ 23 using the measurement invariance across gender.

This study has shown one item, namely, AWS2 did not operate the same way across gender. AWS 2 which stated as "*I was generally given enough time to understand the things I had to learn*" deserves further investigation. The female and male undergraduate students may have misinterpreted the phrases. This phrase could be broadly

conceptualized and not well defined which carries different interpretation from female and male students based on their preferences of program studied. In this regard, item developers could revise and specify AWS2 based on a preference of program studied among the undergraduate students.

However, this study was limited to the invariance of factor loadings, representing a minimum condition for multigroup analysis. Therefore, it would be insightful to extend the present study that involved the testing of scalar invariance, factor covariances, and error covariance to achieve a full multigroup analysis of CEQ 23 in future studies. This study is also limited with the small sample size of the number of male students which only half of the female students and only involved two public universities. Future studies can be extended by including a larger sample size across gender. In addition to gender, measurement invariance across ethnics can be another avenue that deserves to be explored in future studies. Viewing from methodology perspective, the testing of measurement invariance is based on the chi-square statistic which are known to be sensitive to sample size. For this reason, using alternative fit indices in measurement invariance investigations is recommended for future research endeavor (Cheung & Rensvold, 2002).

For methodology implication, this study has shown the satisfactory psychometric properties of CEQ 23 with only one item that was found not convey the same meaning across gender. The findings of this present study would serve as the empirical evidence that can be used to compare and contrast with the findings of future empirical studies related to CEQ. More importantly, the method and procedures used of multigroup analysis as shown in this study could serve as an exemplar and replicable in future studies in addition to contribute methodological knowledge in higher education literature.

## References

Ainley, J., & Long, M. (1994). *The course experience survey 1992 graduates*. Canberra: Australian Government Publishing Service.

Bradley, K. (2000). The incorporation of women into higher education: Paradoxical outcomes? *Sociology of Education, 73*, 1–18.

Bentler, P. M. (2004). *EQS: Structural equations programs manual*. Encino, CA: Multivariate software.

Broomfield, D., & Bligh, J. (1998). *Knowing and reasoning in college: Gender-related patterns in students' intellectual development*. San Francisco: Jossey-Bass.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of accessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equations models* (pp. 136-162). Newbury Park, CA: Sage.

Byrne, B. M. (2004). Testing for multigroup invariance using AMOS graphics: A road less traveled. *Structural Equation Modeling, 11*(2), 272–300.

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicotherma, 20*(4), 872-882.

Byrne, B. M. (2010). *Structural Equation Modeling with AMOS: Basic concepts, applications and programming* (2nd ed.). NY: Routledge.

Byrne, B. M., & Flood, B. (2003). Assessing the teaching quality of accounting programmes: An evaluation of the course experience questionnaire. *Assessment and Evaluation in Higher Education, 28*, 135-145.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement equivalence revisited. *Journal of Cross-cultural Psychology, 34*, 155-175.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.

Eley, M. (2001). The course experience questionnaire: Altering question format and phrasing could improve the CEQ's effectiveness. *Higher Education Research and Development, 20*(3), 293-312.

Entwistle, N. J., & Tait, H. (1990). Approaches to learning, evaluations of teaching and preferences for contrasting academic environments. *Higher Education, 19*, 169-194.

Ginns, P., Prosser, M., & Barrie, S. (2007). Students' perceptions of teaching quality in higher education: The perspective of currently enrolled students. *Studies in Higher Education, 32*(5), 603-615. doi: 10.1080/03075070701573773.

Grebennikov, L., & Skaines, I (2009). Gender and higher education experience: A case study. *Higher Education Research and Development, 28*(1), 71-84.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis.* (7th ed.). New Jersey: Prentice Hall.

Ho, A. S. P. (1998). *Changing teachers' conceptions of teaching as an approach to enhancing teaching and learning in tertiary education* (Unpublished doctoral thesis). The University of Hong Kong, Hong Kong.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Ismail, L. (2015). Gender gap in higher education: Perspective on factors influencing enrolment in Malaysian universities. *The Online Journal of Quality in Higher Education, 2*(4), 35-45.

Law, D. C. S., & Meyer, J. H. F. (2011). Adaptation and validation of the course experience questionnaire in the context of post-secondary education in Hong Kong. *Quality assurance in Education, 19*(1), 50-66. doi: 10.1108/09684881111107753.

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The course experience questionnaire. *Studies in Higher Education, 16*, 129-150.

Ramsden, P., & Entwistle, N. J. (1981). Effects of academic departments on students' approaches to studying. *British Journal of Educational Psychology, 51*, 368-383.

Ramsden, P., Martin, E., & Bowden, J. (1989). School environment and sixth Form pupils' approaches to learning. *British Journal of Educational Psychology, 59,* 129-142.

Rasoolomanesh, S. M., Jaafar, M., Kock, N., & Ramayah, T. (2015). A revised framework of social exchange theory to investigate the factors influencing residents' perceptions. *Tourism Management Perspective, 16*, 335-345.

Richardson, J. T. E. (1994). A British evaluation of the course experience questionnaire. *Studies in Higher Education, 19*, 59-68.

Thien, L. M., & Ong, M. Y. (2016). The applicability of Course Experience Questionnaire (CEQ) for a Malaysian university context. *Quality Assurance in Education, 24*(1), 41-55.

Trigwell, K., & Prosser, M. (1991). Improving the quality of student learning: The influence of learning context and student approaches to learning on learning outcomes. *Higher Education, 22*, 251-266.

Vandenberg, R. J. (2002). Towards a further understanding of and improvement in measurement invariance methods and procedures. O*rganizational Research Methods, 5*(2), 139-158.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance

literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-69.

Wilson, K. L., Lizzio, A., & Ramsden, P. (1997). The development, validation and application of the course experience questionnaire. *Studies in Higher Education, 22*, 33-53.

Yorke, M. (2009). 'Student experience' surveys: Some methodological considerations and an empirical investigation. *Assessment & Evaluation in Higher Education, 34*(6), 721-739.