Volume: 2 Issues: 6 [December, 2017] pp.57-70]
International Journal of Education, Psychology and Counseling
eISSN: 0128-164X
Journal website: www.ijepc.com

MEASURING TEACHING AND LEARNING PERFORMANCE IN HIGHER EDUCATION

Habsah Muda[1],
Mass Hareeza Ali[2]
Mazuki Jusoh[3]

[1]Department of Management Sciences, Faculty of Economics and Management Sciences, Universiti Sultan Zainal Abidin, Gong Badak, 21300 Kuala Terengganu, Terengganu, Malaysia. Tel: 09-6688299,
Email: habsah_muda@unisza.edu.my. H-phone: 012-9387808
[2]Department of Management, Faculty of Economics and Management, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor. Email: mass@upm.edu.my
[3]Asian Institute of Finance, Kuala Lumpur, Malaysia.
Email: keyjus@yahoo.com

Accepted date: 12 November 2017 **Published date**: 19 December 2017

To cite this document: Muda, H., Ali, M.H., and Jusoh, M. (2017). Measuring Teaching and Learning Performance in Higher Education. *International Journal of Education, Psychology and Counselling*, 2 (6), 57-70.

Abstract:
The purpose of this paper was to establish a psychometric scale for measuring teaching and learning performance in the institution of higher learning. It is enlisted based on Kirkpatrick's four levels of evaluation model. The researcher also establishes the content validity through a series of expert review, pre-test and pilot testing using Exploratory Factor Analysis (EFA). The EFA provides dimensionality of measures for teaching and learning performance and the scale alpha coefficient (Cronbach's alpha) in the range of 0.894 to 0.904 provides a reliable measure of internal consistency in pilot testing. Later, the confirmatory factor analyses (CFA) using Structural Equation Model (SEM) provide empirical evidence of the unidimensionality, convergent, discriminant, and construct validity together with the normality of data and reliability of the teaching performance scales in the survey of 592 students' evaluation. Based on the field study data, the confirmatory factor analysis (CFA) for the measurement model in SEM verified the construct of the teaching and learning performance scales. Establishing a psychometric scale of teaching and learning performance using Kirkpatrick's four levels of evaluation is an innovative step towards teaching quality and efficiency. Theoretically, this study had provided a new dimension for the psychometric scale of teaching and learning performance based on Kirkpatrick's model of the reaction (level 1) and learning (level 2). Practically, both trainers and academic staff may use this measurement to assess their performance based on student evaluation in the context institution of higher education.

Keywords: *Confirmatory Factor Analysis (CFA), Higher Learning, Kirkpatrick's Evaluation Model, Psychometric Scale, Structural Equation Modelling (SEM) and Teaching Performance*

Introduction

Managing performance is associated with the competency proficiency and behaviours of an individual in strategic planning, monitoring and measuring their performance (Stiffler, 2006). The current practices of performance management in many organizations embrace various topics such as balanced scorecard, six sigma, financial reporting, data analysis, business intelligence, performance appraisals, competency management, training evaluation, incentive, compensation, and any other aspect of organizational and the individual performance (Breunig, & Hydle, 2013; Stiffler, 2006).

Managing performance in the context of higher learning commonly related to individual proficiency and competency. The elements of beliefs and values, skills, experience, personality characteristics, motivation, emotional issues and intellectual capabilities will affect competency and proficiency (Zwell & Michael, 2000). What people believe, has a big impact on their behaviour. If people believe that they are not creative or innovative, they will typically not attempt to think about new and different ways of doing things. In addition, Zwell and Michael (2000) stated that the sample key behaviours for managing performance includes; quality and quantity of performance, sets a clear definition of work activity outcomes, and seeks feedback. Furthermore, any differences between an ideal performer and other performers called a gap and usually expressed in behaviours or work outcomes (Rothwell & Graber, 2010).

In the context of higher learning, the primary skills of academics are not the ability to accumulate knowledge, but to communicate that knowledge to other people. Educators must ensure that the teaching and learning material understood, remembered by the students and that they are capable of both acting on the knowledge and recalling it. It related to the academics' presentational behaviour known as verbal and non-verbal behaviour. Whatever the personality of the academics, their behaviour must be helpful to the learners (Rae & Leslie, 2002). According to Jones and Sanghi (2006), the behaviour of an individual communicated through job performance. It represents the culmination of a series of dynamic and complex human processes on the relationship between individual personality traits, behaviour, job performance and environmental influences.

Measuring academic staff's performance in institutions of higher learning implicates an evaluation of teaching and learning, including supervision, research, publication, and community services based on the duties and responsibilities together with other qualities. A systematic evaluation of teaching and learning performance for higher educational setting is relatively limited and immature compared to those in corporate and industrial contexts. Thus scholars suggested borrowing from workplace training to develop a model that can facilitate effective teacher performance (Lawless & Pellegrino, 2007; Naugle, Naugle, & Naugle, 2000; Steinert et al., 2006). The specific model for predicting academic staff teaching and learning performance in institutions of higher learning hardly found in the literature. However, Zwell and Michael (2000) in general, proposed a wide range of factors that determine behaviour and performance and specifically stated that:

> *"The determinant factors include the size and shape of the head, brain weight, skin colour, ethnicity, social class, birth order, handwriting, religion, intelligence quotient (IQ), cultural heritage, astrology, heredity, gender, and so on. Additional factors of technical skills, years of experience, education, certification, and*

personality traits have been used to attempt to predict performance in the workplace."

In the early stage, Bell (1983) used the term "high performance trainer" to describe one who possesses efficiency, effectiveness and excellence in identifying and assessing through planned learning to help develop the key competencies that enable individuals to perform current or future jobs. The following review of the literature reveals a few models and constructs that significantly related to job performance.

Literature Review

Limited of previous studies for the specific teaching and learning performance of academic staff in institutions of higher learning. However, a few studies relevant to job performance (Taylor, 2001; Hubbal & Burt, 2006; Milanowski, 2011) and examined the relationship with personality traits (Van den Berg & Feij, 1993; 2003; Loveland et al, 2005; Chunping, Dengfen & Fan, 2009; James, 2010), supervision (Emillson & Johnson, 2007; Mainhard et al, 2009), and decision making (Sukirno & Siengthai, 2011). Previous studies found that performance indicators (PIs) in institutions of higher learning play an increasingly important role and Taylor (2001) argued that an effective way to improve the application of PIs is to obtain input from university academicians. The study found the reason for their dissatisfaction was due to the inability of current indicators to capture the dimensions of academic work, and preferring research to teaching.

Teaching and learning performance of academic staff in institutions of higher learning also involves measurement of supervision. As, Emilsson and Johnson (2007) and Mainhard et al. (2009) revealed the importance of the supervisor-student relationship to the success of a Ph.D. project. Therefore, information about doctoral students' perceptions of their relationship with their supervisor can be useful for providing detailed feedback to supervisors to improve the quality of supervision. In the finding, Emilsson and Johnson (2007) pointed out the five requirements of the trust, theories, tools, training and time in the supervisor-supervise relationship. Furthermore, there are other constructs has been studied on the relationship between participative decision-making and lecturer performance in institutions of higher learning (Sukirno & Siengthai, 2011). The finding shows that participative decision-making and academic rank have a significant effect on lecturers' performance. The finding indicates that involving lecturers in educational decision-making would be useful in improving not only the lecturer but also the institutional performance.

Kirkpatricks' Four Levels of Evaluation Model

Previous studies reveal specifically a few relevant researches related to performance evaluation. Most of the previous studies had used the Kirkpatrick's Four Levels of Evaluation models for measuring training effectiveness in various contexts and using different methodologies and tools for analysis (Bian, et al, 2015; Rouse, 2011; Chang, 2010; Praslova, 2010). According to Kirkpatrick's model, evaluation includes four levels: reaction level refers to trainees' perception of training; learning level focuses on the measurement of learning outcome in terms of knowledge and skills taught in the training program; behavioural and results level concerns the effects of training on work performance and

productivity gains of the whole organization, respectively. As Praslova (2010) realize, the model is rich, multilevel and finely take turns. It considers not only immediate, but also long-term effects.

Previous studies had also supported the usefulness of adoption and adaptation of Kirkpatrick's four levels of evaluation model (Peirera, et al, 2016; Bian, et al., 2015; Al-Yahya & Mat, 2013; Rouse, 2011; Lin, et al, 2011; Praslova, 2010; Chang, 2010; Jason, et al, 2008; Heidi, et al, 2004; Gomez, 2003; George et al, 1997). The model has been the most studied and useful for assessing workplace training and teaching effectiveness since its instigation in 1959 (Arthur et al, 2003; Praslova, 2010; Roos et al, 2014). For example, Praslova (2010) adapted Kirkpatrick's four levels of the evaluation model for training evaluation criteria and evaluation of learning outcomes in programs found in the institutions of higher learning. Evaluation of the effectiveness of education provides important feedback to institutions of higher learning (Peirera, et al, 2016) especially to external stakeholders (i.e., prospective students, parents, local government and regulatory entities, professional organizations and accreditation. However, selection of an appropriate indicator of the effectiveness of education can be a difficult task if the criteria are not well defined. An adaptation of the Kirkpatrick's model for higher learning helps to clarify the criteria and create plans for assessment of educational outcomes. The specific instruments and indicators accompany the corresponding criteria. It provides a rich context for understanding the role of various indicators in a variety of assessments.

In an empirical study, Rouse (2011) also suggested that an evaluation of the impact and effectiveness of courses is necessary. Therefore, improvements made based on the identification of strength and weaknesses. The study used Kirkpatrick's four levels of the evaluation framework to present a model that instructors can use to improve upon the standard-course evaluation form. The proposed course evaluation model addressed the first three of four levels and focused on the conditions necessary for the transfer of learning knowledge and skills into on-the-job applications. The study provides concrete tips so that health information management (HIM) instructors can apply them in the process of evaluating the effectiveness of their courses and programs. In addition, the empirical studies by Chang (2010) in the hospital industry also support the theory of Kirkpatrick's four levels of the evaluation. Hence, in order to expect organizational results, a positive change in behaviour (job performance) and learning must occur. The examination of Levels 2 and 3 helped to partially explain and predict the Level 4 results. The study examined Kirkpatrick's training evaluation model by assessing a sales training program conducted in an organization. The study assessed the employees' training outcome in aspects such as knowledge and skills, job performance, and the impact of the training upon the organization. The findings had supported the hypotheses in the study tested using the paired-samples t-tests, correlation, and hierarchical regression analysis.

The other example of empirical research by Lin, et al. (2011) also used Kirkpatrick's four-level model of evaluation to study the effect of organizational commitment on employee reactions to educational training. The study aimed to explore the causal relationship among golf club employees' training reactions, training, learning, training behaviours and organizational commitment (level 4) using Kirkpatrick's four-level evaluation model as the foundation of ideas and verification of the causal model. The study used the questionnaire survey method to gather samples in central Taiwan and used SEM for analysis and hypothesis testing of 494 respondents. SEM statistics show that the overall fit indexes had

passed the theory threshold. The findings showed that golf club employees' training reactions affected the result of organizational commitment through training learning and training behaviours. Thus, training, learning and training behaviours are mediating the overall causal model. The conclusion of the study is that golf club employees showed stronger learning achievements and learning behaviours when they had good reactions to education training.

Earlier, Heidi et al. (2004) had demonstrated the use of Kirkpatrick's Level 1 evaluation to assess the effectiveness of an instructional environment in which students learned to use a computer simulation tool to perform engineering design work. The evaluation triangulated with an instructor's assessment of student products generated as the students used the learning modules and designed an ice-cream manufacturing process using computer simulation tools. The tool frequently used in engineering design work and undergraduates are often trained to use these tools when they learn to design systems. Therefore, the use of new tools in the learning environment evaluated to assure that the students are able to use the tools effectively. Next, Buckly et al. (2009) suggested the need for an integrated approach to assessing the effectiveness of training by using the Kirkpatrick's four levels of evaluations; reaction, learning, behaviour and results. At the reaction level, students and trainers evaluated on how they react to the training, their attitudes about the training organization, the contents of the training and the methods employed. Learning is about the principles, facts and techniques learned by the students. The behaviour is the modifications made in job performance resulting from the training. In other words, how students apply their knowledge at the next level. Finally, assessing a result to find out the extent to which training has improved or influenced the performance of an organization (e.g. Reduced costs, improved quality/quantity, increased profits, etc.).

Based on the literature support, it is appropriate to use the first two levels of evaluation (reaction and learning) from the Kirkpatrick's Four Levels of Evaluation model to measure teaching and learning performance constructs in the context of higher learning institutions. Therefore, the specific purpose of this paper was to establish a psychometric property of scales measuring teaching and learning performance using Kirkpatrick's four levels of evaluation in filling the research gap.

Methods

There is a development of different methods in measuring individual and organizational performance. This study focuses on developing teaching and learning performance measures based on one of the duties and responsibilities of academic staff (unit of analysis) in higher education (Sukirno & Siengthai, 2011). Twenty-six items were developed using Kirkpatrick's Four Levels of Evaluation Model at Level 1 (Reaction) and Level 2 (Learning) as the basis of measurement for teaching and learning performance as shown in Table 1.

Table 1: Teaching Performance constructs in higher learning (Source: Kirkpatrick's Four Levels of Evaluation; Sukirno, Sununta Siengthai, 2011; Smeenk et al, 2009; Yu et al, 2009).

Measures of Teaching Performance	Item
1) Level 1 (Reaction)	• Explain the lesson's objectives. • Made it clear what students expected to learn in the lesson. • Covered all topics planned for the class session. • Made use of class time to teach the materials related to learning. • Present the lesson content in an orderly, clear and logical manner. • Use relevant examples to reinforce understanding of the lesson content. • Establish a connection between the theoretical and practical applications. • Used teaching aids effectively. • Stimulate students' interest in learning. • Interact with students. • Promote good class discussion. • Teach at a level / speed appropriate to the students' abilities. • Started and ended the class session on time. • The class session conducted in accordance with the official schedule. • Told students they could consult after class. • The objectives of the lesson achieved. • Show great interest in the lesson. • Participate actively in class discussions.
2) Level 2 (Learning)	• The topics covered are apparently relevant and be used in future subjects or work environments. • Understanding of the lesson improved the result of attending the class session. • Able to define the important concepts, principles, facts and techniques learned in the class. • Able to apply the concepts, principles, facts and techniques learned. • The knowledge has increased after attending the class session. • The skills improved the result of attending the class session. • The environment in the class had motivated students to learn and apply knowledge. • Expect to get a high grade in the course.

Expert review, pretesting and pilot study

The need for expert review in the field of study is important when designing the questionnaire. The questionnaire reviewed and validated by two groups of experts; one is an expert in measurement and evaluation (education) and three experts in human resource development (subject matter). These experts were not part of the population under study. The expert on subject matter consulted for a variety of the following reasons; the content of the questionnaire; importance and meaningfulness of measures to research objectives; wording and terminology of items; respondent/university identification – titles/roles of best respondent; sensitivity/threat of information request; language and format of the questionnaire. Next, pre-testing and pilot study implemented to complete the mission.

The purpose of pre-testing is to ensure that the questionnaire is valid and reliable, appropriate, necessary and sufficient for this study. The questionnaires, self-administered to 20 student respondents who drawn from the population of interest without a random sample. The pre-testing measures how much time it takes to complete each questionnaire and debrief the respondents after they had completed the questionnaire. The researcher analyse the information obtained during pre-testing to clarify the directions and the questionnaire revised

as necessary. According to Cooper and Schindler (2003), "An important purpose of the pre-test is to determine the participant's reactions to the questions." When using pre-designed and pre-verified research instruments, the pilot test strengthens the validity of the tool, thus providing the opportunity to ensure clarity. Later, the pilot test provides an opportunity to improve the questions before the actual test phase (Foreman, 2008).

Later, the researcher performs the pilot study to develop, adapt, or check the feasibility of the techniques, to determine the reliability of the measure, and/or to calculate how big the final sample needs to be. The data collected in the pilot test used a similar approach as that used for actual field study. Hundreds and ten samples of student respondents collected for the pilot study. Answers to the 26 items used a Likert scale with a range of 1 to 5, with five representing 'strongly agree' and one representing 'strongly disagree'. The results of the pilot study (as shown in Table 2) found that well-designed research directions of inquiry, layout, and period had a Cronbach Alpha score of more than 0.7. This score was for the construct of teaching performance (sTP) evaluated by students. However, before performing the reliability analysis, the exploratory factor analysis (EFA) executed, resulting in a few items scoring below 0.6 and need to delete. Hence, the survey questionnaire modified based on the EFA results and proceeded to the actual field study of data collection.

Exploratory confirmatory factor analysis (EFA)

The researcher analyses the pilot data using the principal component analysis (PCA) with the varimax rotation to test the reliability of measures of all items under teaching performance construct based on student evaluation. The results (Table 2) indicated that the Bartlett's Test of Sphericity was significant (Chi-square, p-value<0.000). The measure of sampling adequacy according to the Kaiser-Meyer-Oikin (KMO) standard was above 0.6. The KMO value close to 1.0 and the significance value of the Bartlett's Test close to 0.0 indicated that the data at hand was adequate in order to proceed with the reduction procedure by running the EFA in SPSS.

Table 2: KMO and Bartlett's Test for Students Responses on Academic Staff Teaching Performance

Constructs	Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO)	Bartlett's Test of Sphericity		Sig.
		Approx. Chi-Square	df	
Students Responses on Academic Staff Teaching Performance (sTP)	0.941	3324.112	325	0.000

As a result of EFA, 15 items under two rotated component matrix of academic staff teaching performance (sTP) evaluated by students with factor loadings above 0.6 were retained. Instead of dealing with 26 items, 11 items were excluded from further analysis, namely sTPR8, sTPR9, sTPR11, sTPR13, sTPR14, sTPR15, sTPL1, sTPL2, sTPL4, sTPL5 and sTPL11. Thus, EFA provides the new dimensions of measures for teaching performance at level 1 (Reaction) and level 2 (Learning) using Kirkpatrick's four levels of evaluation as shown in Table 3.

Table 3: The Rotated Component Matrix for Academic Staff's Teaching Performance (sTP) evaluated by students

Code	Students' Responses on Academic Staff's Teaching Performance (sTP) Items	Reaction (sTPR)	Learning (sTPR)
sTPR1	Explained the lesson's objectives.	.733	
sTPR2	Made it clear what students expected to learn in the lesson.	.748	
sTPR3	Covered all the topics planned for the class session.	.642	
sTPR4	Made good use of class time teaching the materials related to learning.	.647	
sTPR5	Presented the lesson content in an orderly, clear and logical manner.	.704	
sTPR6	Used relevant examples to reinforce understanding of the lesson's contents.	.694	
sTPR7	Linked theory and its practical application to the real work environment.	.659	
sTPR10	Interacted with students.	.640	
sTPR12	Taught at a level / speed appropriate to the students' abilities.	.601	
sTPL3	Participated actively in class discussions.		.610
sTPL6	Able to define the concepts, principles, facts and techniques learned in class.		.688
sTPL7	Able to apply the concepts, principles, facts and techniques learned.		.690
sTPL8	Knowledge increased after attending the class session.		.730
sTPL9	Skills on the subject improved because of attending the class session.		.831
sTPL10	The class environment motivated me to learn and apply knowledge.		.769

The next step, the researcher determines the reliability measure for the measuring items under each component of teaching and learning performance evaluated by students (sTP). The Cronbach's alpha of 0.6 or higher for each component under the construct of sTP provides a reliable measure of internal consistency. Table 4 shows that the values of Cronbach's alpha were above 0.6 for both components under the construct of sTP. As recommended by Nunally (1978) and cited by Awang (2012), it is important to note that the reliability measure has exceeded the minimum value of 0.6.

Table 4: Reliability Statistic for items of sTP evaluated by students

Construct	Component	Number of items in a component	Cronbach's Alpha	Cronbach's Alpha based on standardized items
Academic Staff Teaching Performance (sTP) Evaluated by Students	Reaction (sR)	9	0.903	0.904
	Learning (sL)	7	0.894	0.894

Field study

Finally, in the actual field study, the researcher distributes questionnaires and self-administered (self-completed) to 2250 students in the selected faculties (i.e., faculty of business / accounting / management / economics) in institutions of higher learning in Malaysia using multi sampling design. The researcher provides the questionnaire with instructions to complete it. The numbers of questionnaires returned by the respondents were 592, which represented a return rate of 26%. During the data collection process, the researcher gets the permission to collect data, from the deans of the various faculties of business, management, economics, and accounting, at the selective universities. The researcher also obtained the list of academic staff and students from the faculty used as the sampling frame for this study, together with the class schedules of the students. The researcher selects the respondents from the sampling frame using the random numbers generated by the computer. Then the researcher matches those numbers generated by the computer with the list in the sampling frame. The name of academic staff and students that had their numbers on the list and matched the numbers generated by the computer become the respondents in this study. Once the respondents identified, the researcher proceeded to the respective lecture room when the students were having their lecture sessions. The researcher obtained permission from the respective academic staff to distribute the questionnaires to the respective respondents. The respondents had responded to the questionnaire at their convenient time and submit the completed questionnaire using the sealed envelope to the representative of the faculty on the following day. In order to ensure confidentiality of the responses, each questionnaire enclosed with a cover letter that explained clearly the purpose of the study. Moreover, the respondents were not required to identify themselves on the questionnaire.

Results and Discussion

A measurement model: Confirmatory factor analysis (CFA)

The researcher analyses the measurement model using the confirmatory factor analysis (CFA) to testing the goodness of data. In this study, the researcher need to prove the uni-dimensionality and validity of the measurement model of teaching performance. The researcher assesses the uni-dimensionality of measures prior to assessing the validity and reliability. Hence, with the confirmatory factor analysis (CFA), the researcher removes any item that does not fit into its measurement model. The researcher instructs AMOS to calculate the factor loading and R^2 for every item. The researcher deletes the item having a factor loading less than 0.6 from the measurement model. During the item deletion process, the researcher deletes one item at a time, with the lowest factor loading. Then, the researcher re-specify and estimate the new measurement model and the process continues until the uni-dimensionality requirement achieved.

In a measurement model, a series of the goodness-of-fit index reflect the fitness of the model. Hair et al. (1995) and Holmes-Smith (2006) as cited by Awang (2012) recommend that the use of at least three fit indexes by including at least one index from each category of the model fit; absolute fit, incremental fit and parsimonious fit. Table 5 presents the information concerning the fitness index category, their level of acceptance, and comments. The index in bold is recommended since they are highly reported in the literature (Awang, 2012).

Table 5: Index Category and the Level of Acceptance for Every Index

Name of category	Name by index	Level of acceptance	Comments
1. Absolute fit	**RMSEA**	RMSEA < 0.08	Range 0.05 to 1.00 acceptable
Absolute fit	**GFI**	GFI > 0.90	GFI=0. 95 is a good fit
2. Incremental fit	**CFI**	CFI > 0.90	CFI=0. 95 is a good fit
3. Parsimonious fit	**Chisq/df**	Chisq/df < 5.0	The value should be less than 5.0

Source: Awang (2012)

Next, the researcher assesses the validity of the construct. In this study, validity is the ability of the instrument for the construct of academic staff teaching performance (sTP) evaluated by students to measure what is supposed to be measured. The following are three types of validity required for the construct in a measurement model; all items in a measurement model are statistically significant to prove convergent validity. The researcher verifies AVE (Average Variance Extracted) to assess convergent validity. The items in the sTP construct is explained to the average percentage of variation. Hence, AVE should be greater or equal to 0.5 (AVE>=0. 5). Construct validity is achieved when the fitness indexes of TLI=0.90 or higher, CFI=0.90 or higher, RMSEA=0.08 or lower and the ratio of Chisq/df are less than 5.0. Discriminant validity is achieved when all items in a measurement model of sTP are free from redundancy, as the MI (modification indices) in AMOS would report the pair of redundant items in the model.

Again, the researcher measures the reliability of the constructs. In this study, reliability is the extent of how reliable a particular measurement model is in measuring the latent construct of sTP. According to Gall (1996), "Reliability is the extent to which other studies would find similar results if they studied the same case using exactly the same procedures as the first researcher". The following criteria were assessed for the reliability of an instrument in this study; Internal reliability is achieved when the Cronbach's Alpha =0.6 or higher (calculated using SPSS). The Cronbach's alpha of 0.90s might indicate multicollinearity, meaning that the questions on the instrument are measuring the same thing and not different dimensions of the same variable. Construct Reliability is the measure of reliability and internal consistency of the sTP representing a latent construct. A value of CR >= 0.6 is required in order to achieve construct reliability. Average Variance Extracted is the percentage of variation explained by the items in a construct. A value of AVE greater than or equal to 0.5 is required (Table 6).

Table 6: The formula for computing AVE and CR

AVE = $\sum K2 /n$	**K= factor loading of every item and n=number of items in the model.**
CR=$(\sum K)^2/[(\sum K)^2 + (\sum 1-K^2)]$	

The fitness index for the sTP measurement model was extracted and is presented in Table 7. With approximately 90 percent confidence, the population RMSEA for the default model is between 0.069 and 0.084. Under the hypothesis of "close fit" (i.e., that RMSEA is no greater than 0.05 in the population), the probability of getting a sample RMSEA as large as 0.076 is 0.000. Since the entire factor loading of the remaining items, as shown in Figure 2, exceeded

the required value of 0.6, we could assume that the uni-dimensionality of the measurement model of academic staff's teaching and learning performance (sTP) evaluated by students has been achieved. Thus, the researcher was satisfied with the fitness indices. Figure 2 shows the remaining items because of CFA. There are fifteen items remaining for the final measurement model of sTP construct; nine items under level 1 (reaction) and six items under level 2 (learning).

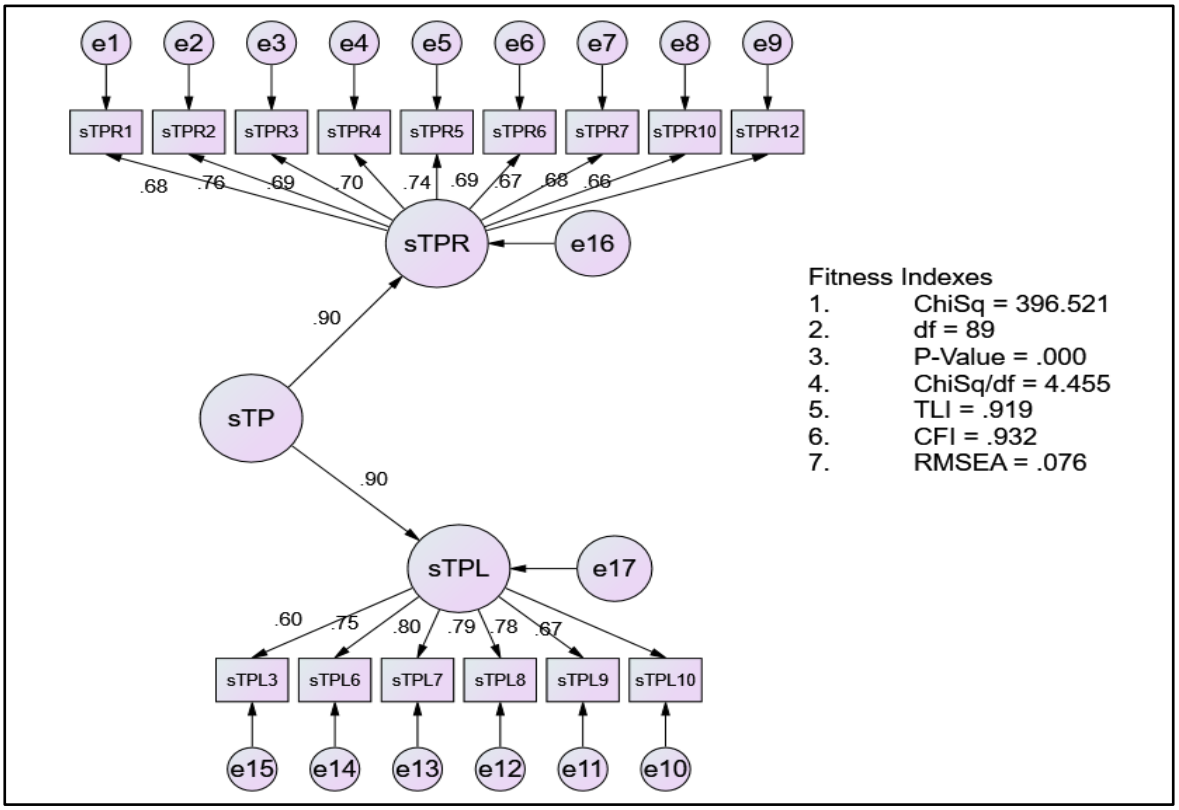

Figure 2: The factor loading of all items of Academic Staff's Teaching Performance evaluated by students (sTP) Construct

Table 7: The Summary of Fitness Indices for the Measurement Model of Academic Staff's Teaching Performance evaluated by students (sTP) Construct

Name of Category	Name on the Index	Index Value	Requirements
1. Absolute fit	RMSEA	0.076	Achieved
2. Incremental fit	CFI	0.932	Achieved
3. Parsimonious fit	Chi sq/df	2.491< 5.0	Achieved

Conclusion

In conclusion, the CFA assesses the unidimensionality, validity and reliability of the teaching and learning performance (sTP) measures evaluated by students. As a result, there are fifteen items remaining for the final measurement model of sTP construct; nine items under level 1 (reaction) and six items under level 2 (learning). Theoretically, the finding of this study offers a modification and new measures for the construct of teaching and learning performance based on student's evaluation. This new measure was based on the

measurement model (CFA) using SEM analysis. As the specific measure of teaching and learning performance developed based on the Kirkpatrick's four levels of evaluation model at level 1 (reaction) and level 2 (learning), the items could be referred as a standard for students' evaluations in measuring teaching and learning performance of academic staff in institutions of higher learning.

References

Al-Yahya, M., & Norsiah, M. (2013). Evaluation of effectiveness of Training and Development: The Kirkpatrick Model. *Asian Journal of Business and Management Science*, 2 (11): 14-24.

Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment centre dimensions. *Personnel Psychology*, *56* (July 1996), 125-154.

Bell, C. R. (1983). The High Performance Trainer. *Training & Development Journal*, 37 (6): 42.

Wu, B., Hu, Y. L., GU, X. Q., & Lim, C. P. (2016). Professional Development of New Higher Education Teachers with Information and Communication Technology in Shanghai: A Kirkpatrick's Evaluation Approach. *Journal of Educational Computing Research*, *54* (4), 531-562.

Buckley, Roger & Caple, Jim. (2009). *The theory & practice of training*. 6th Edition. The online version available at http://common.books24x7.com/toc.aspx?bookid=44862 (Viewed on 26 January 2013).

Breunig, K. J., & Hydle, K.M., (2013). Remote control: measuring performance for value. *Journal of Management & Governance*, 17:559–582.

Chang, Y. E. (2010). An empirical study of Kirkpatrick's evaluation model in the hospitality industry. Unpublished PhD Theses, Florida International University.

Chunping, J., Dengfeng, W., & Fan, Z. (2009). Personality traits and job performance in local government organizations in China. *Social Behavior & Personality: An International Journal*, 37 (4): 451-457.

Cooper, D. R. & Schindler, P. S. 2003. *Business research methods*. 8th Edition, Boston: McGraw Hill Irwin.

Emilsson, U., & Johnsson, E. (2007). Supervision of supervisors: on developing supervision in postgraduate education. *Higher Education Research & Development*, 26 (2): 163-179.

Foreman, S. M. (2008). Kirkpatrick model: Training evaluation practices in the pharmaceutical industry. Unpublished PhD Theses, Capella University.

Gaertner, H. (2014). Effects of student feedback as a method of self-evaluating the quality of teaching. *Studies in Educational Evaluation*, *42*, 91-99. Elsevier Ltd.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational Research.* 6th Edition. White Plains, NY: Longman Publishers USA.

George, M. A., Scott, I. T., Winston, B., Jr, Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, 50 (2): 341-358.

Gomez, A. K. (2003). An analysis of the evaluation practices of employer-sponsored training in the financial services industry. Unpublished PhD Theses, Texas A & M University.

Heidi, A. D., Samant, C., Tristan, E. J., & Debra O'Connor. (2004). Kirkpatrick's level 1 evaluation of the implementation of a computer-aided process design tool in a senior-level engineering course. *Journal of Engineering Education*, 93 (4): 321-331.

Hubball, H., & Burt, H. (2006). The Scholarship of Teaching and Learning: Theory–Practice Integration in a Faculty Certificate Program. Innovative Higher Education, 30 (5): 327-344.

James, J. M. (2010). Examination of the person approach to understanding personality-job performance relationships. Unpublished PhD Theses.

Jason, M. S., Edward, K. B., & Fortuna, A. (2008). Using importance-performance analysis to evaluate training. *Performance Improvement*, 47 (10): 30-35.

Jones, Michael J. & Sanghi, Steve. (2006). Driving excellence: how the aggregate system turned microchip technology from a failing company to a market leader. Available at http://common.books24x7.com/book/id_16777/book.asp (Viewed on May 14, 2013).

Lawless, K. a, & Pellegrino, J. W. (2007). Professional Development in Integrating Technology into Teaching and Learning: Knowns, Unknowns, and Ways to Pursue Better Questions and Answers. *Review of Educational Research*, *77*, 575-614.

Loveland, J., Gibson, L., Lounsbury, J., & Huffstetler, B. (2005). Broad and Narrow Personality Traits in Relation to the Job Performance of Camp Counselors. *Child & Youth Care Forum*, 34 (3): 241-255.

Lin, Y.T, Chen, S.C, & Chuang, H.T. (2011). The effect of organizational commitment on employee reactions to educational training: An evaluation using the Kirkpatrick four-level model. *International Journal of Management*, 28 (3): 926-938.

Mainhard, T., Rijst, R., Tartwijk, J., & Wubbels, T. (2009). A model for the supervisor–doctoral student relationship. *Higher Education*, 58 (3): 359-373.

Milanowski, A. (2011). Strategic Measures of Teacher Performance. *Phi Delta Kappan*, 92 (7): 19-25.

Naugle, K. A., Naugle, L. B., & Naugle, R. J. (2000). Kirkpatrick's evaluation model as a means of evaluating teacher performance. *Education*, *121* (1), 135.

Pereira, Diana, Assunção Floresb, Maria, Veiga Simãoc, Ana Margarida, & Barrosd, Alexandra (2016). Effectiveness and relevance of feedback in Higher Education: A study of undergraduate students. *Studies in Educational Evaluation, 49* (1), 7–14.

Praslova, L. (2010). Adaptation of kirkpatrick's four level models of training criteria to assessment of learning outcomes and program evaluation in higher education. *Educational Assessment, Evaluation and Accountability*, 22 (3): 215-225.

Rae and Leslie (2002). *Trainer Assessment: A Guide to Measuring the Performance of Trainers and Facilitators*. Abingdon, Oxon, GBR: Gower Publishing Limited.

Roos, C. I., Bowman, D. M. J. S., Balch, J. K., Artaxo, P., Bond, W. J., Cochrane, M., D'Antonio, C. M., et al. (2014). Pyrogeography, historical ecology, and the human dimensions of fire regimes. *Journal of Biogeography*. Blackwell Publishing Ltd.

Rothwell, William J. & Graber, Jim M. (2010). Competency-based training basics. Available at http://common.books24x7.com/book/id_37154/book.asp (Viewed on January 12, 2011).

Rouse, D. (2011). Employing Kirkpatrick's evaluation framework to determine the effectiveness of health information management courses and programs. Perspectives in Health Information Management, 1-1c. Available at http://search.proquest.com/docview/86833406 (Viewed on March 8, 2013).

Steinert, Y., Mann, K., Centeno, A., Dolmans, D., Spencer, J., Gelula, M. & Prideaux, D. (2006). A systematic review of faculty development initiatives designed to improve

teaching effectiveness in medical education: BEME Guide No. 8. Medical Teacher, 28 (6), pp. 497-526

Stiffler, Mark A. (2006). Performance: creating the performance-driven organization. Available at http://common.books24x7.com/toc.aspx? id=14019 (Viewed on January 12, 2010).

Sukirno, D. S., & Siengthai, S. (2011). Does participative decision making affect lecturer performance in higher education? *International Journal of Educational Management*, 25 (5): 494-508.

Taylor, J. (2001). Improving Performance Indicators in Higher Education: the academics' perspective. *Journal of Further & Higher Education*, 25 (3): 379-393.

Van Den Berg, P., & Feij, J. A. (1993). Personality traits and job characteristics as predictors of job experiences. *European Journal of Personality*, 7(5):337-357.

Van den Berg, P. T., & Feij, J. A. (2003). Complex relationships among personality traits, job characteristics, and work behaviours. *International Journal of Selection & Assessment*, 11(4): 326-339.

Yu, M., Hamid, S., Ijab, M., & Soo, H. (2009). The e-balanced scorecard (e-BSC) for measuring academic staff performance excellence. *Higher Education*, 57(6):813-828.

Awang, Z. (2012). *Research Methodology and Data Analysis*. 2nd Edition. UiTM Press.

Zwell & Michael. (2000). *Creating a culture of competence*. Available at http://common.books24x7.com/book/id_3784/book.asp (Viewed on January 1, 2011).

Acknowledgement: This paper was based on the Ph.D. scholarship research funded by the Ministry of Higher Education and UniSZA.