Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
EDU2012-0033

The Development of a Teaching Observation Evaluation Instrument for Improving Hospitality Teachers' Teaching Performance in Taiwan

Li-Yia Feng
Professor
Teacher Education Center, National Kaohsiung University of Hospitality and Tourism, Taiwan

Ya-Hui Su
Associate Professor
Teacher Education Center, National Kaohsiung University of Hospitality and Tourism, Taiwan

Chao-Chin Yang
Associate Professor
Department of Chinese Culinary Arts, National Kaohsiung University of Hospitality and Tourism
Taiwan

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209
Email: info@atiner.gr URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm

ISSN **2241-2891**

An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. The papers published in the series have not been refereed and are published as they were submitted by the author. The series serves two purposes. First, we want to disseminate the information as fast as possible. Second, by doing so, the authors can receive comments useful to revise their papers before they are considered for publication in one of ATINER's books, following our standard procedures of a blind review.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

Feng, Li-Yia, Su, Ya-Hui and Chao-Chin, Yang (2012) "**The Development of a Teaching Observation Evaluation Instrument for Improving Hospitality Teachers' Teaching Performance in Taiwan**" Athens: ATINER'S Conference Paper Series, No: EDU2012-0033.

[The Development of a Teaching Observation Evaluation Instrument for Improving Hospitality Teachers' Teaching Performance in Taiwan]

Li-Yia Feng
Professor
Teacher education Center, National Kaohsiung University of Hospitality and Tourism, Taiwan

Ya-Hui Su
Associate Professor
Teacher education Center, National Kaohsiung University of Hospitality and Tourism, Taiwan

Chao-Chin, Yang
Associate Professor,
Chinese Culinary Art Department, National Kaohsiung University of Hospitality and Tourism, Taiwan

Abstract

The construction of teaching observation instruments is necessary for hospitality education quality. When observing teaching is highly authentic and complicated, it often is not economical to observe. This study explored the possibility of achieving a balance between authenticity, reliability and economy in an observation instrument developed for hospitality teachers. The study included (1) suggesting two evaluators who ensure observation validity when considering the economical principle; (2) analysing the validity/reliability of the instrument as tested by comparing evaluators' short-term, professional observations with classroom students' long-term, non-professional observations; and (3) evaluating the effects of the evaluators' training, which is ultimately acceptable but still needs improvement.

Keywords: teaching observation, evaluation instrument, hospitality education

Acknowledgements: The authors would like to thank the National Science Council and the Ministry of Education, R.O.C., for their support of this work through the grant NSC96-2516-S-328-002-MY2 and 100L005-12, respectively.

Contact Information of Corresponding author: liyiafeng@gmail.com

Teacher Evaluation in Taiwan

The Taiwanese government has prioritised the hospitality industry for national development. Public and private sector interest has fuelled the economic development of the hospitality industry, as evidenced by the considerable investment and dramatic transformations in this business (Ministry of Education, 2008). This emphasis on the hospitality industry's growth has stimulated the development of hospitality education courses and programs in the vocational education sector. Thus far, the population of hospitality teachers has become the second largest group of teachers in Taiwan (Ministry of Education, 2007), and a quality hospitality education results from having highly skilled professional teachers. Therefore, establishing a relevant teacher evaluation system for professional development, which includes the construction of evaluation criteria and instruments, is necessary to ensure the quality assurance of hospitality education.

Discourses on quality have been part of the educational policy for the last 20 years, both globally and in individual nations (Thomas, 2008). One of the keys to achieving quality education is the promotion of teachers' professional competence. The Taiwanese government has emphasised the importance of using teacher evaluations for professional development within the K-12 system and has attempted to create evaluation administrative regulations and funding subsidy conditions (Ministry of Education, 2006a). In 2006, 16 high schools and vocational schools (which account for 5% of Taiwan's high schools) joined this evaluation on a trial basis. By 2011, 293 high schools and vocational schools (representing 90% of Taiwan's high schools) had enrolled in the evaluation (Ministry of Education, Taiwan, 2011).

Globally, the majority of the evaluators assessing teachers are principals and assistant principals. Examples of this practice are found in the USA, the UK, Japan, and Singapore (Bridges & Groves, 1999; Danielson, 2002; Steiner, 2010; Yu, 2011). Some research has shown that even though administrators spend a significant amount of time on evaluations (more than 10 hours per teacher each year), they cannot effectively use these evaluations to improve teaching and learning (Schomker, 2001; Kersten & Israel, 2005). Many nations and school districts have connected teacher evaluations and professional development, often through the use of self-evaluation tools and peer observation, including systematic classroom observations and professional meetings among teachers (Ovando, 2001; Kassabian, 2009a). Some research has found that the accountability portion of teacher evaluations is invalid (Kersten & Israel, 2005; Marshall, 2005) and thus suggested that the top-down bureaucracy control model should be abandoned in favour of peer evaluation as an alternative approach (Mayo, 1997; Goldstein, 2007). Many teacher evaluation studies in Taiwan have indicated that peer evaluation is the most appropriate approach (Ministry of Education, 2006b; Pan, Wang, Chang & Lin, 2007; Feng, Yang & Su, 2009).

Taiwan's teacher evaluation is school-based and lacks a uniform set of rules for the evaluation instrument, process, and number of evaluators. For example, the number of evaluators for an observation can range from one to six (Feng, Yang & Su, 2009), but the participating teachers being evaluated rely less on trust and fairness, which are key to professional dialogue and growth between peers, and focus more on accuracy, validity, and practicality. As a result, the teacher performing the evaluation requested an advanced training program for evaluators and also asked for evaluation instruments that consider

differences when evaluating, for instance, science teachers and liberal arts teachers, whose teaching practices may differ because of their varied subject matter (Feng, 2010; Tseng, 2011).

This study explored the possibility of achieving balance between the authenticity, reliability and economy in an observation instrument for hospitality teachers. The study included (1) investigating the number of evaluators to ensure observation validity when considering the economical principle; (2) analysing the validity/reliability of the instrument by comparing evaluators' short-term, professional observations with classroom students' long-term, non-professional observations; and (3) evaluating the effects of evaluator training.

Literature Review

Evaluation Instruments

Danielson & McGreal (2000) claim that teacher evaluation criteria should be included in the domains of teaching practices. For example, teacher performance evaluation in Texas has eight domains: (1) active, successful student participation in the learning process; (2) learner-centred instruction; (3) evaluation and feedback on students' progress; (4) management of student discipline, instructional strategies, time and materials; (5) professional communication; (6) professional development; (7) compliance with policies and requirements; and (8) improvement of academic performance of all students within the class (Ovando, 2001). Ten standards were constructed for the Malden Public Schools Teacher Performance Evaluation. The standards refer to (1) planning skills; (2) instructional skills; (3) classroom management skills; (4) motivation skills; (5) material that is current and consistent with the curriculum; (6) promotion of equality and respect for diversity; (7) operational duties; (8) professional relationships; (9) relationships with parents and community; and (10) teachers' own professional development (Kassabian, 2009b).

The California Standards for the Teaching Profession, as modified for use in the San Francisco United School District, includes the following standards: (1) engaging and supporting all students' learning; (2) creating and maintaining an effective environment for learning; (3) understanding and organising subject matter knowledge; (4) planning, designing, and delivering learning experiences for all students; (5) assessing student learning; and (6) developing as a professional educator (Louw, 2006). The Professional Standards for Teachers proposed 15 criteria for performance management arrangements in UK. These criteria fit into three categories as follows: (1) attributes, which include relationships with children and young people, frameworks, communicating and working with others, and professional development; (2) knowledge and understanding, which includes teaching and learning, assessment and monitoring, subjects and curriculum, literacy, numeracy, information communication technology, achievement and diversity, and health and well-being; and (3) skills, which include planning, teaching, assessing, monitoring and giving feedback, reviewing teaching and learning, the learning environment, teamwork and collaboration (TDA, 2007). The standards of teacher evaluation in Japan are classroom

management, teaching of subjects, counselling, professional development, and school duties (Miyagi, 2001).

In Taiwan, the Ministry of Education (2006a) proposed 4 teacher evaluation domains: curriculum design and teaching, classroom management and counselling, research development and further studies, and commitment and attitudes to teaching.

Based on the above descriptions, teacher evaluations from different countries typically include four characteristics: (1) they focus on important teaching skills, (2) they are linked to teacher development; (3) classroom observation is the most popular evaluation instrument used; and (4) the evaluation standards and criteria are not subject-specific. Because there are many differences among education systems in Taiwan, evaluation instruments should include items from general teaching categories and specific evaluation items for each discipline (Ministry of Education, 2006b).

The educational goals, curriculum, learning contexts, and student background of vocational schools are not the same as those of elementary and general high schools. In 2008, Feng, Su, and Yang created a survey of teaching evaluation tools used in Taiwan, including tools that measure the behaviour of tourism and hospitality teachers. Two teaching observation checklists were developed, one for the theory curriculum and the other for the hands-on curriculum. The two instruments are widely used in Taiwan's elementary and high schools (accounting for 20% of Taiwan's elementary and high schools) (Feng, 2011).

In the field of vocational education, a variety of terms describe student experiences, including internships and practicums (Busby, 2003). Hassanien (2006) and Lugosi (2010) noted that group work remains a fundamental component of hospitality education. The validity and reliability of the evaluation instruments needs to be studied, as does the number of evaluators, especially considering the evidence connecting teacher performance and students' learning conditions. This necessity remains true while observing teachers who are implementing a practical, hands-on curriculum in a large laboratory room.

Evaluator Training

Bandura's self-efficacy theory suggests that evaluators' beliefs in their own efficacy regarding their evaluation tasks may affect their motivation to perform their supervisory tasks. Painter (2000) found that school principals place a high value on the task of evaluating low-performing teachers and that they believe that they are well equipped to do so. The data also suggest that principals with long-term training believe that they are adequately addressing the problem of poor performance.

Milanowski & Heneman (2001) surveyed teacher reactions to a standards-based teacher evaluation system in a medium-sized Midwestern school district in Wisconsin. Thirty-six teachers received three days of training to be evaluators for the new peer evaluation system. This report suggested that this training was necessary to enhance the teachers' understanding of the evaluation system and to improve their skills. Teachers viewed the evaluators as capable and objective.

In Taiwan, basic training for evaluators began in 2006, and advance training began in 2009. K-12 teachers all receive the same training. Feng, Chang, & Lin (2007) indicated that

the basic evaluator training can improve an evaluator's competence and acceptance of the evaluation system.

However, reports by Gau & Chiang (2009) and Pan, Wang, Chang & Lin (2007) indicated that there was a moderate level of acceptance of the evaluators along with some concern regarding the evaluators' lack of subject-matter knowledge. Additionally, some teachers questioned the consistency among evaluators in terms of the process used and the scores assigned. Wang (2009) suggested that the classroom observation training was insufficient to provide an adequate assessmen. Before implementing a teaching observation evaluation instrument in vocational schools, the Evaluator Training Centre in Kaohsiung City, Taiwan agreed to a pilot test in six vocational schools. The test schools were chosen by the evaluator training centre based on administrator interest. Of the 72 teachers in the six test schools who had been recommended to receive advanced evaluator training, 6 were hospitality teachers. Based on the previous evaluator training studies, the advanced training was 15 hours long. The initial training sessions described the domains and standards, reviewed the evaluation forms, taught how to observe and coach and included practice evaluations for both theory curriculum and hands-on curriculum using videotaped vignettes of teacher performance. Later sessions covered rating bases, scoring portfolios, and comparing their scores with expert raters.

Methodology

The purpose of this research is to develop a teaching observation evaluation instrument that balances authenticity, reliability and economy while considering the effects of evaluator training.

Samples

As mentioned previously, the evaluator training centre selected 6 vocational schools for the pilot test during the 2009-2010 school year. All 72 teachers at the 6 schools were selected to be peer evaluators.

We were able to survey 60 of the 72 peer evaluators. In addition, we invited the 6 hospitality teachers to participate in the experimental study to test the reliability and economy of the teaching observation evaluation instrument. After a practice evaluation of four videotaped vignettes of teaching performance, the 6 hospitality teachers used the two teaching observation checklists (one for theory curriculum and one for hands-on curriculum) that were developed by Feng, Su, and Yang to observe and evaluate three teachers' classroom teaching videotape. The teachers' performances were labelled by each evaluator as either 'recommended', 'pass', or 'improvement needed'. This study collected the 6 evaluators' written reports to understand the effects of evaluator reliability and the number of evaluators.

To test validity, we invited 42 hospitality teachers to video record their teaching performances of one teaching unit, based on their principals' recommendation and assistance. Students in the observed classes were asked to complete 'Student Ratings of Teacher Performance' after the teaching unit, which had question items similar to those in

the teaching observation checklists but rewritten for students. The basic data of observed teachers are presented in Table 1.

Survey Items

Survey items were constructed to assess the overall evaluators' efficacy beliefs about peer evaluation, need for professional development, acceptance of peer evaluation, growth of observation skills and feedback skills, and acceptance of the standards and instruments. Items were developed by collecting a bank of over 66 items from prior studies on employee reactions to evaluators' training and then adapting some of those items and writing some new ones. Items were evaluated using 5-point Likert scales, which ranged from 'strongly disagree' (1) to 'strongly agree' (5) (see Table 7).

Analysis

The Aiken's homogeneity coefficient and generalisability coefficient were conducted to test evaluator reliability. Criterion-related validity was tested by comparing evaluators' short-term, professional observations with classroom students' long-term, non-professional observations. The survey analysis included the calculation of scale, means and standard deviations.

Development of a Teaching Observation Evaluation Instrument

Evaluator reliability

This study used the homogeneity reliability coefficients that were proposed by Aiken to determine the absolute consistency of teaching assessments that were provided by six evaluators. According to Shavelson and Webb (1991), by using Aiken homogeneity reliability coefficients that are based on clinical application standards, a value between 0.75 and 1.0 indicates that an evaluator is reliable. Table 2 shows the results of the reliability analysis of the assessments that were provided by the six evaluators. The reliability coefficients between the ratings of each individual evaluator and the average of all of the other evaluators were 0.77, 0.76, 0.81, 0.96, 0.96, and 0.96. They all fell within a range of values that indicate high reliability. The correlations for evaluators A and B were the smallest (0.77 and 0.76), indicating that their opinions differed from other evaluators. Therefore, they are less suitable to serve as evaluators for a formal study. Four indicators had a homogeneity coefficient of H=1, indicating identical assessments by each evaluator. The homogeneity coefficients for the three remaining indicators were H=0.72, indicating that the evaluations were similar, as shown in Table 2.

Generalizability Coefficient

This study used a generalisability study (G) and a decision (D) study in the field of generalisability theory to determine the number of evaluators that are required to fairly and

economically evaluate teacher performance using observational methods. The 'face of differentiation' of this study was the in-class performance of teachers, which had been previously recorded. The measurement tools, also known as the 'face of instrumentation,' were evaluators and a checklist of the seven standards of teaching. Four evaluators had similar homogeneity coefficients; two of which were randomly selected to evaluate the teaching performance of three hospitality teachers. The teachers' performances, which were previously recorded, were evaluated using a checklist of the seven standards of teaching. From the evaluation scores, we found that a generalisability coefficient that is greater than 0.8 was achieved using only two evaluators, indicating that the criteria of fairness and economic efficiency were met, as shown in Tables 3, 4, and 5.

The notations that are related to the calculation of the generalisability coefficient are defined as the following:

σ_p^2 is the variance of the teaching performances.

σ_r^2 is the variance of the evaluation scores.

σ_i^2 is the variance of the indicators, i.e., the seven standards of teaching.

σ_{pr}^2 is the covariance between the teaching performances and the evaluation scores.

σ_{pi}^2 is the covariance between the teaching performances and the indicators.

σ_{ri}^2 is the covariance between the evaluation scores and the indicators.

σ_e^2 is the covariance between the teaching performances, the evaluation scores, and the indicators.

The variance components (σ) are given in Table 4 and were calculated by setting the mean square (MS) as equal to the expected mean square (EMS). The MS and EMS for each variance component are given in Table 3, in which nr is the number of evaluators and ni denotes the number of evaluation indicators. The error variance component listed in Table 4, $\sigma_e^2 = 0.133$, is trivially obtained from Table 3, in which MS = 0.133 = EMS = σ_e^2. The covariance between the teaching performances and the indicators (p×i) is 0.363 (0.860 = σ^2_e + nr σ_{pi}^2; 0.860 = 0.133 + 2 σ_{pi}^2, σ_{pi}^2 = 0.363), and the teaching performances variance component is 0.382 (6.328 = σ^2_e + nr σ^2_{pi} + ni σ^2_{pr} + ni × nr σ_p^2; 6.328 = σ^2_e + 2 × 0.363 + 7 × 0.018+ 7 × 2 σ_p^2; σ_p^2 = 0.382).

A D study based on the variance components previously calculated in the G study is presented in Table 5. The variances of the two evaluators and seven indicators design are as follows: teaching performances = 0.382 (0.382/1=0.382), evaluation = 0.005 (0.01/2=0.005), standard = 0.020 (0.141/7 = 0.020), p×r = 0.009 (0.018/2 = 0.009), p×i= 0.052 (0.363/7 = 0.052), r×i=0(0/14 = 0), and residual = 0.010 (0.133/14=0.010).

According to classic test theory, the reliability coefficient is defined as $\sigma_{XT}^2 / \sigma_{XX'} = \sigma_T^2 / (\sigma_T^2 + \sigma_e^2)$. For the generalisability theory, the reliability coefficient is obtained by the following substitutions: $\sigma_T^2 = \sigma_p^2$, σ_{XT}^2 is replaced by either the G coefficient or ϕ, $\sigma_{Rel}^2 = \sigma_e^2$, and because $\sigma_{Rel}^2 = \sigma_{pr}^2 / nr' + \sigma_{pi}^2 / ni' + \sigma_{pri,e}^2 / (ni'nr')$, $\sigma^2_{Abs} = \sigma^2_{Rel} + \sigma_r^2 / nr' + \sigma^2_i / ni' + \sigma^2_{ri} / (ni'nr')$. Therefore, σ^2_{Rel} = 0.071 (0.018/2 + 0.363/7 + 0.133/(2×7) = 0.009 + 0.052 + 0.010 = 0.071), σ^2_{Abs} = 0.096 (0.071 + 0.010/2 + 0.141/7 = 0.071 + 0.005 + 0.020 = 0.096), G coefficient = 0.843(0.382/(0.382 + 0.071)), and ϕ coefficient = 0.799 (0.382/(0.382+0.096)).

In Table 5, nr` and ni` represent the number of evaluators and the number of indicators, respectively. Because this study used the seven standards of teaching as evaluation

indicators and none of them could be eliminated, the variance component was fixed as seven in the D study. The index of dependability, ϕ, indicates the consistency between the evaluation scores given by all of the evaluators for this study and the scores given by all of the qualified evaluators. The G coefficient shows the consistency of the relative scores given by the evaluators for this study and those given by all of the qualified evaluators. Therefore, the G coefficient is larger than ϕ. By analysing the estimated variance components in Table 5, we found that the G coefficient for the one evaluator design was 0.811, which is consistent with the actual scores that were given by all of the qualified evaluators. The consistency of the absolute scoring was 0.762. Therefore, in the future, the formal study can be conducted accurately using only one evaluator.

Criterion-Related Validity

When considering the real use and objectivity of the teaching observation instrument, this study focuses on inter-judge reliability.The inter-judge reliability is investigated by comparing short-term evaluator observations (by an external evaluator) with long-term participant observations (by students within the class). The results of the canonical correlation are shown in Table 6 and Figure 1.

Table 6 shows the canonical correlations between the views of students and evaluators using two canonical factors and two canonical coefficients. The correlation coefficients (ρ=0.91 and 0.75) achieved a significance level of 0.05, and the Λ values were 0.08 and 0.43. The canonical variables, $\chi 1$ and $\chi 2$, for the student ratings of teacher performances explained 82% and 57% of the total variance of the canonical variables, $\eta 1$ and $\eta 2$, for the results that were obtained using the teaching observation checklist. The variables $\eta 1$ and $\eta 2$ explained 92% and 13% of the total variance for the teaching observation checklist scores. The overlap indices on $\eta 1$ and $\eta 2$ between two subscale scores of the student ratings and two subscale scores of the teaching observation checklist were 0.75 and 0.07, for a total of 0.82. It was found that two subscale scores for the student ratings, through two sets of canonical factors, explain 82% of the total variance of two subscale scores for the teaching observation checklist. Thus, the degree of similarity between the scale scores is 0.82, which is very high. These results are similar to those of several previous studies (Feng, 2003; Jewell, 1990), meaning that the results obtained by qualified evaluators using effective teaching observation evaluation instruments are very similar to students' views that are formed by spending a long period of time in the classroom. In other words, systematic observational evaluations by well-trained evaluators are highly correlated with evaluations by students.

Evaluators Training

Referring to the descriptive statistics (Table 7), the average scores show that the effects of the evaluators' training were all above the neutral point of 3.0, including the overall evaluators' efficacy beliefs about teacher evaluation, need for professional development, understanding of peer evaluation, growth of observation skills and feedback skills, and acceptance of the standards and instruments.

Discussion and Implications

This study developed teaching observation checklists for assessing hospitality teachers in Taiwan. The two checklists—one for the theory curriculum and the other for the hands-on curriculum—incorporate items for evaluation based on the reliability and validity as determined by this study. The findings of the study conform to the reliability and validity expectations of the instruments.

Based on the presented analysis, each observation requires one evaluator. However, the data from evaluators' training surveys suggest that the evaluators lack the confidence to complete an evaluation independently. Due to the unique field requirements for teaching writing courses, we suggest that two evaluators assess writing courses. Training is necessary to improve the evaluators' competence (efficacy beliefs; observation skills; feedback skills; and understanding of peer evaluation, standards, and instruments) to achieve a high standard of evaluator reliability. Although these evaluation instruments are constructed specifically for hospitality teachers in Taiwan, the instruments may be useful in evaluating vocational teachers in Taiwan and hospitality teachers in other countries who are open to changes in a contextual application. Placing peer observation at the centre of teacher evaluation and providing inclusive evaluation criteria that are applicable to the theory curriculum and the hands-on curriculum can be valuable for developing teacher evaluation in other countries. We suggest the following directions for future research:

1. The credibility, validity, and framework of the teaching observation checklist for the theoretical curriculum and the hands-on curriculum should be tested regularly. The teaching observation checklist used in this study is credible and valid. The predictive validity can be used to determine if the evaluations accurately predict the teaching efficacy of hospitality teachers. Cross-validation can be used to regularly test the validity.

2. The views of teachers and their supervisors on the teaching observation checklist for the theoretical curriculum and the hands-on curriculum should be explored further. We introduced the teaching observation checklist to more than 360 teachers in 10 vocational schools in Southern Taiwan. We found that a majority of schools require two evaluators, whereas some schools allowed a single evaluator. These practices agree with the conclusions that were drawn from this study. However, the effect of two evaluators on the evaluation results and the execution of the evaluation process must be better understood.

3. The appropriate environments for implementing the teaching observation checklists for theory curriculum and hands-on curriculum must be identified. In addition, any restrictions on using the checklist should also be analysed. Every instrument has its restrictions, and the instruments from this study cannot be applied to every teaching environment.

4. A norm and scoring standard must be established. A passing score for teachers who are evaluated should be determined, along with whether passing is based on the total score or the lowest criteria of "the standards". This topic should be explored in future research.

5. Other recommendations include regular evaluations of the evaluation methods and tools and adopting lessons that are learned from schools and teachers during implementation. This study used full-time hospitality teachers as study subjects due to

the restrictions on the researchers' capabilities, time, and finances. For future research, intern hospitality teachers and substitute teachers can be added as study subjects to determine the applicability of evaluation tool.

Table 1. Basic Data for the Criterion-Related Validity of the Samples

Teacher Experience		n	%	Teacher Experience		n	%
Length of Teaching Time	Less than 3 years	3	7.1	Department	Tourism	12	28.6
	3-10 years	25	59.6		Food and Beverage	30	71.4
	More than 10 years	14	33.3	Industry Experience	Yes	27	4.8
Educational Level	Associate Degree	2	4.8		No	15	69.0
	Bachelor	29	69.0	Gender	Male	10	23.8
	Graduate	11	26.2		Female	32	76.2

Table 2. The Reliability Analysis of Observational Teaching Assessments by Six Evaluators

	Evaluators						
Indicators	A	B	C	D	E	F	Homogeneity Coefficient (H)
A-1	3	3	3	3	3	3	1.00
A-2	3	2	2	2	2	2	0.72
A-3	3	2	3	3	3	3	0.72
A-4	2	2	2	2	2	1	0.72
A-5	3	3	3	3	3	3	1.00
B-1	3	3	3	3	3	3	1.00
B-2	3	3	3	3	3	3	1.00
Corrected Item-Total Correlation	0.77	0.76	0.81	0.96	0.96	0.96	

Table 3. Analysis Results of Measured Variances.

Source of Variance	SS	df	MS	EMS
Teacher (p)	12.656	2	6.328	$\sigma^2_e + nr\,\sigma^2_{pi} + ni\,\sigma^2_{pr} + ni \times nr\,\sigma_p^2$
Evaluator (r)	0.341	1	0.341	$\sigma^2_e + np\,\sigma^2_{ri} + ni\,\sigma^2_{pr} + np \times ni\,\sigma_r^2$
Indicator (i)	5.792	6	0.965	$\sigma^2_e + np\,\sigma^2_{ri} + nr\,\sigma^2_{pi} + np \times nr\,\sigma_i^2$
pxr	0.518	2	0.259	$\sigma^2_e + ni\,\sigma_{pr}^2$
pxi	10.323	12	0.860	$\sigma^2_e + nr\,\sigma_{pi}^2$
rxi	0.457	6	0.076	$\sigma^2_e + np\,\sigma_{ri}^2$
Error (pri)	1.601	12	0.133	σ^2_e

Note: np: (Teachers) = 3; nr: (Evaluators) = 2; ni: (Indicators) = 7.

Table 4. Estimation Results of Measured Variance Components.

Source of Variance	MS	σ	%
Teacher (p)	6.328	0.382	36.485
Evaluator (r)	0.341	0.010	0.955
Indicator (i)	0.965	0.141	13.467
pxr	0.259	0.018	1.719
pxi	0.860	0.363	34.670
rxi	0.076	-0.057	0
Error (pxrxi)	0.133	0.133	12.703

Note: *negative values are set to zero

Table 5. Estimation of Variance Components for the G and D studies.

Source of Variance	Estimation of variance components for the G Study	Estimation of variance components for the D study			
	nr'=1	1	2	3	4
	ni'=1	7	7	7	7
Teacher (p)	0.382	0.382	0.382	0.382	0.382
Face of differentiation					
Evaluator (r)	0.010	0.010	0.005	0.003	0.003
Indicator (i)	0.141	0.020	0.020	0.020	0.020
pxr	0.018	0.018	0.009	0.006	0.005
pxi	0.363	0.052	0.052	0.052	0.052
rxi	0	0	0	0	0
Error (pxrxi)	0.133	0.019	0.010	0.006	0.005
$\sigma\delta_{Rel}^2$	0.514	0.089	0.071	0.064	0.062
σ_{Abs}^2	0.665	0.119	0.096	0.087	0.085
G coefficient (relative)	0.426	0.811	0.843	0.856	0.860
ϕ coefficient (absolute)	0.365	0.762	0.800	0.814	0.812

Notes:

$\sigma_{Rel}^2 = \sigma_{pr}^2/nr' + \sigma_{pi}^2/ni' + \sigma_{pri,e}^2/(ni'nr')$

$\sigma_{Abs}^2 = \sigma_{Rel}^2 + \sigma_r^2/nr' + \sigma_i^2/ni' + \sigma_{ri}^2/(ni'nr')$

nr` = Number of evaluators

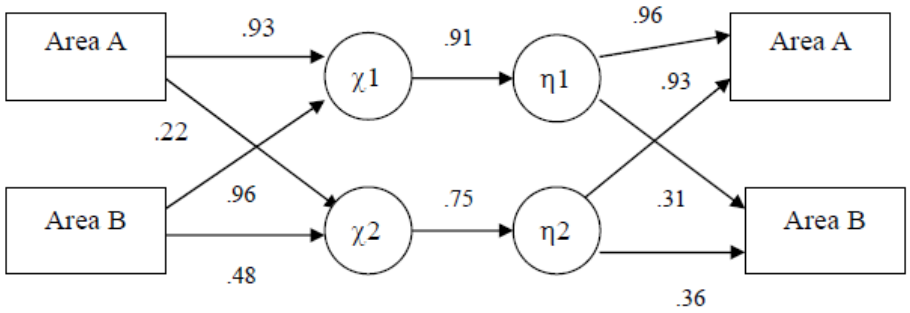
ni` = Number of Indicators

Table 6. Canonical Correlation between 'Student Ratings of Teacher Performance' and 'Teaching Observation Checklist'

Covariates Variables (students' ratings)	Canonical Factor		Dependent Variables (evaluators' observations)	Canonical Factor	
(X Variable)	$\chi 1$	$\chi 2$	(Y Variable)	$\eta 1$	$\eta 2$
Curriculum Design and Teaching	0.93	0.22	Curriculum Design and Teaching	0.96	0.31
Classroom Management and Student Counselling	0.96	0.48	Classroom Management and Student Counselling	0.93	0.36
Variance in covariates explained by canonical variables	0.90	0.21	Variance in dependent explained by canonical variables	0.92	0.13
Variance in dependent variables (%)	0.74	0.12	Variance in canonical variables (%)	0.75	0.07
ρ^2	0.82	0.57			
ρ	.91***	0.75***			

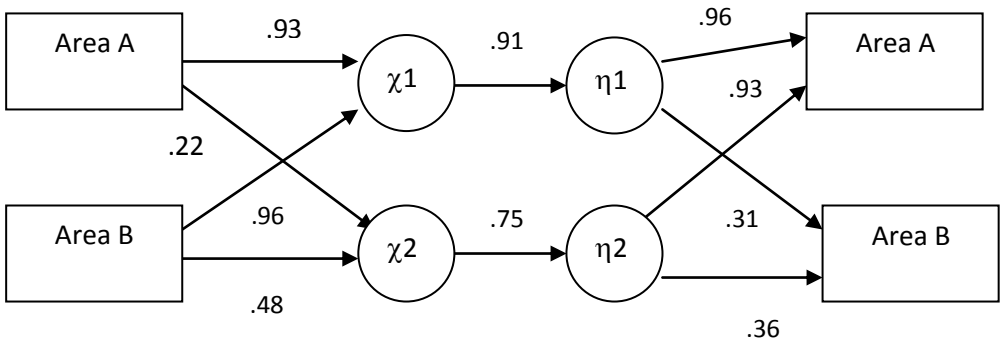
*** P<.001

Figure 1. The 'Canonical Correlation' Route of 'The Teaching Effectiveness Student Scale' and 'The Teaching Observation Checklist'



Note: Area A, Curriculum Design and Teaching; Area B, Classroom Management and Student Counselling

Figure 1. The 'Canonical Correlation' Route of 'The Teaching Effectiveness Student Scale' and 'The Teaching Observation Checklist'



Note: Area A, Curriculum Design and Teaching; Area B, Classroom Management and Student Counselling

Table 7. The Effects of the Evaluators' Training

The Effects of the Evaluators' Training	Mean	Std. Dev.
Overall evaluators' efficacy beliefs about teacher evaluation	3.61	0.34
Need for professional development	3.75	0.48
Understanding of peer evaluation	3.91	0.47
Growth of observation skills and feedback skills	4.39	0.12
Acceptance of the standards and instruments	4.01	0.34

References

Bridges, E. M., & B. R. Groves (1999). 'The macro-and micropolitics of personnel evaluation: A framework.' *Journal of Personnel Evaluation in Education* 13(4): 321-337.

Busby, G. (2003). 'Tourism degree internships: a longitudinal study.' *Journal of Vocational Education &Training* 55: 319-334.

Busby, G., P.Brunt, & S. Baber (1997). 'Tourism sandwich placements: An appraisal.' *Tourism Management* 18: 105-110.

Danielson, C. (2002). *Enhancing student achievement: A framework for school improvement*. Alexandaria, VA: ASCD.

Danielson, C., & T. L. McGreal (2000).*Teacher evaluation to enhance professional practice*. Alexandria,VA: ASCD.

Feng, L.Y. (2003). 'A study of the relationships among teacher self-evaluation, classroom observation, and student ratings of teaching effectiveness.' *Journal of National Taipei Teachers College* 6(1): 201-228. [In Chinese].

Feng, L.Y. (2010). *Hospitality teacher evaluation system for professional development in vocational school*. Kaohsiung: Fu-Won. [In Chinese].

Feng, L.Y. (2011). 'A study of the effects of the training courses of advanced evaluators in teacher evaluation: Kaoshiung City, Kaoshiung County, and Pingtong County Cases.' *Journal of Research on the elementary education* 8: 11-25. [In Chinese].

Feng,L.Y., S. J. Chang, & Y. F. Lin (2007). 'A study of teacher evaluator's training: A case of Kaohsiung City.' In: National Teacher Education Association (ed.), *Teacher evaluation and professional development,* 61-92. Taipei: Psychology. [In Chinese].

Feng, L. Y., C.C. Yang, & Y. H. Su. (2009). *A menta-evaluation of teacher evaluation for professional development in Kaohsiung city*. Kaohsiung: Education Bureau. [In Chinese].

Feng,L.Y., Y. H. Su,& C.C. Yang (2011). 'The construction of professional evaluation instruments for hospitality teachers in Taiwan's vocational schools.' *Journal of Teaching in Travel & Tourism* 11(3): 229-252.

Gau, S. J., & C.C. Chiang (2009). 'A meta-evaluation of teacher evaluation for professional development in elementary schools.' Paper presented at International Conference on Teacher Evaluation with Professional Growth Cycle. December 3-4, in Kaohsiung,Taiwan. [In Chinese].

Goldstein,J.(2007). 'Easy to dance to: Solving the problems of teacher evaluation with peer assistance and review.' *American Journal of Education* 113 (May): 475-508.

Hassanien, A. (2006). 'Student experience of group work and group assessment in higher education.' *Journal of Teaching in Travel & Tourism* 6(1): 17-39.

Jewell, L. R. (1990). 'Evaluating teaching effectiveness of secondary vocational educators.' *Journal of Vocational and Technical Education* 6(2): 3-17.

Kassabian, N. (2009a). 'The professional growth cycle in Malden public schools: the standards.' Paper present at International Conference on Teacher Evaluation with Professional Growth Cycle, December 3-4, in Kaohsiung, Taiwan.

Kassabian, N. (2009b). 'Malden public schools supervision and evaluation: The professional growth cycle.' Paper presented at the International Conference on Teacher Evaluation with Professional Growth Cycles, December 1-5, in Taipei, Taiwan.

Kersten, T. A., & M. S. Israel (2005). 'Teacher evaluation: Principals' insights and suggestions for improvement.' *Planning and Changing*; 36(1/2): 47-67.

Louw, H. C. (2006). 'Teacher professional evaluation standards in USA.' Symposium conducted at the 2006 International Conference on Teachers Professional Development Evaluation, July 4-5, in Kaohsiung City, Taiwan.

Lugosi, P. (2010). 'Computer assisted self and peer assessment: Applications, challenges and opportunities.' J*ournal of Hospitality, Leisure, Sport & Tourism Education,* 9(1): 85-91.

Marshall, K. (2005). 'It's time to rethink teacher supervision and evaluation.' *Phi Delta Kappan*, 86(10): 727-731.

Miyagi Prefectural Government(2001).Teacher evaluation form. Available at http://www.pref.miyagi.jp/sibun/reiki_int/reiki_honbun/aa70012721.html [11 May 2011]

Mayo, W. R. (1997). 'Trends in teacher evaluation.' *The Clearing House,* 70 (5): 269.

Milanowski, A.T., & H. G. Heneman III. (2001). 'Assessment of teacher reactions to a standards-based teacher evaluation system: a pilot study.' *Journal of Personnel Evaluation in Education,* 15(3): 193-212.

Ministry of Education (2006a). *A brochure of evaluation on elementary and junior high school teachers' performance for promoting their professional development*. Taipei: Ministry of Education. [In Chinese].

Ministry of Education (2006b). 'A new development of teacher professional development evaluation in Taiwan.' Symposium conducted at the 2006 International Conference on Teachers Professional Development Evaluation, October 8-9, in Kaohsiung City, Taiwan. [In Chinese].

Ministry of Education (2007). *National statistics of vocational school students.* Taipei: Ministry of Education. [In Chinese].

Ministry of Education (2008). *2008 Consultation conference brochure on experimental evaluation for promoting teacher professional development*. Taipei: Ministry of Education.

Ministry of Education (2011). *Teacher evaluation for professional development*. Available at http://tepd.moe.gov.tw/chinese/04_situation/00_overview.php [11 December 2011] [In Chinese]

Ovando, M. N. (2001). 'Teachers' perceptions of a learner-centered teacher evaluation system.' *Journal of Personnel Evaluation in Education* 15(3): 213-231.

Pan, H. L., L. U. Wang, S. J. Chang,& C. W. Lin (2007).*A meta-evaluation of Teacher Evaluation for Professional Development in Taiwan*. Taipei: Ministry of Education. [In Chinese].

Painter, S. R.(2000). 'Principals' efficacy beliefs about teacher evaluation.' *Journal of Educational Administration* 38(4): 368-378.

Schmocker, M. (2001).*The results fieldbook: Practical strategies from dramatically improved schools*. Alexandria, VA: ASCD.

Steiner, L. (2010). Using competency-based evaluation to drive teacher excellence: Lessons from Singapore. NC, Chapel Hill: Public Impact.
Available at http://publicimpact.com/opportunityculture.org/images/stories/singapore_lessons_2010.pdf [20 September 2011]

TDA (2007). *Professional Standards for Teachers*. Available at http://www.tda.gov.uk/standards [11 March 2012]

Thomas, S. (2008).'Leading for quality: questions about quality and leadership in Australia.' *Journal of Education Policy,* 23(3): 323-334.

Tseng, H. C. (2011).'The review and future of teacher evaluation for teacher development in Taiwan.' Paper presented at the 2011 International symposium on Teacher Evaluation for Professional development, July 7-8, in Taitung, Taiwan. [In Chinese].

Wang, J. H. (2009).'A reflection of promotion of teacher evaluation for professional development.' Paper presented at International Conference on Teacher Evaluation with Professional Growth Cycle, December 3-4, in Kaohsiung, Taiwan. [In Chinese].

Yu, L. Y. (2011). 'A study of Teacher education policy in Japan.' Paper presented at the Conference on International Comparison of Teacher Quality, November 26-27, in Kaohsiung, Taiwan.