



# Effective restaurant rating scale development and a mystery shopper evaluation approach



Chih-Hsing Sam Liu<sup>a,1</sup>, Ching-Shu Su<sup>b</sup>, Bernard Gan<sup>c</sup>, Sheng-Fang Chou<sup>b,\*</sup>

<sup>a</sup> Department of Leisure and Recreation Administration, School of Tourism, Ming Chuan University, Taoyuan, Taiwan

<sup>b</sup> Department of Hospitality Management, School of Tourism, Ming Chuan University, Taoyuan, Taiwan

<sup>c</sup> School of Management, UNSW Business School, University of New South Wales, West Lobby Level 5, Australian School of Business Building, Sydney, NSW 2052, Australia

## ARTICLE INFO

### Keywords:

Restaurant  
Restaurant rating  
Mystery shopper  
Service quality  
Scale development

## ABSTRACT

Dining is a major attraction for tourists visiting Taiwan. In 2008, the Taipei City Government commissioned a project to produce a list of recommended restaurants in Taipei for tourists. To facilitate the selection process, we developed a restaurant rating scale using a mystery shopper approach. This study entailed 20 questions that covered all elements of the consumers' dining experience in a table-service restaurant setting, as well as a comprehensive mystery shopper training program to ensure consistency of quality in the evaluation process. This research involved the top 500 restaurants in Taipei. The findings revealed that our rating scale achieves internal consistency, validity and model fit. This study provided an important tool for further industrial applications and research opportunities. Further, this study proposes future research directions.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

A special and unique memorable gastronomic experience is an indispensable asset to any successful tourist destination (Fox, 2007). Gastronomy tourism development initiatives utilize locally produced foodstuffs and beverages to strengthen regional tourism products, enhance visitors' experiences, and help maintain and enhance the viability of local food production and processing sectors (Boyne et al., 2003). According to the above-cited studies, delicacies are a key reason that tourists visit different countries. For example in the West, slow food tourism is widely promoted in Italy (Nilsson et al., 2011), and wine and food are an integral part of the French culture. In Asia, tea tourism in China is popular (Cheng et al., 2010). Singapore invited hotels and restaurants to join the World Gourmet Summit, as well as promoted itself as an important Asian food center through organizing many fusion cuisine festivals (Chaney and Ryan, 2012).

Between 2008 and 2009, the inbound tourists of Taiwan grew by 14%, thereby achieving the highest tourism growth rate in Asia (UNWTO, 2010). According to a survey by the Taiwan Tourism Bureau, "scenery" and "dining" are the primary factors that attract tourists. Moreover, "dining" is a primary factor for the tourists coming from Japan, Korea, Hong Kong and Macau (Taiwan Tourism Bureau, 2014). Because most tourists include Taipei as an integral part of their itinerary to Taiwan, the tourists' dining experience in Taipei is crucial to the tourism industry and its development. Thus, recommending good restaurants to tourists has become a priority for the government. Given its rich historical and geographical environment, food culture in Taipei is vibrant and diversified. In 2008, the Taipei City Government sponsored a project to select and compile a list of recommended restaurants to enhance the quality of food and service for the restaurants in Taipei; however, there is no reliable measurement tool to evaluate restaurants nor is there a standardized restaurant rating system available in Asia.

In the foodservice industry, the Michelin Red Guide is the one of the most authoritative and popular restaurant guides. Despite its extensive influence in the foodservice industry globally, the Michelin Red Guide has never disclosed the written criteria for the quality level required for the various star levels (Johnson et al., 2005), which represents a significant research gap in the establishment of reliable and consistent criteria or scales for restaurant rating. Furthermore, a reliable measurement tool is just one important component of restaurant evaluation. The implementation process

\* Corresponding author at: 5 De Ming Road, Gui Shan District, Taoyuan County 333, Taiwan. Tel.: +886 3 3507001; mobile: +886 911 787982.

E-mail addresses: [phd20110909@gmail.com](mailto:phd20110909@gmail.com) (C.-H.S. Liu), [sallysu@mail.mcu.edu.tw](mailto:sallysu@mail.mcu.edu.tw) (C.-S. Su), [kcgan@unsw.edu.au](mailto:kcgan@unsw.edu.au) (B. Gan), [dodo.chou@gmail.com](mailto:dodo.chou@gmail.com) (S.-F. Chou).

<sup>1</sup> Tel.: +886 919880096.

is equally important; in particular, we need an effective and consistent evaluation process. The Michelin Red Guide deploys mystery shoppers as “inspectors” to evaluate restaurants. The mystery shopper, also known as the secret, ghost or anonymous shopper, is a market survey technique that is applied in various industries (Cobb, 1995; Dwek, 1996). Through over a century of development and refinements, the mystery shopper technique used to compile the Michelin Red Guide now has a significant impact on the foodservice industry, particularly on restaurants. Thus, the mystery shopper approach was used to evaluate restaurants after developing the scale.

This study makes several contributions. First, although the Michelin Red Guide and restaurant reputation have received practitioners' and scholars' attention in the literature, inconsistent findings and the lack of a theoretical foundation suggest a need to develop a deeper understanding of how the Michelin Red Guide evaluation process relates to restaurant performance. Thus, our study extends beyond previous studies (e.g., Marini et al., 2003; Snyder and Cotter, 1998) by analyzing how the Michelin Red Guide affects restaurant performance and by developing an effective evaluation tool for restaurant ratings in Taiwan. Further, almost every model developed to explain restaurant performance (e.g., Cheng et al., 2012; Wu and Liang, 2009) is grounded in the customer's perspective. Unfortunately, the experts' comment links proposed by prior researchers have received limited empirical attention, especially in restaurant evaluation. Our study attempts to fill the gaps of prior studies by further integrating the mystery shoppers' observations to analyze the critical indicators that form a restaurant rating measurement tool based on the Michelin Red Guide.

The structure of this study is as follows: first, we develop and then analyze the indicators that form the measurement tool for restaurant ratings. The selection of mystery shoppers and their training are critical for ensuring reliability and consistency in the evaluations. After a 3-day training program, we deploy 48 mystery shoppers to undertake empirical research using the newly developed measurement tool. Finally, we test the empirical data to validate our rating scale. Thus, by developing and then validating a standardized restaurant rating scale using the mystery shopper approach, this study has significant managerial implications for restaurant management, and it provides further scope for research in the foodservice industry and beyond. In summary, the Taipei City Government's project presented an invaluable opportunity for this study to develop a new measurement tool for rating restaurants using the mystery shopper approach and to validate this tool using the rich data collected from 500 listed restaurants.

## 2. Literature review

### 2.1. World-renowned authoritative restaurant guides

Presently, the two most representative restaurant rating systems are the French Michelin Red Guide and the American Zagat Restaurant Survey, originating in New York. In this study, we draw reference primarily from the former. Given the growing prominence of Asia in the global economy, in recent years, the *Michelin Red Guide* has issued a 2007 Tokyo Michelin Guide and a 2008 Hong Kong and Macau Michelin Guide. The introduction of these guides indicates that Asian cuisines are now more broadly accepted and popular with international tourists (Lane, 2010).

The founder of Michelin Tire Company, French industrialist André Jules Michelin, published the first Michelin Red Guide in 1900. The guide has since become the most influential restaurant guide in France and beyond. To date, total publications have reached 1.5 million copies and covered 12 countries in Europe, four cities in the United States (US), and cosmopolitan Asian cities including Tokyo, Hong Kong and Macau. The Michelin Red Guide

evaluates restaurants by anonymously assigning inspectors to dine in restaurants (Callan, 1990; Lee, 2013). In analyzing how the Michelin Red Guide affected restaurants, Johnson et al. (2005) proposed that its influence on restaurants included changing price settings, affecting the level of revenue, and elevating chefs' prestige and social position. Thus, these comprehensive considerations could provide customers with a reliable reference tool when selecting a restaurant.

#### 2.1.1. The Zagat Restaurant Survey

Americans Tim and Nina Zagat co-founded the first Zagat Restaurant Survey in 1979 by compiling their friends' surveys of New York City restaurants. Zagat's has since become one of the most trusted restaurant surveys and guides for tourists in the US and globally. Unlike the Michelin Red Guide, which uses anonymous inspectors to conduct evaluations, the Zagat survey advocates shared opinions. It relies on consumers to evaluate restaurants and then collates the results to provide the most objective and accurate information (Zagat Survey, 2007).

#### 2.1.2. The Thai select certification system

Thai Select is a certification system promoted globally by the Thai government. A certified Thai cuisine restaurant will be issued a Thai Select label by the Thai government. This label has become the Thai government's official support for high-quality restaurants that serve Thai cuisines. The main purpose of this label is to recognize and appreciate exceptional Thai restaurants as well as to encourage Thai restaurants to achieve international standards. Presently, Thai restaurants are commonly found everywhere around the world. The Thai Select system aims to provide consumers beyond the shores of Thailand with an authentic experience of Thai cuisines. Given the proliferation of Thai Select around the world, Taiwan adopted this certification system in 2007. As of 2008, there were already 21 Thai restaurants in Taiwan that had been issued the Thai Select label (<http://thaiselect.com.au>).

#### 2.1.3. Taiwan restaurant rating reviews

Presently, restaurant ratings in Taiwan are largely based on media reports. There is an obvious gap in exploring the classification, evaluation standards and credibility in Taiwan restaurant ratings. The existing Taiwanese restaurant rating system includes a ranking of restaurants in *Vision Magazine*. The magazine employs SGS (Societe Generale de Surveillance) to undertake evaluations using the mystery shopper approach. That said, these ratings only include service quality evaluations and not all aspects of restaurant evaluations. In recent years, the Taipei City Government tried to mitigate the shortfalls in the Taiwan restaurant evaluation system by embarking on mystery shopper training and establishing a more robust method of restaurant evaluations. In addition, the Taipei Restaurant Evaluations implemented by Jiao Tong largely followed the Michelin restaurant guide and involved gourmet experts evaluating various recommended restaurants over a period of months. To ensure the objectivity of the evaluations, the experts dined at their own expense and did not reveal their purpose of their visit. Each expert provided a score for each item including food, hygiene, overall environment, service, wine and beverages. Finally, they decided how many stars to award each restaurant by consensus.

### 2.2. The indicators from various restaurant rating systems

The Michelin Red Guide employs a symbol system to select the best hotels and restaurants in each classification of comfort and price. For the restaurant category, the guide proposes the following primary indicators: quality of ingredients and produce; cooking skill and talent; meal characteristics, including taste, clarity and fusion; well-balanced menu; value for the money; and consistent

**Table 1**  
Comparison of various restaurant rating systems.

|                      | Michelin Red Guide                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Zagat Restaurant Survey                                                                                                                                                                                                                                                                           | Thai Select                                                                                                                                                                                                               |
|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Origins              | France, Paris                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | United States, New York                                                                                                                                                                                                                                                                           | Thailand, Bangkok                                                                                                                                                                                                         |
| Area coverage        | 12 European Countries, four United States Cities (New York, San Francisco, Los Angeles, Las Vegas), Asian Region (Tokyo, Hong Kong and Macau)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Various cities in the United States, 27 European cities, Asia including four Japanese cities, and three Chinese cities – Beijing, Shanghai and Hong Kong                                                                                                                                          | Europe, United Kingdom, United State, Japan, South Korea and Taiwan                                                                                                                                                       |
| Evaluators           | Mystery Shoppers                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Consumers                                                                                                                                                                                                                                                                                         | Thailand representatives based overseas                                                                                                                                                                                   |
| Evaluation methods   | Mystery inspectors undertaking multiple anonymous evaluations                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Consumers evaluating their actual experience                                                                                                                                                                                                                                                      | <ul style="list-style-type: none"> <li>• Mystery investigations</li> <li>• Unannounced visit</li> </ul>                                                                                                                   |
| Evaluation criterion | <ul style="list-style-type: none"> <li>• Chef culinary skills</li> <li>• Quality of restaurant</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | <ul style="list-style-type: none"> <li>• Food</li> <li>• Decor</li> <li>• Service</li> </ul>                                                                                                                                                                                                      | <ul style="list-style-type: none"> <li>• Quality</li> <li>• Ingredients</li> <li>• Taste</li> <li>• Presentation</li> <li>• Cultural authenticity</li> <li>• Pass or fail</li> <li>• Issue “Thai Select” label</li> </ul> |
| Evaluation outcomes  | <ul style="list-style-type: none"> <li>• Chef culinary skills (represented by the symbol ★, three ★ being the best): one ★ denotes “Good Restaurant worth visiting”, Two ★ denotes “Excellent Culinary Skills, Worthy of a devour”, Three ★ denotes “Perfect Culinary Skills, Worth a special trip and premium spending for the gastronomic experience”</li> <li>• Quality of restaurant (represented by the symbol of Fork and Spoon , five  being the best): One symbol denotes “Comfortable Restaurant”, Two symbols denote “Cosy Restaurant”, Three symbols denote “Very Comfortable Restaurant”, Four symbols denote “High Class Restaurant”, and five symbol denotes “Luxurious Restaurant”</li> </ul> | <ul style="list-style-type: none"> <li>• Three criterion – food, decor and service – where the highest score for each criterion is 30 points: 26–30 points denotes “Exceptional to Perfect”, 20–25 points denotes “Very good to Outstanding”, 16–19 points denotes “Good to Very Good”</li> </ul> |                                                                                                                                        |
| Evaluation frequency | Michelin-star restaurants are evaluated several times every year. Non-Michelin star Restaurants with Forks and Spoons symbols are evaluated every 1.5–2 years                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | No set evaluation frequency. Consumers evaluate each restaurant on their own time                                                                                                                                                                                                                 | Thai Select Restaurants are evaluated every three years                                                                                                                                                                   |

Source: Compiled by the Authors.

experience in each visit (Henderson, 2011). The number of stars, ranging from one to three stars, in the Michelin Red Guide represents the level of culinary skill. The guide focuses on the cuisine rather than interior decoration, service quality, and the setting of the table. Further, in rating a restaurant's level of comfort in terms of environment and quality of service, the Michelin Red Guide uses the symbol of a crossed spoon and fork () , for which a minimum of one symbol denotes “quite comfortable” and a maximum of five denotes “luxury in a traditional style”.

The Zagat Restaurant Survey evaluates restaurants based on three criteria – food, decoration and service – and the highest score for each criterion is 30 points. The Thai Select system also employs a stringent set of certification criteria. Apart from reviews by Thailand's representatives based overseas, the certification criteria include food safety, hygiene conditions, sources of raw materials, chef's quality, restaurant decor, service quality, dining atmosphere, menu variety, and integration with Thai culture. The entire certification process takes approximately six months. Once approved, the Thai Select label is valid for three years. Upon expiry, restaurants must undergo the entire certification process again. Based on established accreditation standards, Thai select restaurants are classified into traditional Thai cuisine, modern Thai flavor, Specialty Thai cuisines and franchise Thai restaurants. Comparisons of the three established restaurant rating systems are listed in Table 1.

In addition, much research focuses on the relationships between service quality and restaurant ratings because service quality has long been recognized to be an important external source of restaurant selection (Zhang et al., 2010). Litvin et al. (2008) suggested that tourists' restaurant selections depend on service quality and recommendations by friends. According to Han and Kim (2009), restaurant service quality is highly connected to perceived value and dining experience. Perceived value refers to a comprehensive evaluation of food, service, environment, price, etc., and is

the greatest contributor to the customer's behavioral intentions and emotional responses (Liu and Jang, 2009). Empirically, Han and Kim (2009) investigated perceived price and perceived value as the key factors behind customers' restaurant selection. Thus, perceived value not only influences the market ratings of restaurants, but also affects customer's restaurant selection. Further, dining experience, physical facilities, ambience and service are critical indicators that influence service quality in restaurants (Johns and Pine, 2002). Previous studies show that dining experience, including food quality, meal variety, service delivery, and meal presentation, were all associated with positive customer emotions (Nield et al., 2000). For instance, after a thorough literature review, Knutson (2000) suggested that dining experience was associated with positive behavioral intentions, such as cleanliness, friendliness, price, prompt service, menu consistency and variety, location, atmosphere and promotion. Based on the aforementioned discussions, service quality significantly affects restaurant market reputation and ratings. Therefore, reputation has been verified to be an appropriate explanation for the effect of restaurant ratings on consumer dining experience and value perceived in different consumption settings.

Drawing from the literature noted above as well as from the Michelin Red Guide and the Zagat Restaurant Survey, the five main indicators of exceptional restaurants summarized include (1) meals, (2) service, (3) dining environment and atmosphere, (4) price and (5) reputation/popularity. Thus, these five indicators serve as the foundation on which to develop the Mystery Shopper measurement/evaluation tool for rating restaurants.

### 2.3. Mystery shopper

The mystery shopper is a technique applying the anonymous observation of a service delivery procedure from a customer

perspective (Wilson, 1998a, 1998b). The primary issues highlighted in the recent research on this topic include the training of mystery shoppers (Beck et al., 2004), the relationship with service quality (Beck and Miao, 2003), the performance of service (Wilson, 2001), and the measurement of service delivery (Wilson, 1998a). Wilson (1998b) proposed that mystery shoppers could contribute to the improvement of service performance in the short run. It was also proposed that mystery shopping could be a useful instrument in addition to the more common survey methods to measure customer satisfaction (Hessellink and Wiele, 2003; Wiele et al., 2005). Furthermore, mystery shoppers are not representative of the restaurants' typical customer. However, well-trained mystery shoppers provide a cost-effective evaluation method and different feedback than customers, and mystery shoppers have been widely used as a diagnostic tool for restaurants service quality evaluations (Lynn, 2002; Mathe and Slevitch, 2013; Oronsky and Chathoth, 2007).

The mystery shopper process involves a group of people who are selected and trained by an enterprise or organization that provides a mystery shopper service. These mystery shoppers disguise themselves as customers to gather information (Cobb, 1995; Dwek, 1996). After visiting the restaurant, the mystery shopper completes a series of detailed reports to provide the principal with relevant information, such as front-line staff performance, product quality, service delivery procedures and other potential problems involving the evaluated organization (Finn and Kayande, 1999; Finn, 2001). Moreover, mystery shopper methods were applied to explore the influence of emotion displayed by contact personnel on consumption expenditure and re-patronage intentions in bakery chains (Wang et al., 2012). In addition, the mystery shopper method has also been applied in other service industries to evaluate service quality. For example, Tarantola et al. (2012) used the mystery shopper method to collect information about loan products and services in Greek Banks. In this study, they also used multivariate statistics such as Bayesian networks to analyze the data. Furthermore, mystery shoppers provided useful information about the service quality at various levels. The mystery shoppers evaluated the sales effectiveness of bank staff on several factors including whether the staff listened to their customers, if the tellers were friendly, and how long it took to serve a customer, as well as other aspects of the banking services.

In the restaurant industry, mystery shopper's observations provides good index for measuring restaurants' service quality (Lynn, 2002). The advantages of mystery shopper are that the anonymous inspector which analyzes the needs of all customers and could collect a large amount of information of customer reaction for restaurants (Oronsky and Chathoth, 2007). For another perspective, compare to customer surveys, mystery shopping also provides a cost-effective method and different feedback for restaurants (Mathe and Slevitch, 2013). This is due to before the evaluation process, the mystery shopper need to experience the necessary professional training and therefore will provide the feedback necessary to appraise restaurants service (Urban, 2013). Therefore as suggested by the literature above, this paper used mystery shopper's observations when conducting restaurants rating.

### 3. Methodology

#### 3.1. Measurement tool

##### 3.1.1. Design of the scale

This study combines qualitative and quantitative methods of research design. The research team collected famous gourmet

restaurants guides around the world, namely, Zagat Survey, Michelin Red Guide, Bottin Gourmand, Champerard, Gault & Millau, Thai Select, and other restaurant guides published by governments and organizations. Further, we also analyzed the related literature for samples of top quality restaurant service collated using gourmet restaurant mystery shopper surveys. In developing in-depth interview outlines based on the literature reviews, we found that the gourmet restaurant mystery shopper evaluation criteria can be summarized in the following four dimensions: (1) cuisines: food style, taste, quality of ingredients, decoration of dish, exquisiteness of cuisine, complexity and subtlety, overall coordination; (2) cellar and wine: wine evaluation; (3) atmosphere: table setting, design and decoration; and (4) other variables: comfort and price. Next, the research team constructed a restaurant evaluation scale through in-depth interviews with 14 experts and the results of the literature review. The experts were divided into three categories. Group A included experts from the catering and gourmet tourism marketing field; Group B included experts from the restaurant services quality evaluation field; and Group C included experts with extensive experience in mystery shopper evaluation. A total of 14 experts were invited. On average, each had 24 years of working experience in their respective field (as detailed in Appendix 1).

Through the use of qualitative research methods, in-depth interviews, and content analysis, and by reviewing, conceptualizing, coding, classifying, and narrowing our interpretation of the connotation of excellent restaurant service quality, the concept for the mystery shopper restaurant rating was formed. In addition, through further discussion, the important factors and code classifications of the mystery shopper restaurant rating indicators were established and subsequently developed into scale items.

#### 3.2. Procedure of the scale development

##### 3.2.1. Initial questionnaire

The restaurant rating scale was derived from expert interviews and focus group meetings. The project team further discussed and fine-tuned the content of the measurement tool. Each item consists of a 5-point Likert scale, where the range is "1" to "5" ("very poor" to "very good").

The restaurant rating scale development for this study included multiple stages of qualitative and quantitative surveys. In the first stage, this study analyzed relevant documents including the Michelin Red Guide, Gault Millau, the Zagat Survey, Thai Select and reviews. In this stage, we aimed to identify the primary dimensions and measurements of a restaurant rating scale. After completing the original questionnaire, we invited 10 experts (three academic experts and educators, four gourmets and reporters, and three delegates from the Taipei City Government) in related restaurant rating scales to review the questionnaire and eliminate unclear questions. Based on the experts' opinions and experiences, we revised the questions.

##### 3.2.2. Expert reviews, reliability and validity analysis and questionnaire revision

In the second stage, we confirmed the primary dimensions of the restaurant rating scale. The project team invited 10 experts from the foodservice industry and from restaurant rating to review each scale item and to evaluate the theoretical framework compliance as well as the meaning of the questions and the appropriateness of the text. Following this review and verification process, we conducted questionnaire surveys of 20 restaurants to examine the reliability and validity of the questionnaire tool. Finally, based on the results

**Table 2**  
Mystery shopper training program.

| Day (time)                                           | Course schedule                                                             |
|------------------------------------------------------|-----------------------------------------------------------------------------|
| Day 1 (8 h)                                          | Food and service knowledge and techniques                                   |
| Basics in evaluating food and service in restaurants | Restaurant quality management (food quality and service quality evaluation) |
|                                                      | The art and aestheticism of restaurant                                      |
|                                                      | The basics of evaluation                                                    |
|                                                      | Explanation of evaluation criteria, and sharing of evaluation experience    |
| Day 2 (8 h)                                          | Introduction to Western Cuisine                                             |
| Western Cuisine                                      | Assessment practices in Western Cuisine                                     |
| Evaluation Program                                   | Overall onsite dining experience                                            |
| Day 3 (8 h)                                          | Introduction to Chinese Cuisine                                             |
| Chinese Cuisine                                      | Assessment practices in Chinese Cuisine                                     |
| Evaluation Program                                   | Overall onsite dining experience                                            |

of the meeting with the 10 experts, we created the four dimensions and the 20-item questionnaire.

### 3.3. The research objects

#### 3.3.1. The pilot study

In the pilot study, we tested our restaurant rating tool in 20 restaurants recommended by gourmets and academic experts. These 20 restaurants included Chinese, Western, Taiwanese, Japanese, and various types of foreign cuisines. In this stage, we further corrected any ambiguities in the questionnaire that could be misunderstood by the mystery shoppers.

#### 3.3.2. The mystery shopper profiles and selection process

Through open enrolment, the Taipei City Government recruited 68 applicants to undertake mystery shopper training before participating in the restaurant rating exercise. A total of 48 out of the 68 applicants completed the three-day training course (refer to Table 2 for the course outline).

The 48 applicants consist of 35 females (73%) and 13 males (27%). In terms of educational qualifications, 24 applicants have bachelor degrees (50%), and 20 applicants have master or doctoral degrees (41%). The applicants are between 25 and 55 years old. They come from a variety of occupational backgrounds such as senior executive, journalist, university lecturer, office worker, hotel manager, bartender and consultant. Every applicant is passionate about food, and patronizes restaurants ranging from once to seven times per week. A total of 10% of the applicants have a strong preference for salty and spicy food, while the remaining 90% of the applicants do not have any strong preference. The restaurants they frequently patronize include Taiwanese, Cantonese, Shanghainese, Japanese, Thai, French, Italian, American cuisines.

#### 3.3.3. The restaurants selection process

To enhance restaurant service quality in Taipei, the Taipei City Government implemented mystery shopper restaurant rating evaluations following the restaurant shortlisting process described below:

- (1) First, the Taipei City Government invited 10,000 companies operating in the Taipei restaurant industry via mail to register online to partake in the massive evaluation exercise and also to conduct an initial self-evaluation exercise. This request resulted in 2100 companies completing the registration.
- (2) Because each evaluation would pay basic expenses (report writing and part of the main meal) to the mystery shopper, subject to limited funding, this study invited six experts (including a gourmet author, reporter, consultant and scholar; overall, the experts had an average of 15 years working experience) to set

the standard to reduce the sample size. The select standard follows:

- (a) The restaurants selected for evaluation were primarily table-service restaurants. Thus, fast food, snacks, bars, etc., were not included in the sample pool.
  - (b) For franchised restaurants, headquarters nominated a representative franchisee to participate in the evaluation.
  - (c) The short-listing process eliminated not-profitable restaurants and those with poor self-evaluation results.
  - (d) Priority selection was given to award-winning, gourmet restaurants or those providing evidence of quality through related reports.
- (3) Finally, based on these above selection criteria, the researcher selected the samples from the complete list of 2100 enrolled restaurants; 728 restaurants met the conditions (500 on the priority list and 228 on the waiting list). The researcher assigned mystery shoppers to evaluate the restaurants via random lottery selection methods.

If the mystery shoppers could not complete an evaluation, such as if the restaurant was closed or for other reasons, then researcher would assign another restaurant on the waiting list until all 500 evaluations were complete.

#### 3.3.4. The mystery shopper restaurant rating process

The mystery shopper restaurant rating process spanned a period of three months from May to August 2009. Depending on availability and patronizing frequency, each individual mystery shopper was given three to five randomly allocated restaurants (from the list of 500 pre-selected restaurants) to conduct anonymous restaurant ratings. Each mystery shopper consented and signed off on a confidentiality agreement in advance. They were also required to provide relevant supporting photography or a photocopy of restaurant bills and to file an evaluation report within six days of visiting the restaurant. Finally, the project team vetted each evaluation report for inconsistency and also terminated the employment of those mystery shoppers who failed to do a proper evaluation job. Upon completion of the evaluation process, the project team announced the final top 100 restaurants that would be widely promoted via official restaurant guides, press conferences, and various media coverage as the highlights of Taipei Tourism. The mystery shopper restaurant rating process is shown in Fig. 1.

### 3.4. Statistical analysis plan

The purpose of this study was to develop a restaurant rating scale. After gathering the data, the editing and classifying of the data began. The data collected in this study was recorded in each mystery shopper evaluation report and entered into SPSS 18.0.0 software. This study used the following SPSS 18.0.0 feature. The statistical analysis plan can be divided into two stages: (1) formal scale established and (2) scale reliability and validity test.

#### 3.4.1. Formal scale established

The purpose in this stage was to confirm the reliability and validity of each item in the mystery shopper restaurant rating scale. The researcher used item analysis and exploratory factor analysis (EFA) to achieve this goal.

- (a) *Item analysis.* This study used item analysis to delete unsuitable items. The deletion criterion included significant skewness, not reaching statistical level ( $p < .001$ ), and the removal reduced the Cronbach's  $\alpha$  coefficient.
- (b) *Exploratory factor analysis (EFA).* Principle components analysis was used because the primary purpose was to identify and compute the factors underlying the mystery shopper restaurant

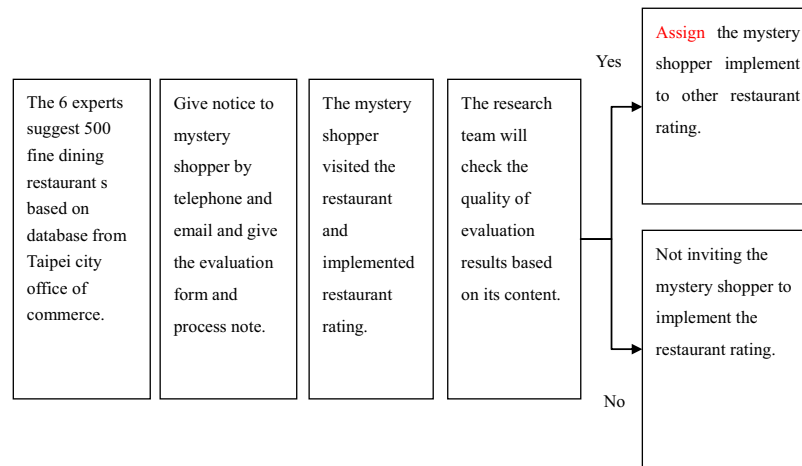


Fig. 1. Mystery shopper restaurant rating process flow.

rating scale. After calculating the factor loadings for each question item, they were summarized into different factors, such as higher loading factors and other factors to be retained, and the factor structure of the scale was confirmed.

#### 3.4.2. Scale reliability and validity test

The purpose of this stage was confirming the mystery shopper restaurant rating scale's internal consistency and construct validity (convergent validity and discriminant validity). We use Cronbach's  $\alpha$ , confirmatory factor analysis (CFA) and Analysis of Variance (ANOVA).

(a) *Cronbach's  $\alpha$* . The internal consistency and homogeneity of the test scale are good. Nunnally (2010) and Cronbach (1951) suggest that Cronbach's  $\alpha$  should be greater than 0.70.

(b) *Confirmatory factor analysis (CFA)*. This study analyzed the good fit results through the CFA to proof the scale of convergent validity. This study used Joreskog's (1969) constructed SEM basic model.

(c) *Analysis of Variance (ANOVA)*. We used ANOVA to test the discriminant validity of the scale. Past research found that customer perceptions for the quality of food, service, and the physical environment are affected by the perceived price 'indirectly affecting customers' satisfaction and recommending intentions (Liu and Jang, 2009; Ryu and Jang, 2007, 2008b). In addition, restaurant capacity is also a key factor affecting restaurant ambience (Liu and Jang, 2009; Ryu and Jang, 2008a). This study used ANOVA to test different amounts for the average check against the recommending intention and different capacities.

Table 3  
Mystery shopper restaurant rating factor analysis results ( $N=500$ ).

| No.                               | Item description                                                                                               | Factor loadings |          |          |          | Communality |
|-----------------------------------|----------------------------------------------------------------------------------------------------------------|-----------------|----------|----------|----------|-------------|
|                                   |                                                                                                                | Factor 1        | Factor 2 | Factor 3 | Factor 4 |             |
| 1                                 | Food is served at the right temperature                                                                        |                 |          | 0.66     |          | 0.47        |
| 2                                 | Excellent food presentation                                                                                    |                 |          | 0.69     |          | 0.58        |
| 3                                 | Appropriate cooking method and food is delicious                                                               |                 |          | 0.83     |          | 0.69        |
| 4                                 | Excellent taste of food leaves long lasting impression                                                         |                 |          | 0.79     |          | 0.71        |
| 5                                 | The food look and smell delicious                                                                              |                 |          | 0.82     |          | 0.70        |
| 6                                 | Preparation of food meets health trends                                                                        |                 |          | 0.65     |          | 0.48        |
| 7                                 | The dish and utensils are clean and hygienic                                                                   |                 |          | 0.54     |          | 0.43        |
| 8                                 | Efficient and effective booking service                                                                        | 0.68            |          |          |          | 0.49        |
| 9                                 | Efficient and effective process in the welcoming and ushering of the customers                                 | 0.83            |          |          |          | 0.70        |
| 10                                | Efficient and effective explanation of the menu                                                                | 0.79            |          |          |          | 0.63        |
| 11                                | Efficient and effective process in delivery of food                                                            | 0.80            |          |          |          | 0.65        |
| 12                                | Efficient and effective billing process                                                                        | 0.78            |          |          |          | 0.61        |
| 13                                | Excellent interaction with the customers in the entire process                                                 | 0.85            |          |          |          | 0.72        |
| 14                                | The creation of a superior theme or ambience using table decoration, restaurant decor, appearance, and signage |                 | 0.87     |          |          | 0.76        |
| 15                                | The design of the menu is congruent with the restaurant features and ambience                                  |                 | 0.80     |          |          | 0.66        |
| 16                                | An excellent and appropriately themed website and publicity materials                                          |                 | 0.80     |          |          | 0.64        |
| 17                                | Presence of pleasant ambient stimuli such as lighting, music and scent                                         |                 | 0.82     |          |          | 0.70        |
| 18                                | Reasonable pricing                                                                                             |                 |          |          | 0.81     | 0.72        |
| 19                                | Overall perceived as value for money                                                                           |                 |          |          | 0.82     | 0.78        |
| 20                                | Restaurant is well known, has rich history and culture                                                         |                 |          |          | 0.38     | 0.36        |
| Eigenvalues                       |                                                                                                                | 6.27            | 5.03     | 5.88     | 1.91     |             |
| Variance explained (%)            |                                                                                                                | 40.63           | 9.12     | 7.49     | 5.21     |             |
| Cumulative variance explained (%) |                                                                                                                | 40.63           | 49.74    | 57.23    | 62.45    |             |

KMO (Kaiser–Meyer–Olkin) measure of sampling adequacy = 0.925.

**Table 4**  
Mystery shopper restaurant rating reliability and AVE (N = 500).

| Factors         | Cronbach's $\alpha$ | CR   | AVE  |
|-----------------|---------------------|------|------|
| Service         | 0.88                | 0.88 | 0.52 |
| Ambience        | 0.83                | 0.91 | 0.62 |
| Meals           | 0.84                | 0.89 | 0.68 |
| Value for price | 0.82                | 0.73 | 0.50 |
| Overall scale   | 0.92                |      |      |

### 3.5. Data analysis

In the proposed measurement model, we employ confirmatory factor analysis (CFA) to analyze the data. CFA not only provides a substitute measure of internal consistency but also assesses the external consistency of the scale items (Sethi and King, 1994). We then use the AMOS software package to test the 20 measurement scale items using four factor structures.

## 4. Findings

### 4.1. Item analysis

Item analysis was utilized to eliminate the items with lower reliability. According to the result of the item analysis, all 20 items could be kept in the restaurant rating scale for mystery shoppers because the Cronbach's  $\alpha$  coefficients for the factors did not become too high if any item was deleted.

### 4.2. Exploratory factor analysis

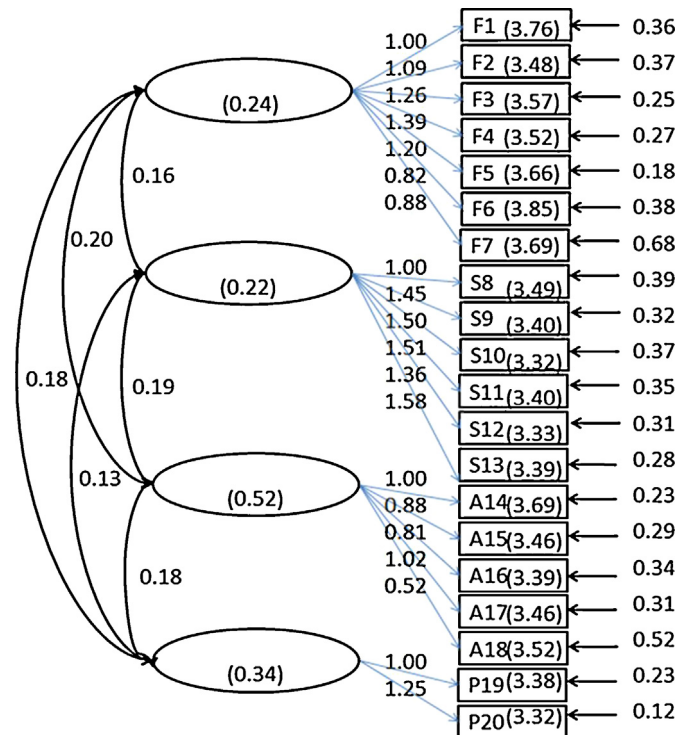
While employing factor analysis, a factor loading cut-off point of 0.40 was used to retain items, and factors with an eigenvalue equal to or greater than 1 were retained. The Kaiser–Meyer–Olkin (KMO) is 0.925 and Bartlett's test of sphericity is 5015.08 (chi-square value), reaching the significant level (refer to Table 3). A principal component factor analysis with Varimax rotation on statements regarding restaurant ratings for mystery shoppers identifies four factors. As a result, the 20 items were reduced to four factors, which explained 62.45% of the total variance. The four restaurant rating factors were Service (factor 1), Ambience (factor 2), Meals (factor 3) and Value for the price (factor 4).

### 4.3. Reliability and validity of the scale

After verifying the items in the questionnaire, we examined their reliability and validity. There were 20 question items in total, which included six items for Service, four items for Ambience, seven items for Meals, and three items for Value for the price.

In this research, we used Cronbach's  $\alpha$  to test the internal consistency of the mystery shopper restaurant rating scale. Cronbach's  $\alpha$  coefficients for the four restaurant rating factors are between 0.82 and 0.88. In terms of the entire scale, Cronbach's  $\alpha$  coefficient is 0.92 and represents good reliability (refer to Table 4).

We used confirmatory factor analysis (CFA) to test the proposed measurement model. CFA not only provides an alternative measure of internal consistency, but also assesses the external consistency of the scale items (Sethi and King, 1994). As shown in Table 2, the composite reliability (CR) of the four constructs are between 0.73 and 0.91, which is higher than 0.6 (Bagozzi and Yi, 1988). The average variance extracted (AVE) reflects the overall amount of variance in the indicators accounted for by the latent construct. All of the AVE values are higher than 0.5, which suggests that the indicators are truly representative of the construct. This result confirms the discriminate validity (Fornell and Larcker, 1981); thus, the measurement scale is valid.



**Fig. 2.** Mystery shopper restaurant rating factor structure.  $\chi^2/df = 3.13$  ( $\chi^2 = 512$ ,  $df = 164$ ;  $p = .000$ ), NFI = .90, RFI = .87, IFI = .93, CFI = .93, RMSEA = .07.

### 4.4. Overall model fit by confirmatory factor analysis (CFA)

The results obtained using CFA indicate that the proposed measurement model fits the data (refer to Fig. 2). The fit indices of the model demonstrate that  $\chi^2/df = 3.125$  ( $\chi^2 = 512$ ,  $df = 164$ ;  $p = .000$ ), NFI = 0.899, RFI = 0.871, IFI = 0.929, CFI = 0.929, RMSEA = 0.065. These indices indicate that the proposed measurement model has good fit.

### 4.5. The results of ANOVA

To understand mystery shopper restaurant rating discriminant validity, we employ ANOVA to examine the differences in mystery shopper restaurant ratings among the restaurants with different average bills and capacities. The ANOVA is the most commonly used method to examine the difference between two groups with respect to a particular variable. To use the model legitimately, the normal assumption has to be validated to distinguish whether the difference is statistically significant. Bonett and Woodward (1990) suggest undertaking the skewness and kurtosis tests to examine whether the normal assumption in large samples ( $N > 50$ ) is sufficient. The normality check results of our study are as follows: Service (skewness =  $-0.258$ , Std error = 0.109; kurtosis = 0.65, Std error = 0.218), Ambience (skewness =  $-0.548$ , Std error = 0.109; kurtosis = 0.606, Std error = 0.218), Meals (skewness =  $-0.089$ , Std error = 0.109; kurtosis =  $-0.258$ , Std error = 0.109) and Value for the price (skewness = 0.058, Std error = 0.114; kurtosis =  $-0.117$ , Std error = 0.227). Based on the above results, we found that the results of the data are close to a normal distribution.

#### 4.5.1. The restaurants with different average bills

The study used a one-way ANOVA to test mystery shopper restaurant rating scale discriminant validity. We classify the surveyed restaurants into four groups with different average checks (less than NTD\$300, NTD\$301 to NTD\$500, NTD\$501 to NTD\$1000, more than NTD\$1001) (Note: 1 US dollar is approximately 29

**Table 5**  
The result of ANOVA among different average checks.

| Variables       | SS    | df  | MS   | F        | Post hoc tests  |
|-----------------|-------|-----|------|----------|-----------------|
| Service         | 11.50 | 496 | 3.83 | 8.31***  | 1, 2 < 3, 4     |
| Ambience        | 12.03 | 496 | 4.01 | 9.65***  | 1 < 2, 3 < 3, 4 |
| Meals           | 14.11 | 482 | 4.70 | 15.23*** | 1, 2 < 2, 3 < 4 |
| Value for price | 0.36  | 496 | 0.12 | 0.32     |                 |

\*  $p < .05$ .  
 \*\*  $p < .005$ .  
 \*\*\*  $p < .001$ .  
 Note. 1 = <NT\$300, 2 = NT\$301–500, 3 = NT\$501–1000, 4 = >NT\$1000.

New Taiwan Dollars (NTD)). The results reveal that there is significant difference in Service ( $F = 8.31, p < .001$ ), Ambience ( $F = 9.65, p < .001$ ) and Meals ( $F = 15.23, p < .001$ ), but not in Value for the price ( $F = 0.32, p > 0.05$ ) among the restaurants with different average checks. Subsequently, using posteriori comparison, we find that the restaurants with higher average checks have better performance in Service, Ambience and Meals. The results provide evidence for mystery shopper restaurant rating scale discriminant validity (refer to Table 5).

Michelin restaurant research demonstrates that higher priced restaurants have more resources to maintain service and food and create an atmosphere of quality (Cotter and Snyder, 1996, 1998; Snyder and Cotter, 1998). Other study results also indicate that Michelin star restaurant chefs have higher standards, propose more innovative products, select ingredients, control costs and profits, and consider menu pricing (Ottenbacher and Harrington, 2007).

We use two-way ANOVA to confirm scale discriminant validity. We divide the samples into two groups based on the respective mystery shopper's intention to recommend or not recommend. For these two groups, we compare average checks with the four dimensions of mystery shopper restaurant ratings by applying posteriori comparison; we find that there is no significant difference between average checks and service, ambience, meals, or value for the price. In a two way ANOVA main effects for the Service variable, the results reveal that there is a significant difference for different average checks ( $F = 2.61, p < .05$ ) and recommending intention ( $F = 84.16, p < .005$ ) but not for average check times recommending intention. For the Ambience variable, the results reveal that there is significant difference for the different average checks ( $F = 5.69, p < .005$ ) and recommending intention ( $F = 40.62, p < .005$ ) but not for average check times recommending intention. For the Meals variable, the results reveal that there is a significant difference for different average checks ( $F = 8.15, p < .005$ ) and recommending intention ( $F = 142.31, p < .005$ ) but not for average check times recommending intention. For the Value for the price variable, the results reveal that there is significant difference only for recommending intention ( $F = 142.31, p < .005$ ) but not for different average check or for average check times recommending intention.

In Table 6, by applying posteriori comparison, we find that the Service variables average check at the <NTD\$300 and <NTD\$301–500 levels are significantly different from the <NTD\$501–1000 and <NTD\$ > 1000 levels. We found that the average NTD\$ 500 was a peak, with an average check of >NTD\$ 500 indicating restaurant service performance significantly higher than check <NTD\$ 500. In addition, for the ambience variable, the average check at the <NTD\$300 level was significantly different from the average check at the <NTD\$301–500, <NTD\$501–1000 and NTD\$>1000 levels. For the meals variable, the average checks at the <NTD\$300 and <NTD\$301–500 levels were significantly different from the average check at the <NTD\$501–1000 and NTD\$>1000 levels. Finally, for the Value for the price variables, “not recommended” was significantly different from “recommended”. The cross tabulation result reveals the effect of the average check level on the intention to recommend the restaurant.

**Table 6**  
The result of cross tabulation between average check and recommending intention.

| Variables              | SS    | df | MS    | F        | Post hoc tests |
|------------------------|-------|----|-------|----------|----------------|
| <i>Service</i>         |       |    |       |          |                |
| AVC                    | 3.05  | 3  | 1.02  | 2.60*    | 1, 2 < 3, 4    |
| R.I.                   | 32.83 | 1  | 32.83 | 84.16**  | 0 < 1          |
| AVC × R.I.             | 1.16  | 3  | 0.39  | 0.99     |                |
| <i>Ambience</i>        |       |    |       |          |                |
| AVC                    | 6.34  | 3  | 2.11  | 5.69**   | 1 < 2, 3 < 4   |
| R.I.                   | 15.08 | 1  | 15.08 | 40.62**  | 0 < 1          |
| AVC × R.I.             | 0.11  | 3  | 0.04  | 0.10     |                |
| <i>Meals</i>           |       |    |       |          |                |
| AVC                    | 5.49  | 3  | 1.83  | 8.15**   | 1, 2 < 3 < 4   |
| R.I.                   | 31.95 | 1  | 31.95 | 142.31** | 0 < 1          |
| AVC × R.I.             | 0.04  | 3  | 0.01  | 0.06     |                |
| <i>Value for price</i> |       |    |       |          |                |
| AVC                    | 1.04  | 3  | .35   | 1.04     |                |
| R.I.                   | 16.42 | 1  | 16.42 | 49.28**  | 0 < 1          |
| AVC × R.I.             | 0.25  | 3  | 0.08  | 0.25     |                |

AVC: average check; R.I.: recommending intention.  
 Whether to recommend: 0, not recommended; 1, recommended.  
 \*  $p < .05$ .  
 \*\*  $p < .005$ .  
 \*\*\*  $p < .001$ .  
 Note. 1 = <NT\$300, 2 = NT\$301–500, 3 = NT\$501–1000, 4 = >NT\$1000.

**Table 7**  
The result of ANOVA among different capacities.

|                 | SS    | df  | MS   | F      | Post hoc tests |
|-----------------|-------|-----|------|--------|----------------|
| Service         | 1.93  | 495 | 0.48 | 1.00   |                |
| Ambience        | 13.92 | 495 | 3.48 | 8.43** | 1 < 2, 3, 4, 5 |
| Meals           | 0.78  | 481 | 0.19 | 0.58   |                |
| Value for price | 3.51  | 495 | 0.88 | 2.36   |                |

\*  $p < .05$ .  
 \*\*  $p < .005$ .  
 \*\*\*  $p < .001$ .  
 Note. 1 = <50 seats, 2 = 51–100 seats, 3 = 101–150 seats, 4 = 151–200 seats, 5 = >201 seats.

According the result, we find that there is no significant difference between the average check and Service, Ambience, Meals, and Value for the price. For the four dimensions of brand equity, the recommended restaurants are obviously higher than those not recommended. In addition, we also find that restaurants with average bills of over NTD\$1000 and not recommended are rated lower scores on Service (refer to Appendix 2).

#### 4.5.2. The restaurants with different capacities

Restaurant capacity is also a key factor affecting customer perception and restaurant ambience performance (Liu and Jang, 2009; Ryu and Jang, 2008a). We classify the surveyed restaurants into five groups with different capacities (0–50, 51–100, 101–150, 151–200 and more than 201 seats). The results reveal that there is only one significant difference on Ambience ( $F = 8.43, p < .001$ ) among the restaurants with different capacities. Subsequently, by using posteriori comparison, we find that the restaurants with larger capacity of over 50 seats have better performance on Ambience (refer to Table 7). The above ANOVA results enhance the discriminant validity of the mystery shopper restaurant rating scale.

## 5. Conclusions

### 5.1. Implications and contributions

Food is an important tourism resource. As part of a tourism strategy, we can make great restaurants into a tourism topic through the restaurants' evaluation and promotion. However, restaurants

are numerous and widespread. It is very difficult to evaluate their value and quality. It is therefore very important to understand how to identify a restaurant with great service quality. This study makes a number of significant contributions. First, this is a pioneering study in Asia that develops a standardized restaurant rating system using the mystery shopper approach. Second, by establishing the restaurant rating scale, this study provides the Taipei Tourism Bureau with an effective tool for screening and recommending different types of restaurants to consumers. The index for evaluating restaurants could be a reference for investigating restaurant quality in other countries. Compared with the general public, well-trained mystery shoppers can perform better and more professional evaluations (Beck et al., 2004; Finn, 2001). The study combined interviews and 500 restaurant surveys made by 48 mystery shoppers and passed tests for reliability and validity. The evaluation can be used as a tool to promote tourism in other countries.

Third, this evaluation tool encompasses all aspects important to consumers such as the quality of food, service, ambience and price. This tool requires the deployment of trained mystery shoppers, which in turn provides a more objective result. Fourth, this study provides the Taipei Tourism Bureau with important insights into the local foodservice industry as well as a benchmarking resource to help poorly performing restaurant operators improve their businesses. The possible areas for improvement may include the development of new products, improvement of service procedures, design of space, and adjustment of pricing to meet consumers' expectations. Finally, the validation of this evaluation tool will help the development of restaurant reviews and the implementation of future mystery shopper restaurant ratings. This tool has a positive impact on improving the overall quality of the foodservice industry.

## 5.2. *For future research and Limitations*

From the perspective of future research, each item in this evaluation tool is simple and clear, which will aid in its wide usage. This study suggests that relevant tourism bureaus provide more in-depth training for mystery shoppers. In-depth training can include food tasting techniques as well as differentiating the service and dining ambience of different restaurants (Butcher et al., 2009; Luria and Yagil, 2008). Further, the restaurant 'manufacturing' process consumes much energy, which in turn causes pollution and greenhouse gas emissions. Thus, due to the global warming phenomenon, we recommend that future evaluations include 'impact on the environment' considerations. In addition, one primary attraction in the Taiwanese food and beverage industry is the snacks served by street vendors in the various night markets. We suggest that the relevant tourism bureaus adapt this evaluation tool to be applied beyond service restaurants to evaluate other types of dining and night markets. We further suggest that the relevant bureaus release the evaluation results with the aim of improving the overall quality of the local food and beverage industry.

For the long-term sustainability of the local food and beverage industry, we need to consider a full range of important factors such as the quality of food, service, ambience and price. Apart from utilizing this evaluation tool in restaurant management, it has wide applications as a tool for self-evaluation, designing restaurant management training courses, evaluating the catering industry, developing teaching materials to strengthen the management capabilities of restaurant operators and creating more new products that meet consumer expectations. In addition, the food and beverage industry can utilize certain parts of the evaluation tool to target specific problems or gaps in a restaurant's performance. Finally, there remains much room for further research in the area of mystery shopper evaluations.

As a multi-faceted and comprehensive evaluation tool, this tool is consistent with previous similar studies (Knutson, 2000; Tam and Yung, 2003; Johns and Pine, 2002). To further test the reliability and validity of this evaluation tool using the mystery shopper approach, we recommend that future studies include various types of restaurant evaluations in China, other parts of Taiwan, Europe, the United States, Japan and Southeast Asia. Similarly, studies could be applied to Chinese restaurants in different countries such as Mainland China, Singapore and Thailand. Finally, we hope that this study encourages more ongoing investigations in the testing of this mystery shopper evaluation tool to further develop its potential through more trials.

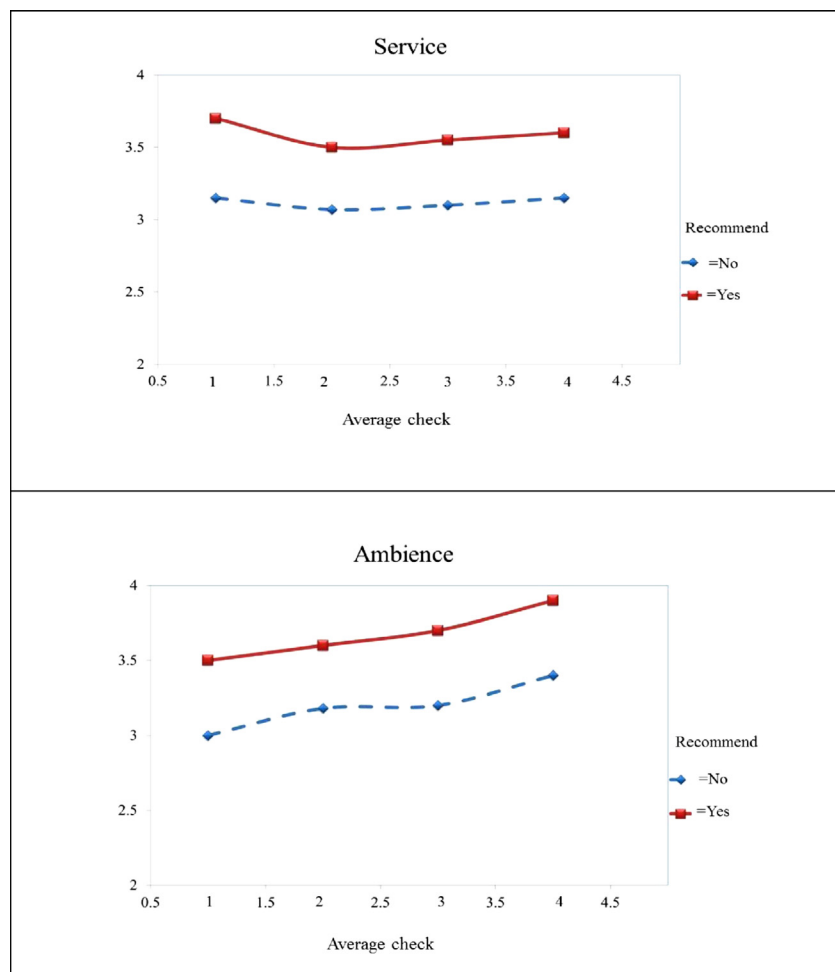
Despite the contributions made by this study, there also exist several limitations and suggestions for future research. First, as not every mystery shopper who partook in this study shares a similar background and their food and beverage preferences may be very different, this variation may have affected the outcome of their evaluations in a service restaurant environment. Because it is not easy to change a person's preferences in a short period of time, there could be some level of subjectivity when these mystery shoppers conduct their evaluations. Thus, problems of common method variance may arise. Second, consumers often assume that a higher bill means a higher rating. Thus, the mystery shoppers may have this preconception while rating the restaurants. In other words, the evaluator's assumptions or expectations may affect his/her evaluation results. Third, this study's sample focuses primarily on full-service restaurants. Restaurant typologies are wide and varied. Some examples include bakery cafes, buffet restaurants, coffeehouses, diners, drive-in restaurants, ethnic restaurants, fast-food restaurants, gluten-free restaurants, ice cream and yogurt parlors, juice bars, kosher restaurants, steakhouses, tea houses, special-themed restaurants and vegetarian restaurants. It would be valuable to analyze other restaurant types using the newly developed restaurant rating scale.

## Acknowledgment

The authors would like to thank the anonymous reviewers for useful suggestions and the Taipei City Government.

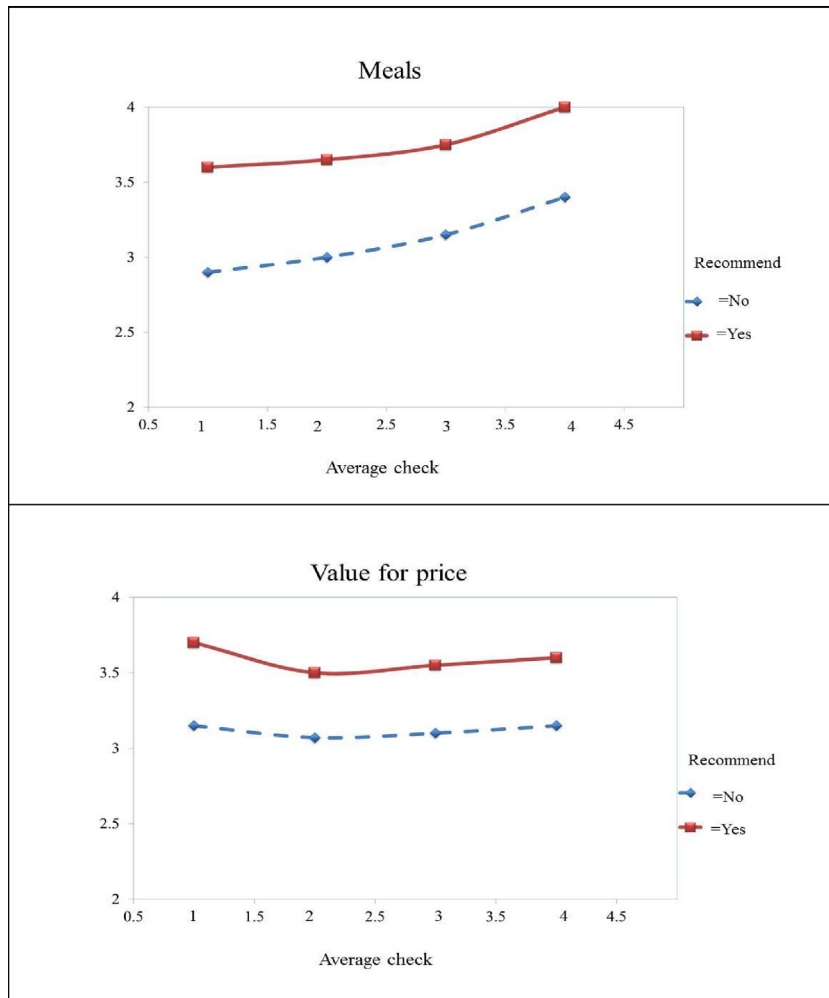
**Appendix 1. Experts background**

| No. | Title of position         | Experience | Background                                           |
|-----|---------------------------|------------|------------------------------------------------------|
| A1  | Professor                 | 15         | Tourism marketing and promotion                      |
| A2  | Assistant manager         | 25         | Restaurant marketing, transnational business         |
| A3  | Professor                 | 30         | Tourism marketing and promotion                      |
| A4  | Secretary Director        | 35         | External Trade Development Council                   |
| B1  | Chief Executive Officer   | 12         | Gourmet restaurant selection website                 |
| B2  | Chef                      | 25         | International hotel restaurant management            |
| B3  | Chef                      | 30         | International hotel restaurant management            |
| B4  | Director                  | 25         | Restaurant management                                |
| B5  | Chairman of the institute | 25         | Restaurant management                                |
| C1  | Professor                 | 20         | Restaurant management and mystery shopper research   |
| C2  | Consultant of government  | 28         | National hospitality service certification committee |
| C3  | Director of the institute | 40         | Mystery shopper research                             |
| C4  | Secretary Director        | 15         | Service quality and mystery shopper evaluation       |
| C5  | Secretary Director        | 16         | Service quality and mystery shopper evaluation       |

**Appendix 2. The result of cross tabulation between average check and recommending intention – Service and Ambience**

Note: Average Check: 1=< NTD\$300, 2= NTD\$301-500  
3=NTD\$501-1000, 4=> NTD\$1,000

### Appendix 3. The result of cross tabulation between average check and recommending intention – Meals and Value for price



Note: Average Check 1=<NTD\$300, 2=NTD\$301-500  
3=NTD\$501-1000, 4=>NTD\$1,000

### References

- Bagozzi, R.P., Yi, Y., 1988. On the evaluation for structural equation models. *J. Acad. Mark. Sci.* 16, 74–94.
- Beck, J., Miao, L., 2003. Mystery shopping in lodging properties as a measurement of service quality. *J. Qual. Assur. Hosp. Tour.* 4 (1–2), 1–21.
- Beck, J., Lalopa, J., Hall, J., 2004. Insuring quality service: training mystery shoppers. *J. Hum. Resour. Hosp. Tour.* 2 (2), 41–56.
- Bonett, D.G., Woodward, J.A., 1990. Testing residual normality in the ANOVA model. *J. Appl. Stat.* 17 (3), 383–387.
- Boyne, S., Hall, D., Williams, F., 2003. Policy, support and promotion for food-related tourism initiatives. *J. Travel Tour. Mark.* 14 (3–4), 131–154.
- Butcher, K., Sparks, B., McColl-Kennedy, J., 2009. Predictors of customer service training in hospitality firms. *IJHM* 28 (3), 389–396.
- Callan, R.J., 1990. Hotel award schemes as a measurement of service quality – an assessment by travel industry journalists as surrogate consumers. *IJHM* 9 (1), 45–58.
- Chaney, S., Ryan, C., 2012. Analyzing the evolution of Singapore's World Gourmet Summit: an example of gastronomic tourism. *IJHM* 31, 309–318.
- Cheng, C.C., Chen, C.T., Hsu, F.S., Hu, H.Y., 2012. Enhancing service quality improvement strategies of fine-dining restaurants: new insights from integrating a two-phase decision-making model of IPGA and DEMATEL analysis. *IJHM* 31 (4), 1155–1166.
- Cheng, S.W., Xu, F.F., Zhang, J., Zhang, Y.T., 2010. Tourists' attitudes toward tea tourism: a case study in Xinyang, China. *J. Travel Tour. Mark.* 27 (2), 211–220.
- Cobb, R., 1995. Magical mystery lure. *Marketing* 26 (October), 45.
- Cotter, M., Snyder, W., 1996. How mobile stars affect restaurant pricing behavior. *Cornell Hotel Restaurant Admin. Quart.* 37 (2), 34–41.
- Cotter, M., Snyder, W., 1998. How guide books affect restaurant behavior. *J. Restaurant & Foodser. Mark.* 3 (1), 69–75.
- Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16 (3), 297–334.
- Dwek, R., 1996. Magic of mystery shopping. *Marketing* 170, 41–44.
- Finn, A., 2001. Mystery shopper benchmarking of durable-goods chains and stores. *J. Serv. Res.* 3 (4), 310–320.
- Finn, A., Kayande, U., 1999. Unmasking a phantom: a psychometric assessment of mystery shopping. *J. Retail.* 75 (2), 195–217.
- Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18 (1), 39–50.
- Fox, R., 2007. Reinventing the gastronomic identity of Croatian tourist destinations. *IJHM* 26 (3), 546–559.
- Han, H., Kim, W., 2009. Outcomes of relational benefits: restaurant customers' perspective. *J. Travel Tour. Mark.* 26 (8), 820–835.
- Henderson, J.C., 2011. Celebrity chefs: expanding empires. *Br. Food J.* 113 (5), 613–624.

- Hesselink, M., Wiele, T.V.D., October 2003. Mystery Shopping: In-Depth Measurement of Customer Satisfaction. SSRN eLibrary [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1098525](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1098525)
- Johns, N., Pine, R., 2002. Consumer behavior in the food service industry: a review. *IJHM* 21, 119–134.
- Johnson, C., Surlmont, B., Nicod, P., Revaz, F., 2005. Behind the stars: a concise typology of Michelin restaurants in Europe. *Cornell Hotel Restaur. Adm. Q.* 46 (2), 170–187.
- Jöreskog, K.G., 1969. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* 34 (2), 183–202.
- Knutson, B.J., 2000. College students and fast food – how students perceive restaurant brands. *Cornell Hosp. Q.* 41 (3), 68–74.
- Lane, C., 2010. The Michelin-starred restaurant sector as a cultural industry: a cross-national comparison of restaurants in the UK and Germany. *Food Cult. Soc.* 13 (4), 493–519.
- Lee, F.S.J., 2013. Hospitality products and the consumer price – perceived quality heuristic: an empirical perspective. *Serv. Mark. Q.* 34 (3), 205–214.
- Litvin, S.W., Goldsmith, R.E., Pan, B., 2008. Electronic word-of-mouth in hospitality and tourism management. *Tour. Manage.* 29 (3), 458–468.
- Liu, Y., Jang, S.S., 2009. The effects of dining atmospherics: an extended Mehrabian–Russell model. *IJHM* 28 (4), 494–503.
- Luria, G., Yagil, D., 2008. Procedural justice, ethical climate and service outcomes in restaurants. *IJHM* 27 (2), 276–283.
- Lynn, M., 2002. Turnover's relationships with sales, tips and service across restaurants in a chain. *IJHM* 21 (4), 443–447.
- Mathe, K., Slevitch, L., 2013. An exploratory examination of supervisor undermining, employee involvement climate, and the effects on customer perceptions of service quality in quick-service restaurants. *J. Hosp. Tour. Res.* 37 (1), 29–50.
- Marini, G., Adriani, F., Scaramozzino, P., 2003. The Inflationary Consequences of a Currency Changeover: Evidence from the Michelin Red Guide. CEIS Tor Vergata – Research Paper Series No. 27.
- Michelin Red Guide, 2007. France: Hotels & Restaurants 2007. Michelin Travel Publications, France.
- Nield, K., Kozak, M., LeGrys, G., 2000. The role of food service in tourist satisfaction. *IJHM* 19, 375–384.
- Nilsson, J.H., Svärd, A.C., Widarsson, Å., Wirell, T., 2011. 'Cittáslow' eco-gastronomic heritage as a tool for destination development. *Curr. Issues Tour.* 14 (4), 373–386.
- Nunnally, J.C., 2010. *Psychometric Theory 3E*: Tata McGraw-Hill Education.
- Oronsky, C.R., Chathoth, P.K., 2007. An exploratory study examining information technology adoption and implementation in full-service restaurant firms. *IJHM* 26 (4), 941–956.
- Ottensbacher, M., Harrington, R.J., 2007. The innovation development process of Michelin-starred chefs. *Int. J. Contemp. Hosp. M.* 19 (6), 444–460.
- Ryu, K., Jang, S.C.S., 2007. The effect of environmental perceptions on behavioral intentions through emotions: The case of upscale restaurants. *J. Hospitality & Tourism Research* 31 (1), 56–72.
- Ryu, K., Jang, S.C., 2008a. DINESCAPE: A scale for customers' perception of dining environments. *J. Foodservice Business Res.* 11 (1), 2–22.
- Ryu, K., Jang, S.C., 2008b. Influence of restaurant's physical environments on emotion and behavioral intention. *Serv. Ind. J.* 28 (8), 1151–1165.
- Sethi, V., King, W.R., 1994. Development of measures to assess the extent to which an information technology application provides competitive advantage. *Manage. Sci.* 40 (12), 1601–1624.
- Snyder, W., Cotter, M., 1998. The Michelin Guide and restaurant pricing strategies. *J. Restaurant Foodservice Mark.* 3 (1), 51–67.
- Tam, W.Y., Yung, N.L.A., (MBA thesis) 2003. Managing customer for value in catering industry (fast food) in Hong Kong. The Chinese University of Hong Kong.
- Tourism Bureau Ministry of Transportation and Communications, Republic of China (Taiwan), 2014. Executive Information System. <http://admin.taiwan.net.tw/index.aspx>
- Tarantola, C., Vicard, P., Ntzuofras, I., 2012. Monitoring and improving Greek banking services using Bayesian networks: an analysis of mystery shopping data. *Expert Syst. Appl.* 39 (11), 10103–10111.
- UNWTO, 2010. Tourism Highlights 2010 Edition, Available from: [http://www.unwto.org/facts/eng/pdf/highlights/UNWTO\\_Highlights10\\_en\\_LR.pdf](http://www.unwto.org/facts/eng/pdf/highlights/UNWTO_Highlights10_en_LR.pdf)
- Urban, W., 2013. Perceived quality versus quality of processes: a meta concept of service quality measurement. *Serv. Ind. J.* 33 (2), 200–217.
- Wang, E.S.T., Tsai, B.K., Chen, T.L., Chang, S.C., 2012. The influence of emotions displayed and personal selling on customer behaviour intention. *Serv. Ind. J.* 32 (3), 353–366.
- Wiele, T.V.D., Hesselink, M., Iwaarden, J.V., 2005. Mystery shopping: a tool to develop insight into customer service provision. *Total Qual. Manage.* 16 (4), 529–541.
- Wilson, A.M., 1998a. The role of mystery shopping in the measurement of service performance. *Manage. Serv. Qual.* 8 (6), 414–420.
- Wilson, A.M., 1998b. The use of mystery shopping in the measurement of service delivery. *Serv. Ind. J.* 18 (3), 148–163.
- Wilson, A.M., 2001. Mystery shopping: using deception to measure service performance. *Psychol. Mark.* 18 (7), 721–734.
- Wu, C.H.J., Liang, R.D., 2009. Effect of experiential value on customer satisfaction with service encounters in luxury-hotel restaurants. *IJHM* 28 (4), 586–593.
- Zagat, S., 2007. *New York City Restaurants 2007*. Zagat Survey, LLC, New York.
- Zhang, Z., Ye, Q., Law, R., Li, Y., 2010. The impact of e-word-of-mouth on the online popularity of restaurants: a comparison of consumer reviews and editor reviews. *IJHM* 29 (4), 694–700.