Research article

GPTCCC: An instrument for measuring group package tour service

Kuo-Ching Wang[a], An-Tien Hsieh[b], Shu-Hui Chou[c],*, Yu-Shan Lin[d]

[a]*Graduate Institute of Recreation, Tourism, and Hospitality Management, National Chiayi University, Taiwan*
[b]*PH. D Program in Management, Da-Yeh University, Taiwan*
[c]*Department of Tourism, Providence University, Taiwan*
[d]*Graduate Institute of International Business Administration, Chinese Culture University, Taiwan*

Abstract

The aim of this paper is to develop and empirically validate an instrument (scale) which measures the group package tour (GPT) service. This study employed multistage steps for investigation; both qualitative and quantitative approaches were used to develop a new customer comment card (CCC) for GPT (GPTCCC). For qualitative data, literature review, three focus groups, and collection of 20 existing CCCs were employed. For quantitative data, responses from 521 and 312 tourists were used to assess the reliability and validity by exploratory and confirmatory factor analysis. Through a rigorous instrument development process, six factors were finally identified: hotel, transportation, shopping arrangement, optional tour, tour leader, and local guide. The study provides insight into the nature of the measurement of GPT service and how it might be used in practice. Implications of these findings for travel managers and areas for future research are also provided.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Service quality; Group package tour (GPT); Customer comment card (CCC)

1. Introduction

There has been dramatic growth in outbound travel from Asian countries in recent years, fueled by the region's rapid economic growth and rising income levels (China National Tourism Administration, 2005; March, 1997). The international tourism industry is now faced with an increasing number of inbound travelers from Asia, such as Australia (Reisinger & Turner, 2002) and Guam (Iverson, 1997). Besides, as a result of easing restrictions on outbound travel by China, the Chinese are likely to be enthusiastic tourists in the future. A recent article in *The Economist* indicated that destination countries like Germany could expect one million Chinese organized tour tourists by the end of the decade (Anonymous, 2003).

Although the travel market seems to be flourishing in Asia, as indicated in *The Annual Survey Reports on R.O.C. Outbound Travelers* (Tourism Bureau, 2002), and in Dev and Olsen's (2000) report of a think-tank process held in Manila, customers' satisfaction regarding the services provided by the hospitality and travel industries is declining. A similar situation has also been found in China (Wei & Clark, 2002).

There are numerous factors that might cause the reported decline of customer satisfaction, and the quality of the measurement tool could be one of the key factors (Stevens, Knutson, & Patton, 1995). In the service and hospitality industries, to measure customer's satisfaction and service quality, one widely used approach for obtaining customer feedback and other information is the customer comment card (hereafter abbreviated as CCC) (Barsky & Huxley, 1992; Cawley, 1998; Marvin, 1992; Wisner & Corney, 1997). Although, other tools for measuring customer satisfaction and service quality are also available such as focus groups, employee feedback, management observation, "800" telephone numbers, and sales data (Barsky & Huxley, 1992), Marvin (1992) once noted that the CCC plays an important role in the relationship between the service providers and their customers.

---

*Corresponding author.

E-mail addresses: wangkc@mail.ncyu.edu.tw (K.-C. Wang),
shchou@pu.edu.tw (S.-H. Chou).

Likewise, travel managers also utilize CCC to measure and manage the service quality of group package tours (hereafter abbreviated as GPT). As stated by several studies, GPT is one of the main modes of outbound travel in most Asian countries and areas, such as China, Japan, Taiwan, Korea, etc. (March, 2000; Prideaux, 1998; Wang, Hsieh, & Chen, 2002; Wang, Hsieh, Yeh, & Tsai, 2004; Wang & Sheldon, 1995; Yamamoto & Gill, 1999). About 40% of the international vacation travels made by Taiwanese people are GPTs (Tourism Bureau, 2005). Another example is in China, according to *The Year Book in China Tourism*, the number of outbound tourists in China has reached 28.85 million in 2004. Of these, 5.58 million travelers participated in GPT (China National Tourism Administration, 2005). In a typical GPT, as far as the present authors are aware, the CCC is one of the important methods used to measure service quality.

Despite the importance of CCC in travel management, there is a lack of empirical support regarding the content and effectiveness of CCC. To investigate this, sets of primary testimony were collected by this present study, which were composed of 10 in-depth interviews with senior travel managers and 20 CCCs collected from Taiwan's major travel operators (see Appendix). Two interesting phenomena were identified from the above exploratory investigations. The first one is that the existing CCCs used by travel operators seem to have several drawbacks, such as (1) rating points; (2) lack of clarity and precision; (3) too many double-barreled questions; (4) low return rate; (5) tour leaders' interference in the return delivery; (6) appearance design; and (7) data analysis. For example, during the in-depth interviews, one participant made the following comment "…as far as I am aware, most of the major wholesale travel operators, they do not have sufficient employees to take care of the data coding and analysis. Therefore, normally the managerial effectiveness of CCCs is not what it should be".

Another example of these drawbacks is about the tour leader's interference "…our company's CCC is fairly simple, about A4 size with a single-side typed format, duplication is mainly via the copy machine. Typically, we [the travel operator] will ask tour leaders to distribute the CCC to his/her tourists on the last day of the GPT itinerary and right after the GPT tourist has filled out the CCC, the tour leader is asked to get it back immediately".

The second noteworthy phenomenon is that the design of the CCC is commonly based on either a manager's objections/experiences or imitations of competing travel operators. For instance, several participants indicated that "…our company's CCC is designed and worked out solely by one senior employee from the marketing department…" or "imitation is quite a common phenomenon in practice, our CCC is no exception, we just look at how the other travel operators designed their CCCs…" It is just as Vogt and Fesenmaier (1995) stated that achieving service quality requires the provider to accurately interpret customer's needs and satisfactions.

Nevertheless, the service providers do not understand the level at which customers evaluate their experience and tend to under-rate the customer experience.

The preceding discussion implies that it is still doubtful whether the existing CCCs used by major travel operators for evaluating GPT service quality really can reflect the customers' opinions. This situation is fairly consistent with Mattsson (1994), where he argued that many mangers in travel agencies seem to have only an intuitive understanding of their customers' expectations, and fail to identify the specifics of the service process. Consequently, for travel managers, under the above circumstances, it seems that poor decision-making would be probable if based on the information provided by an ineffective CCC.

2. Research problem

Baker and Crompton (2000) indicated that performance quality is under management's control and it is the more useful measure. Since the existing CCCs entail many drawbacks, this study seeks to find some alternatives from the literature to solve the service quality measurement problems in GPT. As the literature reveals that many researchers have put their focus on analyzing the concepts of quality and consumer satisfaction such as Importance-Performance Analysis, SERVPERF, and SERVQUAL (Hudson, Hudson, & Miller, 2004).

Some researchers have modified ways of SERVQUAL or propose other models to measure service quality, but most of them still used the items developed from SERVQUAL (Boulding, Kalra, Staelin, & Zeithaml, 1993; Brady, Cronin, & Brand, 2002; Brown, Churchill, & Peter, 1993; Cronin & Taylor, 1992). The 22 items of SERVQUAL have become the most extensively used measures of service quality and have been applied in a variety of industries (Carman, 1990; Cook & Thompson, 2000; Cronin & Taylor, 1992; Fick & Ritchie, 1991).

However, the scale dimensionality of SERVQUAL appears to have a weak standing and the five dimensions are not as distinct and independent as one would wish (Llosa, Chandon, & Orsingher, 1998). In addition, in Hudson et al.'s (2004) measurement of service quality in the tour operating study, they did not force the attributes into the five original SERVQUAL dimensions. Furthermore, for the following discussions, the items of SERVQUAL seem to be an inappropriate method for specifically measuring service quality in GPT.

First, previous studies that utilized the scale dimensionality of SERVQUAL focused only on certain sectors of the tourism industries, e.g. airline, hotel, ski area, restaurant, and travel agent services (Albrecht, 1992; Bigné, Martínez, Miquel, & Andreu, 2003; Bojanic & Rosen, 1994; Fick & Ritchie, 1991; LeBlanc, 1992; Lee & Hing, 1995; Lam & Zhang, 1999; Ryan & Cliff, 1997; Saleh & Ryan, 1991; Tsai, Ryan, & Lockyer, 2002; Weiermair & Fuchs, 1999).

Although GPT is one sector of tourism industries, the services and characters of GPT are more complex

comparatively. For example, in Wang, Hsieh, and Huan's (2000) study on service features in GPT, they stated that one of the most distinctive characteristics of GPT is that it typically involves *many* travel-related industries all together, such as: local travel agencies, airlines, restaurants, souvenir stores, bus companies, etc. Reimer (1990) also indicated that tour operators combine various tourism products into a single entity and Swarbrooke and Horner (1999) argued that product quality can be seen as a jigsaw, with many equally important, but different sized pieces, that must all fit together perfectly in order to satisfy the tourist.

More specifically, the measurement of GPT includes all the services provided by travel operators itself and local service providers (hotel, restaurant, coach, and local guide). However, some service dimensions cannot be directly controlled by travel operators but are still important for the tourists' perceptions and experiences of the trip (Wang et al., 2000). Hence, the services provided by the travel operator itself and by overseas local travel agencies are both important to the tourists. The above-mentioned distinctive nature of the GPT results in the difficulty of service quality measurement. If the items of SERVQUAL are employed for the service quality measurement in GPT, it seems an inappropriate method for such a complex nature of GPT product. Specifically, as Dabholkar, Thorpe, and Rentz (1996) once indicated, although SERVQUAL has been empirically tested in a number of studies involving *pure* service settings; it has not been successfully adapted to and validated in a more complicated environment.

Consequently, questions like "XYZ has up-to-date equipment" and "XYZ's physical facilities are visually appealing" in SERVQUAL are too *pure* to measure the service quality in GPT. Most important of all, it could not cover all the related entities in GPT, if the items of SERVQUAL are adapted and intended for use as a quality measure for GPT, the scale items might be enormously complex.

Second, the development and confirmation of the SERVQUAL dimensions applied mostly to short-term service encounter, in which the interaction is limited between customers and service providers. For example, appliance repair and maintenance, a long-distance telephone provider, retail banks, and credit cards companies (Parasuraman, Zeithaml, & Berry, 1985), a dental school patient clinic, business school placement center, tire store, and acute care hospital (Carman, 1990), fast food, banking, pest control, and dry cleaning (Cronin & Taylor, 1992), spectator sports and entertainment, health care, long-distance carriers, and fast food (Brady et al., 2002), travel agencies (Albrecht, 1992; Bigné et al., 2003; Lam & Zhang, 1999; LeBlanc, 1992; Ryan & Cliff, 1997), real estate brokers (Johnson, Dotson, & Dunlop, 1988), public recreation programs (Crompton & Mackay, 1989), hotels (Fick & Ritchie, 1991), restaurants (Bojanic & Rosen, 1994; Fick & Ritchie, 1991; Lee & Hing, 1995), airlines

(Fick & Ritchie, 1991), ski area service (Fick & Ritchie, 1991), and retail stores (Dabholkar et al., 1996).

In encounters such as listed above, the interaction between consumer and service provider is limited; typically during the entire service encounter it would not directly get too many other tourism-related service providers involved. On the contrary, GPT is generally a long and continuous process. Governmental statistics show that the median length of stay overseas for GPT is 5.7 nights in Taiwan (Tourism Bureau, 2003). Moreover, during the entire itinerary, one prominent phenomenon is that two important service providers, namely, the tour leader and local guide interact with the tourists closely and hourly. In general, the tour leader is expected to offer 24 h service over the entire tour.

In short, although these early and recent service quality studies provided useful measurement for service managers, still this present study finds that prior studies could not fully solve the problem in measuring overall service quality in GPT. Therefore, this study attempts to capture the precise criteria which are used by consumers to evaluate the quality of GPT service among numerous dimensions, facilities, and service providers.

The relationship between managements' interaction with customers and their feedbacks has been the subject of number of research studies over the years. All of the research had found that CCC was necessary for service improvement. Lewis and Pizam (1981) pointed out guest surveys could offer crucial information for evaluating policies and making management decisions. They showed how traditional surveys fail and outlined a new method for getting the facts. McCleary and Weaver (1982) stated that customer questionnaires were suggested to identify customer service elements and to evaluate customer satisfaction provided by employees. Further, Sampson (1998) compared the nature of Web-based feedback collection over the Internet to the features of conventional (paper) comment cards. Evenson (2001) gave some suggestions how to do the survey with comment cards.

For travel managers, accurate information could be gained from valid service measurement items; further, an attractive appearance could also show customers how the travel operators valued their opinions and would encourage tourists to fill in the CCC more carefully (Cawley, 1998). On the other hand, despite the fact that the current CCCs used by most of the travel operators may still entail many drawbacks, Wisner and Corney (1997) stated that the CCC offers managers the opportunity to design the comment instrument to elicit information most important to the organization; it should be a more appropriate tool for service quality measurement in GPT. Consequently, the specific purpose of this study is to develop a service quality measure, namely the CCC, for GPT that would have desirable reliability and validity properties.

Churchill suggests a eight-stage paradigm for developing a scale, starting with stage one "specify domain of the construct" and ending with "developing norms." Many

researches employ Churchill's paradigm (e.g. Fodness, 1994; Geyskens & Steenkamp, 2000; Parasuraman, Zeithaml, & Berry, 1988; Sweeney & Soutar, 2001). Thus the process follows Churchill's (1979) approach for developing measures of multiple-item GPT constructs. After the development of an initial set of items, two-scale purification stages were undertaken. While the first stage used a student sample, the second stage used a more diverse sample of consumers.

3. Methodology

According to the purpose of this study, the overall process of developing the CCC was divided into four separate parts. *Part One*: generate sample of items. By means of a literature search, three focus groups, and 20 existing CCCs used by major travel operators. *Part Two*: collect data and purify measures. Based on Part One, an initial scale is generated; in this part, data are collected from the student samples for purifying the measure. *Part Three*: collect data from a more diverse sample. In this

part, another set of samples were used to further assess the CCC's reliability and validity. *Part Four*: collect data and comparison. Based on Part Three, a new designed CCC is developed. In this part, the samples were used to evaluate and compare the existing CCCs used in practice, with this new designed CCC via a questionnaire in light of the design appearance, content, practical utility, etc. Each of the above four parts are depicted in Fig. 1 and details are shown as follows:

3.1. Part one: items generation

To establish content validity, a comprehensive and representative set of important items of GPT consumer's concern is developed. According to Churchill's (1979) prescription, the research must be exacting in delineating what is included in the definition and what is excluded. Therefore, in this part this present study tries to generate a complete item list via three different ways: literature review, three focus groups, and 20 existing CCCs used by major travel operators.

Part One

✓ Specify domain of GPT service quality construct
✓ Generate item samples
 ● Literature review
 ● Three focus groups
 ● Twenty collected CCCs

✈ Initially, a pool of 359 items were generated
✈ The items were subsequently categorized by 2 judges
✈ Finally, 76 items scale obtained

Part Two

✓ Data collection (1) and purification of measures
 ● Data collected from 521 student samples
 ● Coefficient alpha
 ● Exploratory factor analysis
 ● Confirmatory factor analysis

✈ Additional questions for testing validity and reliability
✈ Generating a purified measure with 22 items

Part Three

✓ Data collection (2) and reanalysis of GPTCCC
 ● Data collected from 312 experienced GPT customers
 ● Coefficient alpha
 ● Exploratory factor analysis
 ● Confirmatory factor analysis

✈ Additional questions for testing validity and reliability
✈ Final list of 22 scale items (named GPTCCC)
✈ Design a new CCC

Part Four

✓ Data collection (3 and 4) and comparison
 ● Compare the 20 existing CCCs with the newly developed CCC
 ● Data collected from 646 GPT customers
 ● Independent-Samples T test analysis

✈ Significant differences were found
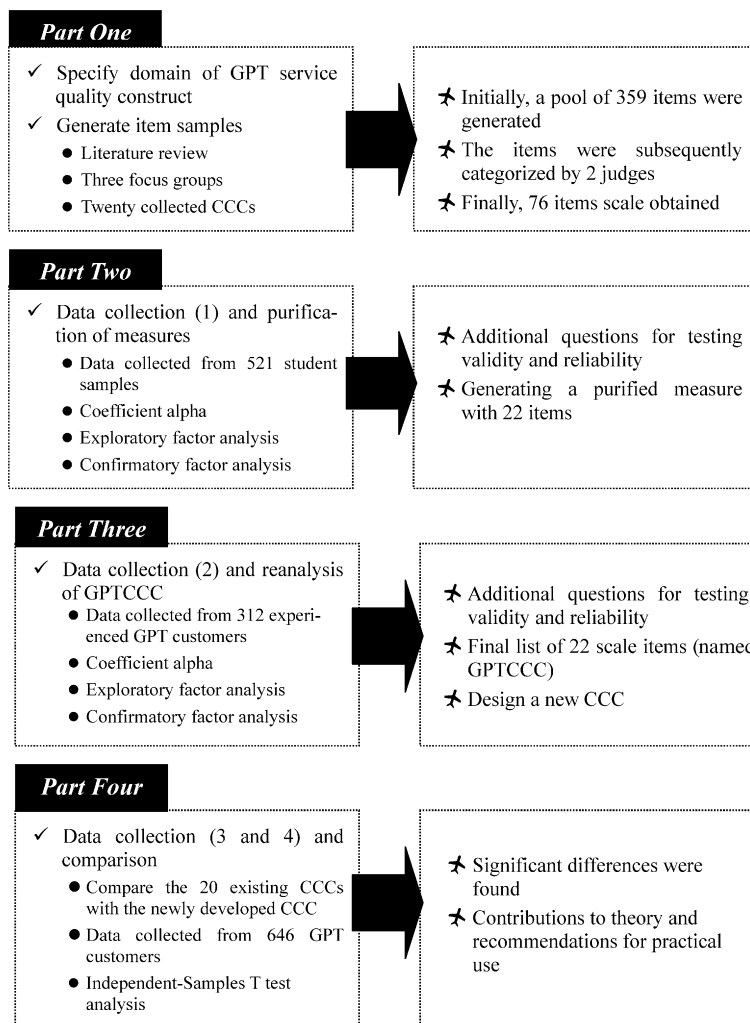✈ Contributions to theory and recommendations for practical use

Fig. 1. Flow chart of the scale development/investigation procedure.

3.1.1. Literature review

Depending on the specific context of the study, researchers differ on which elements they see as comprising GPT service quality measurement. Geva and Goldman's (1991) study of guided tours used 15 specific attributes, such as: meals, entertainment activities, etc. Altogether, these 15 attributes were employed for this present study. Another study conducted by Taiwan Tourism Bureau (1993) focused on the assessment of travel operators' service quality. In that study, measurement items such as language ability of the tour leader, manner of tip collection, arrangement of free time, etc., were used to evaluate service quality during the GPT; 34 items were then considered for the item generation. In addition, Wang et al.'s (2000) critical service features study of GPT found 25 service features from nine GPT sectors. Among the nine GPT sectors, except for the pre-tour briefing (involving four service features), 21 service features were considered relevant and critical to this study. In sum, from the above-mentioned studies, 70 service features were finally used as the sample items.

3.1.2. Three focus groups

As stated by Bowen (2001) and Ryan (1995), qualitative approach is more useful in identifying tourist satisfaction. Accordingly, focus groups were used to explore the variables of tourists concern in GPT. Three focus group interviews were conducted and used to characterize the GPT consumers' point of view. In total, 26 experienced GPT consumers participated in the studies (eight, nine, and nine participants, respectively). Among these participants, 11 were male and 15 were female with ages from 22 to 45 years old. Overall, 76.9% of the participants had taken, at least one GPT in the preceding 12 months and the mean participation in overseas GPTs was 3.7 times.

In the focus group interviews, open-ended questions were asked with respect to customer's GPT experiences. Questions were then developed and divided mainly according to Wang et al.'s (2000) GPT sectors suggestion, which included airport/plane, hotel, restaurant, coach, scenic-spot, shopping, optional tour, and others. Two sectors, tour leader and local guide, were also considered because both are vital to the GPT service quality, as identified in prior studies (Geva & Goldman, 1991; Quiroga, 1990; Schmidt, 1979). The following is an example of a question in the GPT hotel sector:

> According to your personal GPT experiences, what factors will affect your evaluation on the GPT service quality of the hotel sector?

The dialogs in the three focus groups were tape-recorded and then transcribed. As indicated by Kassarjian (1977), the first step in data analysis is to determine the appropriate unit of analysis. Therefore, content analysis method was incorporated here. Two doctoral students (one of them had industrial experience) served as judges and independently coded the transcriptions first into 227 units of analysis and then categorized them into 77 items.

3.1.3. Twenty existing CCCs used by major travel operators

The 20 CCCs represent the practicing managers' viewpoints. CCCs were collected mainly from the major travel operators in Taipei (see Appendix). In total, 212 items were generated from the 20 CCCs.

3.1.4. Content validity of items

Finally, by means of the literature search, three focus groups, and 20 existing CCCs used in practice, a total of 359 items were obtained (70/literature review, 77/three focus groups, and 212/20 CCCs). The single classification concept for category development as recommended by Weber (1990) was employed here. In an iterative process conducted by judges A and B, each of the units was read out, classified, re-read, and re-classified. In the end, 76 items under nine GPT sectors had emerged. Finally, for the intrajudge reliability testing, this study introduced a time lag of 2 weeks (Davis & Cosenza, 1993). The result of intrajudge reliability was above 0.90 and no new GPT sectors or items emerged. The above-mentioned three approaches used to generate items along with the rigorous categorization process suggest the scale has content validity.

3.2. Part two: data collection (1) and purification of measures

3.2.1. Questionnaire development

Initially, 76 items were re-written as questionnaire to form an original 7-point scale. Each item is anchored by "strongly agree" (7) to "strongly disagree" (1), with no verbal labels for scale points 2–6. A performance-based measure was chosen in terms of recommendation both by previous scale development and service quality researchers (Cronin & Taylor, 1994; Sweeney & Soutar, 2001; Yuksel & Rimmington, 1998). Furthermore, to enable an assessment of convergent, discriminant, and criterion-related validity of the constructs, sets of additional questions derived from the literature were added.

First, for convergent validity, nine additional items were used to measure satisfaction for each one of the GPT sectors (e.g. "I am satisfied with the hotel arrangements of my recent GPT" is inserted into hotel sector). In addition, 1 item (I am satisfied with my recent GPT) was included to measure the overall satisfaction (Homburg & Rudolph, 2001; Parasuraman, Zeithaml, & Berry, 1991; Sweeney & Soutar, 2001).

For discriminant validity, as suggested by Brown et al. (1993, p. 133), Five items derived from a consumer discontent scale were then used (e.g. This GPT that I purchased was not the same as advertised) (Lundstrom & Lamont, 1976). Moreover, as suggested by prior studies, Geva and Goldman (1991) and Parasuraman, Zeithaml, and Berry (1988), two items were measured for criterion

(predictive) validity: (1) intention to recommend the travel operator to others and (2) intention to purchase future tours from the same travel operator. Furthermore, a previous study had indicated that social desirability and yea-saying biases could occur when subjects respond to a questionnaire, threatening the reliability and validity of the other measures (Bagozzi & Baumgartner, 1994). Therefore, five items from the Marlowe–Crowne social desirability scale (Crowne & Marlowe, 1964; Paulhus, 1984) and five items from the YN-2 scale (Goldsmith, 1987) were included in the questionnaire (Lafferty & Goldsmith, 1999). All the above measurements were measured by a 7-point scale anchored by "*strongly agree*" (7) to "*strongly disagree*" (1) with no verbal labels for scale points 2–6.

Several questions were included to capture the respondents' demographic attributes. Finally, before data collection was actually conducted, two doctoral and six graduate students were invited to assess the content and relevance of the 76 items. In addition, 30 undergraduate students were also invited to assess whether words and phrases of any of the items were unclear. The student sample applied here as Cooper and Schindler (2003, p. 229, 239) suggested, for reliability, well-trained, supervised, and motivated persons were employed to conduct the investigation to minimize external sources of variation. In addition, the student population also represents an important segment of the tourism market (Chadee & Mattsson, 1996).

3.2.2. Data collection (sample 1)

For further purifying the measure, data were collected from student samples and analyzed using internal consistency analysis and exploratory factor analysis. Four Colleges of Chinese Culture University were chosen randomly to participate in this stage. The Department of Tourism Management was deliberately excluded so as to avoid the possibility of judgments by experts (Freiden, 1984). In total, 43 classes participated in this investigation. Respondents were required to have had at least one GPT experience in the preceding 2 years and each student was asked to recall the most recent GPT experience to answer the questions. From these 43 classes, 555 samples were obtained, of which 34 samples were excluded because they: (i) did not detect the two reverse questions; (ii) did not fit the time (2 years) requirement criterion; or (iii) offered incomplete answers. Finally, 521 respondents were obtained and thus, deemed usable. The profile of participants (*Sample 1*) is shown in Table 1.

3.2.3. Item reduction and exploratory factor analysis

According to the suggestions by Churchill (1979), an iterative-scale purification procedure was used to develop a reduced, more parsimonious scale. First, item-to-total correlations were computed for the 76 items. Items with correlations near zero would be eliminated; further, items that produced a substantial or sudden drop in the item-to-

Table 1
Profile of participants—*Samples 1 and 2*

Sample 1 (N = 521)			Sample 2 (N = 312)		
Variables	N	%	Variable	N	%
Gender			Gender		
Male	175	33.6	Male	104	33.3
Female	346	66.4	Female	208	66.7
Age			Age		
18–20	73	14.1	20 and below	9	2.9
21–23	332	63.7	21–30	159	50.9
24–26	102	19.6	31–40	71	22.5
27–29	14	1.7	41–50	44	14.1
			51–60	24	7.7
Grade			60 and above	5	1.6
Freshman	34	6.5	Highest education level		
Sophomore	59	11.3	Junior high school	7	2.2
Junior	101	19.4	Senior high school	33	10.6
Senior	301	57.8	College	77	24.7
Postgraduate	26	5.0	University	128	41.0
			Graduate school and above	67	21.5
Travel experience					
1–3 times	341	65.5	Travel experience		
4–6 times	127	24.3	1–3 times	166	53.2
7–9 times	27	5.2	4–6 times	82	26.3
10 times and above	26	5.0	7–9 times	34	10.9
			10 times and above	30	9.6

total correlations were deleted (Churchill, 1979). After this, 38 items remained.

Next, for the purposes of deleting a small number of items that had their highest loading on an incorrect factor or an almost equal loading on more than one factor, and to reveal a clearer factor pattern, a principal component analysis with oblique rotation was applied (Gerbing & Anderson, 1988). Then, the study checked for a possible overlap of items across factors. After the iterative deletion of items with absolute factor loading values less than 0.7 (Hair, Anderson, Tatham, & Black, 1991), 22 items remained. Then, exploratory factor analysis (with scree test) confirmed that there were six factors underlying the GPT service quality construct. These factors included hotel, transportation, shopping arrangement, optional tour arrangement, tour leader, and local guide. The factor analysis and associated statistics are presented in Table 2. Combined factor loadings accounted for 81.02% of the total variance in the factor pattern.

In the process of streamlining from 76 items to the present 22 items, we had actually gone through various deletion criteria. However, for practical considerations, the authors viewed GPTCCC as a tool that has to be simple and symbolic. The authors also referenced commonly used service quality measurement-scale items, such as SERVQUAL. Finally, the authors had decided to adopt a high criterion (Hair et al., 1991) so that the items left are less and possess greater variance.

Table 2
Results of exploratory factor analysis and confirmatory factor analysis—*Sample 1* (*N* = 521)

Factor	Items	Factor loading	Mean	Cumulative variance (%)	Individual item reliability	t-value of factor loading	Construct reliability	Average variance extracted	Coefficient α
Hotel sector				45.97			0.90	0.72	0.91
H1	The grade of hotel arrangement is appropriate	0.79	5.22		0.56	17.92			
H2	The hotels have a comfortable staying environment	0.90	5.39		0.83	22.44			
H3	The hotels have a sanitary environment	0.90	5.49		0.79	21.82			
H4	The facilities of the hotel are good	0.87	5.09		0.69	20.16			
Transportation sector				56.36			0.90	0.82	0.90
R1	The coach is clean and tidy	0.93	5.04		0.79	20.30			
R2	Safety of the coach is reliable	0.90	5.09		0.84	20.09			
Shopping arrangement sector				65.70			0.87	0.64	0.90
S1	Frequency of shopping is appropriate	0.82	4.33		0.65	18.65			
S2	Shopping stores fit in with the needs of group members	0.92	4.12		0.74	20.13			
S3	Shopping stores have good reputation	0.76	4.40		0.64	18.56			
S4	Duration in shopping stores is appropriate	0.85	4.12		0.75	20.25			
Optional tour sector				71.42			0.89	0.68	0.89
O1	Provided detailed descriptions of the contents of optional tours	0.77	4.78		0.61	16.94			
O2	The price of optional tours is reasonable	0.95	4.21		0.79	19.01			
O3	The optional tours are safe	0.82	4.57		0.72	18.30			
O4	Appropriate arrangements for those who did not participate in the optional tours	0.81	4.07		0.58	16.55			
Tour leader sector				76.36			0.96	0.79	0.96
T1	The tour leader has a good presentation ability	0.87	5.03		0.82	19.07			
T2	The tour leader has a sense of responsibility	0.89	5.04		0.85	19.34			
T3	The tour leader shows friendliness	0.85	5.13		0.73	18.18			
T4	The tour leader has an interpretive ability	0.92	4.83		0.75	18.23			
T5	The tour leader has a professional ability	0.93	4.85		0.74	18.17			
T6	The tour leader has an ability of coordination within group members	0.85	5.02		0.77	18.57			
Local guide sector				81.02			0.83	0.72	0.82
L1	Local guide has a professional ability	0.90	5.39		0.61	15.35			
L2	Local guide is skillful group leading	0.90	5.16		0.82	14.49			

3.2.4. Confirmatory factor analysis

Subsequently, for the purpose of verifying the reliability and construct validity of the scale, confirmatory factor analysis was employed for parameter estimation (Jöreskog & Sörbom, 1993). Every single factor was then submitted to a confirmatory factor analysis. All factor loadings were found to be significant at the 0.01 level and all individual item reliabilities were above the required value of 0.4

(Bagozzi & Baumgartner, 1994, p. 402). According to the recommendations of Bagozzi and Yi (1988) and Bagozzi and Baumgartner (1994), an average variance extracted of at least 0.5 and a composite reliability of at least 0.7 is desirable, and those requirements were met.

After having assessed the individual factors, the reduced set of items was subjected, all together, to a confirmatory factor analysis using maximum likelihood estimation. The results of analysis are summarized in Table 1, together with some additional information on reliability and validity. Although the χ^2 value was significant (462.04 with 200 df, $p < 0.001$) (Sweeney & Soutar, 2001), other goodness-of-fit measures indicated a good overall fit of the six-factor model to the data: GFI = 0.93, AGFI = 0.91, SRMR = 0.035, RMSEA = 0.05, NFI = 0.95, NNFI = 0.97, RFI = 0.95, and CFI = 0.97. In summary, these criteria seem to suggest that the model fits the data adequately. The 22 detailed items are shown in Table 2, which were named as GPTCCC.

3.2.5. Construct validity

The next step was to assess the convergent and discriminant validity of the GPTCCC. One way to assess convergent validity is to check if all factor loadings are significant (Bagozzi, Yi, & Singh, 1991). As can be seen in Table 2, all factor loadings were significantly different from zero as evidenced by consistently large t-values. Convergent validity is also supported since the average variance extracted clearly exceeded 0.50 for all dimensions (Fornell & Larcker, 1981). Moreover, this study also assessed convergent validity by examining the relationship between each dimension of the GPTCCC scale and the overall satisfaction measure of GPT. The correlation coefficients were all significant at the 0.01 level, and the results are shown in Table 3.

The discriminant validity of the six-dimensional scale was investigated in two ways. First, as suggested by Fornell and Larcker (1981), when taking any pair of constructs, the average variance extracted for each construct should be greater than the squared structural path coefficient between the two constructs. These requirements were met with all

pairs of constructs with the average variance extracted ranging from 0.64 to 0.82 (the maximum of the squared path was 0.59) (see Fig. 2). Second, discriminant validity was also assessed through an examination of the correlations of the six dimensions with the five items of the consumer discontent scale. As shown in Table 3, the correlation coefficients were all significantly negative at the 0.05 level. These results support the distinction of the constructs included in the model.

Finally, the criterion-related validity was assessed by examining its relationship with other conceptually related variables. These included a willingness to recommend and behavioral intention. The correlation coefficients were all significant at the 0.01 level; the results were also demonstrated in Table 3.

Regarding the social desirability and yea-saying measurement, the correlations were small and insignificant at the 0.05 level. This indicated that the result was not contaminated by the social desirability and yea-saying factors. In summary, this study found evidence of convergent, discriminant, and criterion validity and any upward bias was apparently unrelated to individual differences.

3.3. Part three: data collection (2) and reanalysis of GPTCCC

After the above procedures and analyses, a purified 22-item scale has been generated. However, the reliability and validity of this parsimonious scale was evaluated once again on a separate sample as suggested by Churchill (1979). This study reanalyzed this developed scale on the basis of new data to evaluate the robustness of GPTCCC in the new sample.

3.3.1. Data collection (sample 2)

The data were gathered over a period of 2 months. In this stage, student samples were excluded; adults aged 18 and above were solicited to give responses in the major cities of Taiwan. In total, 1000 questionnaires were sent

Table 3
Results of construct validity—*Sample 1* ($N = 521$)

Validity/sector	Hotel	Transportation	Shopping	Optional tour	Tour leader	Local guide	Scale
Convergent							
Overall satisfaction	0.583*	0.496*	0.520*	0.577*	0.662*	0.599*	0.577*
Discriminant							
Consumer discontent	−0.163*	−0.094**	−0.112**	−0.155*	−0.221*	−0.134*	−0.095*
Criterion							
Behavioral intention	0.616*	0.447*	0.490*	0.542*	0.646*	0.586*	0.557*
Recommend	0.563*	0.425*	0.489*	0.503*	0.700*	0.538*	0.516*

*Correlation coefficients are significant at the 0.05 level.
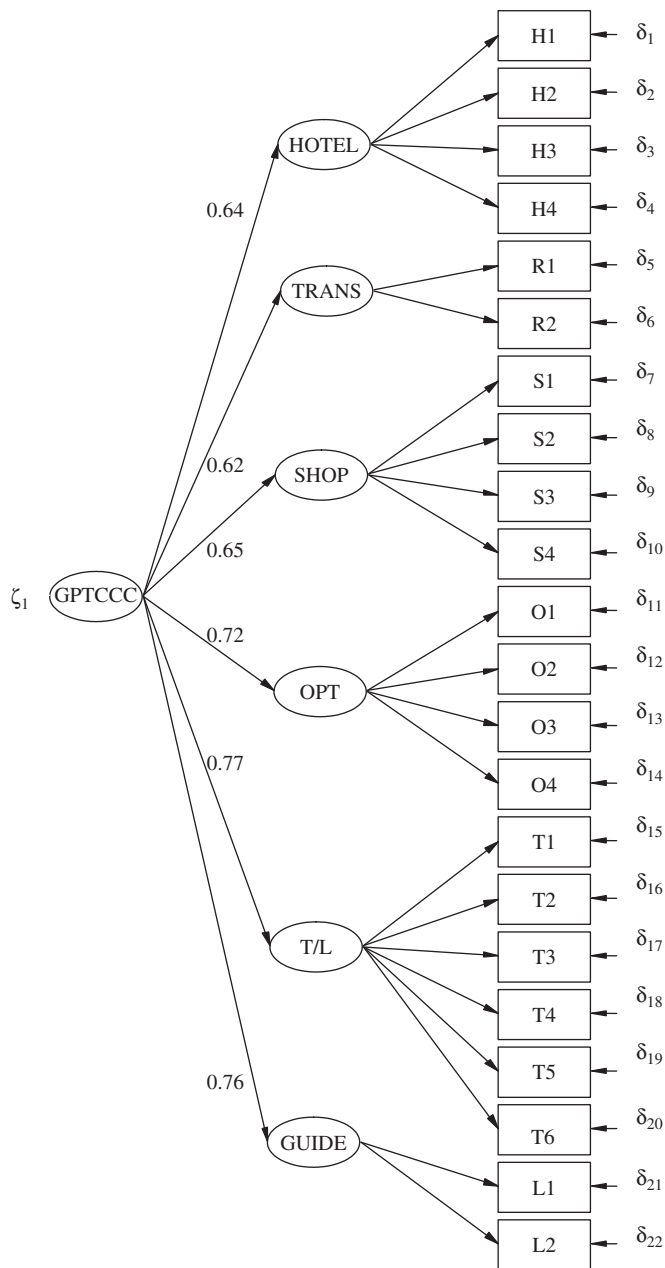**Correlation coefficients are significant at the 0.01 level.

Fig. 2. Completely standardized solution of GPT service measure—Sample 1 ($N = 521$).

3.3.2. Reanalysis of GPTCCC

This study then reanalyzed the developed scale on the basis of the new sample. The procedure involved several steps, similar to those used for *Sample 1*. Table 4 summarizes the results of a confirmatory factor analysis of the 22 GPTCCC items with six underlying factors based on *Sample 2*. All of the measures shown in Table 4 supported the good psychometric properties of the GPTCCC. The overall fit indices for *Sample 2* (χ^2 value of 313.38 with 196 df, $p < 0.001$, GFI = 0.92, AGFI = 0.89, SRMR = 0.036, RMSEA = 0.044, NFI = 0.94, NNFI = 0.97, RFI = 0.93, and CFI = 0.98) are similar to the indices observed in *Sample 1* and provide evidence of a desirable fit of the model in this new sample. Furthermore, convergent validity is evident in *Sample 2*: all factor loadings are highly significant. Discriminant validity has been tested and supported using Fornell and Larcker's (1981) discriminant validity test and alternative measures of consumer discontent. In summary, the six-factor model has also shown sound psychometric properties in *Sample 2*. Other construct validity was also examined, as shown in Table 5.

3.4. Part four: data collection (3 and 4) and comparison

Once the reliability and validity of the CCC were confirmed in Part Three, as suggested by Sampson (1998), the format of customer feedback instruments is essentially important. Therefore, in this part, eight CCCs were first selected from 20 existing CCCs used by major travel operators for the purpose of finding out the advantages of the format and design from the customer's point of view. Finally, a comparison is conducted between eight selected CCCs with the CCC which is developed by this study. The above procedures are discussed in the next section in more detail.

3.4.1. The selection process for eight CCCs

Due to the similar appearance of existing CCCs used by major travel operators, and to reduce the difficulty of comparison, this study took two stages to pick out eight from the 20 CCCs. To avoid likely biases, a rigorous procedure for selecting the eight CCCs was conducted. First, three graduate students and one doctoral student were employed independently as judges, and ranked the 20 CCCs according to established criteria such as content items and appearance. The mean scores of four judges were then computed and the 20 CCCs were ranked from 1 (the best) to 20 (the worst).

Based on the ranking, the 20 CCCs were further distributed into four groups. With each group having five CCCs (e.g. one group consisted of the CCCs ranked 1, 5, 9, 13, and 17). Subsequently, in the second stage, 120 students were invited as judges for each group (30 students per group). They were asked to rank the five CCCs from the best to the worst.

Mean scores were computed for each of the CCCs and those with the highest and the second highest mean scores

out, 378 surveys were returned, and of those, 312 were useable for the purpose of reanalysis. According to Fodness (1994), a common heuristic in marketing research is to use at least five cases per item in factor analysis. The use of 312 cases to purify a 22-item scale fell comfortably in excess of this parameter.

The result showed that among the 312 usable samples, 66.7% of the respondents were female and 33.3% were male. This characteristic of the sample corresponds to the characteristics of GPT tourists in Taiwan, where nearly 60% of GPT tourists are females (Tourism Bureau, 2004). The profile of participants (*Sample 2*) is shown in Table 1.

Table 4
Results of reliability and confirmatory factor analysis—*Sample 2* ($N = 312$)

Factor	Items	Individual item reliability	t-value of factor loading	Construct reliability	Average variance extracted	Coefficient α
Hotel sector				0.76	0.74	0.89
H1	The grade of hotel arrangement is appropriate	0.63	14.24			
H2	The hotels have a comfortable staying environment	0.84	16.69			
H3	The hotels have a sanitary environment	0.76	15.81			
H4	The facilities of the hotel are good	0.71	15.23			
Transportation sector				0.88	0.79	0.87
R1	The coach is clean and tidy	0.76	13.03			
R2	Safety of the coach is reliable	0.76	12.07			
Shopping arrangement sector				0.91	0.72	0.91
S1	Frequency of shopping is appropriate	0.73	14.47			
S2	Shopping stores fit in with the needs of group members	0.68	13.92			
S3	Shopping stores have good reputation	0.69	13.96			
S4	Duration in shopping stores is appropriate	0.80	15.20			
Optional tour sector				0.88	0.66	0.88
O1	Provided detailed descriptions of the contents of optional tours	0.66	11.74			
O2	The price of optional tours is reasonable	0.57	11.11			
O3	The optional tours are safe	0.60	11.32			
O4	Appropriate arrangements for those who did not participate in the optional tours	0.79	12.38			
Tour leader sector				0.95	0.71	0.94
T1	The tour leader has a good presentation ability	0.77	13.58			
T2	The tour leader has a sense of responsibility	0.79	13.78			
T3	The tour leader shows friendliness	0.73	12.09			
T4	The tour leader has an interpretive ability	0.76	13.38			
T5	The tour leader has a professional ability	0.77	13.51			
T6	The tour leader has an ability of coordination within group members	0.78	13.58			
Local guide sector				0.82	0.71	0.98
L1	Local guide has a professional ability	0.61	12.58			
L2	Local guide is skillful group leading	0.82	12.47			

were selected from these four groups. In total, eight CCCs were finally selected. In order to test interjudge reliability, the Kendall coefficients of concordances (W) were computed. The results seem sound for each of these four groups ($p < 0.001$, Kendall W were 0.581, 0.397, 0.693, and 0.589, respectively), and indicated the consistency between judges.

3.4.2. Subject selection for sample 3

For the purpose of finding out the strengths of these eight CCCs, an on-site intercept interview procedure was utilized. Each CCC was evaluated by 50 customers who had at least one GPT experience. The respondents were asked to evaluate the CCCs in terms of appearance,

Table 5
Results of construct validity—*Sample 2* (N = 312)

Validity/sector	Hotel	Transportation	Shopping	Optional tour	Tour leader	Local guide	Scale
Convergent							
Overall satisfaction	0.528*	0.435*	0.522*	0.173*	0.509*	0.225*	0.588*
Discriminant							
Consumer discontent	−0.205*	−0.096**	−0.276*	−0.080	−0.218*	−0.142**	−0.265*
Criterion							
Behavioral intention	0.481*	0.421*	0.443*	0.120**	0.570*	0.187*	0.538*
Recommend	0.516*	0.405*	0.467*	0.179*	0.453*	0.197*	0.544*

*Correlation coefficients are significant at the 0.01 level.
**Correlation coefficients are significant at the 0.05 level.

content, and practical utility (Cooper & Schindler, 2003, p. 230; Devellis, 1991, p. 77). The questionnaire consisted of 19 items on a 5-point Likert-type scale ranging from "*strongly agree*" (5) to "*strongly disagree*" (1).

The result showed that among the 400 usable responses returned, 67.8% were from female and 32.2% were from male. Most of them were 21–30 years old (46.0%), followed by 41–50 years old (19.0%). Over half of the respondents had a college educational background (50.3%), finally, 36.8% of the respondents had 3–5 travel experiences abroad.

The ANOVA method was used to analyze these eight CCCs, and the results provided useful information for designing a new CCC. Some factors are significantly important from the customer's point of view, such as "providing high levels of clarity and precision", "with open-ended comments and enough space", "the size of the printed characters should not be too small", "special designed layout of a printed page (traditional A4 print is not popular)", and "sealable design after completion". In short, apparently the results showed that customers prefer the precise questions and elaborate design and these findings are fairly similar with the Business Research Lab counsels (cf. Cawley, 1998):

"Keep the questionnaire short". Long questionnaires are a turn-off to most people. Make the questionnaire attractive. This includes the layout of the questions, the amount of 'white space' in the document (the more the better), and the font and color used.

Consequently, this present study tried to design a new CCC based on these above-mentioned findings.

3.4.3. Format and content of the new designed CCC
Based on the results of item development and the investigation of the CCCs' format and appearance, the new CCC is designed as a pamphlet with six (double-sided) pages, and A6 sized (41/5 in wide and 6 in high) so as to meet the customer's needs and to increase the managerial efficiency of the travel operator. The elucidations of the original design concepts are as follows.

There are in total 28 closed questions in the newly developed CCC. The question items mainly include six factors with 22 questions derived from the scale development of this study. Moreover, two important questions, the overall satisfaction with this trip and repurchasing intention, were added as suggested by Lewis and Pizam (1981) and Evenson (2001). Two buyer behavioral questions regarding how the customer selected the travel operator and two marketing questions that try to discern the customer's future travel planning (Which country/area would you like to visit next time?) are also employed.

For encouraging feedback and to catch customers' attention, a colorful design (Cawley, 1998) and a personal note from the general manager of the travel operator indicated "we value your opinions" and "your feedback is important" (Sampson, 1998) were employed on the inside cover. An instruction was included for explaining how to fill in this CCC. Moreover, the middle page of the pamphlet also pointed out that the responses and acknowledgements to customers' opinions will be shown on an exclusive website (Sampson, 1998). Finally, the back of the cover displayed the record of awards to improve the travel operator's reputation.

With the assistance of a professional graphic designer, the inner parts of the pamphlet included three innovative features, namely, a computer readable answering card, a coupon, and a security sticker. The computer readable answering card has two sides. The front includes a 5-point Likert-type scale ranging from "*strongly agree*" to "*strongly disagree*" for those closed questions, as suggested by Marvin (1992), the new CCC used the following scale design, which is clearly self-evident and more fun (☺ ☺ ☺ ☺ ☺). In addition, there was a bar code for the group number and the name of the tour leader in order to connect the tour leader's performance with customers' appraisals (Villanova, 1992). In short, standardized information could be obtained from the CCC through this computer readable design, which allows travel managers to perform statistical analysis that helps in directing strategies for service improvement easily and quickly. The reverse side was designed for three open-ended

questions, permitting the customers to put forward points of view not asked by the closed questions. The "white space" for answers was designed to be extensive as possible, in accordance with the finding of *Sample 3* that tourists prefer enough space while filling out their comment card.

Regarding the coupon, as suggested by Lewis and Pizam (1981) and Marvin (1992), management should offer a reward or incentive for completion and return of the survey. Hence, a specially designed coupon was enclosed with the pamphlet, so that after filling in the comment card, the coupon could be taken away and kept by the tourists for their next consumption.

Finally, since the employees have good reason not to tell management about mistakes and customer complaints (Ross, 1994), while filling in this new CCC the tourists will be constructed to follow two steps. First, put the answering card into the pamphlet and second, paste the security sticker on the pamphlet cover and then return it to the tour leader. The security sticker is made from low-viscosity tape to avoid damaging the paper texture, so the pamphlet can be reused. The security sticker will break when removed so this would let the tourists fill in the card with greater confidence. It could also prevent tour leaders from sifting through the pamphlets in search of negative comments. (Copies of the new designed CCC are available from the first author.)

3.4.4. Comparison between the new CCC with the existing eight CCCs

After the new CCC had been designed, this study utilized another sample group (*Sample 4*) to examine and compare this new CCC with the existing eight CCCs used by major travel operators. The sample selection criterion and questionnaire are exactly the same as those used in *Sample 3*. Two hundreds and sixty respondents were invited to evaluate the newly designed CCC. Among the 246 usable responses, 58.5% were female and 41.5% were male. Most of them were between 21 and 30 years old (43.5%), followed by 31–40 years old (24.8%). Nearly 42% of the respondents had a college educational background and travel experiences were between 2 and 4 times (44.9%).

The comparison was conducted via Independent-Samples *t*-test analysis. The existing eight CCCs were aggregated into group 1 and the new CCC became group 2. The comparisons were conducted regarding the format design, content, and overall evaluation. Some significant differences were found as shown in Table 6. Apparently, from the customers' perspectives, the new CCC developed by this present study seemed superior to the existing eight CCCs used by major travel operators.

4. Discussion and conclusions

This study has tried to combine both qualitative and quantitative methods, attempting to establish an instrument suitable for evaluation from the perspectives of consumers and travel operators with a scale based on precise scientific methods. Past studies have either focused on the integral performance of the service quality of the travel operator or they explored the service features of the GPT leader and the characteristics of package tours. They have rarely examined the development and evaluation of specific features of the service quality of GPT. The scale developed in this study, formulated from its initial stage to the final version, has met rigorous criteria for both validity and reliability. The results have important practical and theoretical implications that benefit both practitioners and researchers. There are several noteworthy findings from the construction of the overall scale.

4.1. Items of shopping and optional tour

With respect to the 22 question items found in this study, the results are strikingly different from the existing CCCs used by major travel operators. For example, the CCC question items presently used by travel operators to evaluate the shopping and optional tour sectors of GPT are seriously lacking. For example, only four out of the 20 existing CCCs collected in this study have items to evaluate shopping and optional tour sectors. Nonetheless, the results of this present study showed that, items like appropriate frequency of shopping, whether or not shopping stores fit in with the needs of group members, availability of detailed advance description of the contents of optional tours, reasonable pricing of optional tours, etc. All concern tourists substantially and affect the overall evaluation of the GPT service quality.

Shopping is a common and preferred tourist activity in many destinations (Timothy & Butler, 1995). It seems plausible to explain why shopping and optional tours were perceived as important sectors from traveling risk perspectives. For example, Roehl and Fesenmaier (1992) once used seven different types of risk, namely the equipment risk, financial risk, physical risk, psychological risk, satisfaction

Table 6
The comparison of existing eight CCCs with new CCC—*Samples 3 and 4* ($N = 400, 246$)

	Groups 1 and 2	N	Mean[a]	t
Format design	Existing 8	400	3.64 (0.51)	4.012*
	New	246	3.89 (0.44)	
Content	Existing 8	400	3.53 (0.57)	0.656*
	New	246	3.83 (0.61)	
Overall evaluation[b]	Existing 8	400	6.91 (1.34)	2.359*
	New	246	7.86 (1.20)	

[a] Numbers in parentheses are standard deviations.
[b] The item was ranked on a 10-point bipolar scale anchored by good/bad.
*p values are significant at the 0.001 level.

risk, social risk, and time risk, to measure pleasure tourists' risk perceptions. According to their findings, tourists pay much attention to financial and physical risks. Shopping and optional tours are, without doubt, the main sources of financial and physical risks. In addition, the 1980–2004/ March statistics from the Travel Quality Assurance Association R.O.C. (2004), also reveal that of the 6160 GPT tourists' disputed cases, shopping and optional tours accounted for 34.05%.

In short, this study suggests that travel managers should employ or increase the weight given to shopping and optional tours in GPTs' service quality measurement. Furthermore, from the findings of the focus groups which were conducted in the qualitative stage of this present study, in certain GPT itineraries, such as Southeast Asia and Mainland China, where forced shopping and optional tours are commonly observed, items for measuring shopping and optional tours in GPT are particularly important.

4.2. Dimensions of reduction

The sectors of scenic spots and restaurant were ultimately eliminated in Part Two stage; this unexpected result can be attributed to the following facts:

Relative to other GPT sectors, such as shopping arrangements, optional tours, and transportation, the scenic spots sector has more transparency. For example, various brochures of the scenic spots and the pre-tour briefing are typically provided and conducted by travel operators before the GPT commences. Besides, due to the rapid growth of Internet and public media, GPT tourists now have more channels through which to get the latest information about scenic spots. This situation would probably cause the service perception of scenic spots to differ from other GPT sectors. That is, the detailed information about scenic spots that the GPT tourists could obtain before the GPT begins could possibly reduce the perception of service deviation and failure.

Furthermore, according to the travel laws, any information which is embedded in the travel brochure can be seen as part of the travel contract. In practice, compared to other sectors, the scenic spots are typically introduced in detail in GPT brochure. Hence, after the GPT begins, which and how many scenic spots should be visited are clearly defined by law. A travel operator cannot omit any of the scenic spots arbitrarily; otherwise, tourists could easily report the travel operator to the Travel Quality Assurance Association R.O.C. and claim indemnification in this respect.

In short, compared to other GPT sectors, the scenic spots sector in GPT seems to have the lowest extent of ambiguity and the highest extent of clarity and guarantee. Although the removal of the scenic spots sector may be somewhat unexpected, it seems consistent with what this present study has found in the qualitative stage, that the

problems of the scenic spot sector were the least-mentioned GPT sector (only 7%) in the three focus groups.

With regard to the restaurant sector, typically it is viewed as having a close lineage with the performance of the tour leader and local guide. During the entire GPT itinerary, the tour leader and local guide are the company's representatives, so if something goes wrong, they are expected to show their professional capability to change various components of the tour in order to maintain its quality (Geva & Goldman, 1991, p. 178; Grönroos, 1978). Accordingly, it seems reasonable to infer that if GPT tourists' are dissatisfied with the arrangement of restaurants; they would probably shift the blame on the tour leader and local guide, not on the restaurant itself. This situation may finally reflect the increased relative importance of the tour leader and local guide in the arrangement of restaurants.

The above-mentioned unexpected findings seem to provide a fruitful area for further research. However, in practice, if scenic spots and restaurant sectors still need to be evaluated as part of GPTCCC, this study then suggests that travel managers could employ two open-ended questions such as, "Do you have any suggestions for the arrangement of scenic spots?" and "Do you have any suggestions for the arrangement of restaurants?" on the reverse side of the computer-readable answering card.

4.3. The importance of tour leader/local guide

The significant role of the tour leader/local guide in GPT (Agrusa, 1994; Geva & Goldman, 1991; Mossberg, 1995; Wang et al., 2000) has been once again confirmed by this study.

According to the results of this study, a total of eight items were found to be important for measuring the performance of the tour leader/local guide. They represent 36.4% of the entire GPTCCC. Notably, from the existing 20 CCCs used by the major travel operators, only 29.6% of the items were found to be associated with the tour leader/ local guide. Apparently, more specific items, such as skill of group leading, sense of responsibility, and an interpretive ability still need to be incorporated into practice.

What is most worth noting is that items like friendliness and the ability of coordination between group members are valued by GPT tourists. This result is in consonance with the previous finding that the Taiwanese consumer places importance on friendly behavior; called "Chin-Chieh" in Mandarin Chinese, meaning intimate/cordial/warm politeness (Imire, Cadogan, & McNaughton, 2002). Apparently, it is extremely important for travel mangers to find tour leaders who have the abilities of coordination and friendliness for GPT leading.

4.4. Limitations and directions for future research

The following directions for additional study are mainly stemmed from the limitations of this present study's

investigation plus the desire to know more about the generalization and outcomes of GPTCCC.

In both the qualitative and quantitative stages, the iterative procedure retained only those items that are common and relevant to all GPTs. However, these procedures may have deleted certain important items relevant to some but not all GPTs. Hence, while the findings of this study can be used in its present form to assess and compare service quality of GPTs, appropriate adaptation of the scale may be desirable when investigating some special interest tours, such as in-company GPT or incentive-based GPT.

This research provides the critical items for measuring service quality in GPT. Another avenue to extend this research would be to explore the relation between GPT service quality perceptions and certain variables (e.g. socioeconomic, customer satisfaction, and customer complaining factors). In addition, future research on these issues could compare several cultural units such as Japan, Korea, and mainland China where the GPT is also popular. Such a study approach could help determine the differences in the content of its evaluation items of the service quality of GPT under different cultural contexts in various countries and the cause for these differences. When more cultures are involved in further research, a more rigorous scale of GPTCCC could be developed.

As Olsen and Connolly (2000) once noted, hospitality firms will shift away from competing via traditional methods such as pricing, location, and amenities towards a knowledge-based mode of competition. The most successful competitors will collect and synthesize information about their customers' buying patterns and convert this knowledge into a highly personal level of service. Accordingly, this study believes that development of a scientific tool which is convenient to travel managers is an important foundation for developing knowledge-based competitiveness. The results of this study, both its academic findings and its suggestions for practical designing, reflected this concept completely.

In conclusion, the problems in the GPT faced by the Taiwanese travel managers are unlikely to be unique. The authors believe the problems that have been examined in this study are also common to the GPT operators elsewhere around the world. As Wang and Sheldon (1995) have pointed out, China represents the largest population in the world and is currently experiencing stunning growth in outbound travel as travelers are now discovering the rest of the world. Taiwan and China are similar in race, culture, and language and it is reasonable to believe that China will become the largest outbound GPT export country in the world (Wang et al., 2004). Certainly, it would be worthwhile for destination countries to pay closer attention to this situation, and the findings and ideas of this rigorous study could be generalized to this travel market.

Appendix A

A.1. Ten travel operators for conducting the in-depth interviews

1. Phoenix Tours International, Co., Ltd.
2. Zion International Co., Ltd.
3. Hsi Hung Travel Service Co., Ltd.
4. China Times Travel Service Co., Ltd.
5. Lion Travel Service Co., Ltd.
6. Perfect Travel Agency, Ltd.
7. Four Seasons Travel Service Co., Ltd.
8. Everlight Travel Service Co., Ltd.
9. Fuller Express Corporation
10. Ol'e Travel Service co., Ltd.

A.2. Twenty CCCs collected from Taiwan's major travel operators

1. China Times Travel Service Co., Ltd.
2. Sin Bourn Trave Service Co., Ltd.
3. Perfect Travel Agency, Ltd.
4. Comfort Travel Service Co., Ltd.
5. Lion Travel Service Co., Ltd
6. Toprank Travel Service Co., Ltd.
7. Happy Formosa Travel Service Co., Ltd.
8. Four Seasons Travel Service Co., Ltd.
9. Tien An Travel Service Co., Ltd.
10. Zion International Co., Ltd.
11. Tonan Travel Service Co., Ltd.
12. Wan Sheng Travel Service.
13. All Pass Travel Service Co., Ltd.
14. Regent Travel Service Co., Ltd.
15. Tourex Travel Service.
16. Dadova Travel Service Co., Ltd.
17. Green Mount Int'l Travel Agency.
18. Hsi Hung Travel Service Co., Ltd.
19. Skylark Travel Service Co., Ltd.
20. Europa Travel Service.

References

Agrusa, J. (1994). Group tours in Hawaii. *Annals of Tourism Research*, *21*(1), 146–147.

Albrecht, K. (1992). *The only thing that matters*. New York: Harper Business.

Anonymous. (2003). New routes to the beach. *The Economist*, *368*(8335), 55–56.

Bagozzi, R. P., & Baumgartner, H. (1994). The evaluation of structural equation models and hypothesis testing. In P. R. Bagozzi (Ed.), *Principles of marketing research* (pp. 362–422). Cambridge: Blackwell Publishers.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74–94.

Bagozzi, R. P., Yi, Y., & Singh, S. (1991). On the use of structural equation models in experimental designs: Two extensions. *International Journal of Research in Marketing*, *8*(2), 125–141.

Baker, D. A., & Crompton, J. L. (2000). Quality, satisfaction and behavioral intentions. *Annals of Tourism Research*, *27*(3), 785–804.

Barsky, J. D., & Huxley, S. J. (1992). A customer-survey tool: Using the "quality sample". *The Cornell Hotel and Restaurant Administration Quarterly*, 33(6), 18–26.

Bigné, J. E., Martínez, C., Miquel, M. J., & Andreu, L. (2003). SERVQUAL reliability and validity in travel agencies. *Annals of Tourism Research*, 30(1), 258–262.

Bojanic, D. C., & Rosen, D. L. (1994). Measuring service quality in restaurants: An application of the SERVQUAL instrument. *Hospitality Research Journal*, 18(1), 3–14.

Boulding, W., Kalra, A., Staelin, R., & Zeithaml, V. A. (1993). A dynamic process model of service quality: From expectations to behavioral intentions. *Journal of Marketing Research*, 30(1), 7–27.

Bowen, D. (2001). Research on tourist satisfaction and dissatisfaction: Overcoming the limitations of a positivist and quantitative approach. *Journal of Vacation Marketing*, 7(1), 31–40.

Brady, M. K., Cronin, J. J., & Brand, R. R. (2002). Performance-only measurement of service quality: A replication and extension. *Journal of Business Research*, 55(1), 17–31.

Brown, T. J., Churchill, G. A., & Peter, J. P. (1993). Improving the measurement of service quality. *Journal of Retailing*, 69(1), 127–139.

Carman, J. M. (1990). Consumer perception of service quality: An assessment of the SERVQUAL dimensions. *Journal of Retailing*, 66(1), 43–45.

Cawley, E. (1998). The good, the bad and the ugly: How public comment cards can improve operational effectiveness and public perception. *Assessment Journal*, 5(1), 54–58.

Chadee, D., & Mattsson, J. (1996). An empirical assessment of customer satisfaction in tourism. *The Service Industry Journal*, 16(3), 305–320.

China National Tourism Administration. (2005). *The year book in China tourism in 2004*. Beijing, China: China National Tourism Administration.

Churchill, G. A., Jr. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16(1), 64–73.

Cook, C., & Thompson, B. (2000). Reliability and validity of SERVQUAL scores used to evaluate perception of library service quality. *Journal of Academic Librarianship*, 26(4), 248–258.

Cooper, D. R., & Schindler, P. S. (2003). *Business research methods* (8th ed.). New York: McGraw-Hill.

Crompton, J. L., & Mackay, K. J. (1989). Users' perceptions of the relative importance of service quality dimensions in selected public recreation programs. *Leisure Science*, 11(4), 367–375.

Cronin, J. J., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing*, 56(3), 55–68.

Cronin, J. J., & Taylor, S. A. (1994). SERVPERF versus SERVQUAL: Reconciling performance based and perception-minus-expectations measurement of service quality. *Journal of Marketing*, 58(1), 1 25–131.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive*. New York: Wiley.

Dabholkar, P. D., Thorpe, D. I., & Rentz, J. O. (1996). A measure of service quality for retail stores: Scale development and validation. *Journal of the Academy of Marketing Science*, 24(1), 3–16.

Davis, D., & Cosenza, R. M. (1993). *Business research for decision making*. Belmont, CA: Wadsworth Publishing.

Dev, C. S., & Olsen, M. D. (2000). Marketing challenges for the next decade. *Cornell Hotel and Restaurant Administration Quarterly*, 41(1), 41–47.

Devellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage Publisher.

Evenson, R. (2001). What is the best way to measure customer service? *The American Salesman*, 46(2), 14–16.

Fick, G. R., & Ritchie, J. R. (1991). Measuring service quality in the travel and tourism industry. *Journal of Travel Research*, 30(1), 2–9.

Fodness, D. (1994). Measuring tourist motivation. *Annals of Tourism Research*, 21(3), 555–581.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50.

Freiden, J. B. (1984). Advertising spokesperson effects: An examination of endorser type and gender on two audiences. *Journal of Advertising Research*, 24(5), 33–41.

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marking Research*, 25(2), 186–192.

Geva, A., & Goldman, A. (1991). Satisfaction measurement in guided tours. *Annals of Tourism Research*, 18(2), 177–185.

Geyskens, I., & Steenkamp, J. E. M. (2000). Economic and social satisfaction: Measurement and relevance to marketing channel relationships. *Journal of Retailing*, 73(1), 11–32.

Goldsmith, R. E. (1987). Two studies of Yeasaying. *Psychological Reports*, 60(1), 239–244.

Grönroos, C. (1978). A service oriented approach to marketing of services. *European Journal of Marketing*, 12(3), 588–601.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1991). *Multivariate data analysis with readings* (3rd ed.). New York: Macmillan.

Homburg, C., & Rudolph, B. (2001). Customer satisfaction in industrial markets: Dimensional and multiple role issues. *Journal of Business Research*, 52(1), 15–33.

Hudson, S., Hudson, P., & Miller, G. A. (2004). The measurement of service quality in the tour operating sector: A methodological comparison. *Journal of Travel Research*, 42(3), 305–312.

Imire, B. C., Cadogan, J. W., & McNaughton, R. (2002). The service quality construct on a global stage. *Managing Service Quality*, 12(1), 10–18.

Iverson, T. J. (1997). Decision time: A comparison of Korean and Japanese travelers. *International Journal of Hospitality Management*, 16(2), 209–219.

Johnson, L. L., Dotson, M. J., & Dunlop, B. J. (1988). Service quality determinants and effectiveness in the real estate brokerage industry. *The Journal of Real Estate Research*, 3(2), 21–36.

Jöreskog, K. C., & Sörbom, D. (1993). *LISREL 8 user's reference guide*. Chicago: Scientific Software.

Kassarjian, H. H. (1977). Content analysis in consumer research. *Journal of Consumer Research*, 4(1), 8–18.

Lafferty, B. A., & Goldsmith, R. E. (1999). Corporate credibility's role in consumers' attitudes and purchase intentions when a high versus a low credibility endorser is used in the ad. *Journal of Business Research*, 44(2), 109–116.

Lam, T., & Zhang, H. Q. (1999). Service quality of travel agents: The case of travel agents in Hong Kong. *Tourism Management*, 20(3), 341–349.

LeBlanc, G. (1992). Factors affecting customer evaluation of service quality in travel agencies: An investigation of customer perceptions. *Journal of Travel Research*, 30(4), 10–16.

Lee, Y. L., & Hing, N. (1995). Measuring quality in restaurant operations: An application of the SERVQUAL instrument. *International Journal of Hospitality Management*, 14(3/4), 293–310.

Lewis, R. C., & Pizam, A. (1981). Guest surveys: A missed opportunity. *The Cornell Hotel and Restaurant Administration Quarterly*, 22(3), 37–44.

Llosa, S., Chandon, J. L., & Orsingher, C. (1998). An empirical study of SERVQUAL's dimensionality. *Services Industries Journal*, 18(2), 16–44.

Lundstrom, W. J., & Lamont, L. M. (1976). The development of a scale to measure consumer discontent. *Journal of Marketing Research*, 13(4), 373–381.

March, R. (1997). Diversity in Asian outbound travel industries: A comparison between Indonesia, Thailand, Taiwan, South Korea, and Japan. *International Journal of Hospitality Management*, 16(2), 231–238.

March, R. (2000). The Japanese travel life cycle. *Journal of Travel & Tourism Marketing*, 9(1/2), 185–200.

Marvin, B. (1992). Keeping score on service. *Restaurants and Institutions*, 102(22), 128–142.

Mattsson, J. (1994). Using service process models to improve service quality: Examples from New Zealand. *Managing Service Quality*, 4(1), 47–52.

McCleary, K. W., & Weaver, P. A. (1982). Improving employee service levels through identifying sources of customer satisfaction. *International Journal of Hospitality Management*, *1*(2), 85–89.

Mossberg, L. L. (1995). Tour leaders and their importance in charter tours. *Tourism Management*, *16*(6), 437–445.

Olsen, M. D., & Connolly, D. J. (2000). Experience-based travel: How technology is changing the hospitality industry. *Cornell Hotel and Restaurant Administration Quarterly*, *41*(1), 30–40.

Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991). Refinement and reassessment of the SERVQUAL scale. *Journal of Retailing*, *67*(4), 420–450.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, *49*(4), 41–50.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, *64*(1), 12–40.

Paulhus, D. L. (1984). Social desirable responding: Some new solutions to old problems. In D. M. Buss, & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 201–209). New York: Springer.

Prideaux, B. (1998). Korean outbound tourism: Australia's response. *Journal of Travel & Tourism Marketing*, *7*(1), 93–102.

Quiroga, I. (1990). Characteristics of package tours in Europe. *Annals of Tourism Research*, *17*(2), 185–207.

Reimer, G. D. (1990). Package dreams: Canadian tour operators at work. *Annals of Tourism Research*, *17*(4), 501–512.

Reisinger, Y., & Turner, L. W. (2002). Cultural differences between Asian tourist markets and Australian hosts, Part 1. *Journal of Travel Research*, *40*(3), 295–319.

Roehl, W. S., & Fesenmaier, D. R. (1992). Risk perceptions and pleasure travel: An exploratory analysis. *Journal of Travel Research*, *30*(4), 17–26.

Ross, G. F. (1994). Service quality ideals among hospitality industry employees. *Tourism Management*, *15*(4), 273–280.

Ryan, C. (1995). Learning about tourists from conversations: The over-55s in Majorca. *Tourism Management*, *16*(3), 207–215.

Ryan, C., & Cliff, A. (1997). Do travel agencies measure up to customer expectation? An empirical investigation of travel agencies' service quality as measured by SERVQUAL. *Journal of Travel & Tourism Marketing*, *6*(2), 1–29.

Saleh, F., & Ryan, C. (1991). Analysing service quality in the hospitality industry using the SERVQUAL model. *Service Industries Journal*, *11*(3), 352–373.

Sampson, S. E. (1998). Gathering customer feedback via the internet: Instruments and prospects. *Industrial Management and Data Systems*, *98*(2), 71–82.

Schmidt, C. J. (1979). The guided tour: Insulated adventure. *Urban Life*, *7*(4), 441–576.

Stevens, P., Knutson, B., & Patton, M. (1995). DINESERV: A tool for measuring service quality in restaurants. *Cornell Hotel and Restaurant Administration Quarterly*, *36*(2), 56–60.

Swarbrooke, J., & Horner, S. (1999). *Consumer behaviour in tourism*. Boston, MA: Butterworth Heinemann.

Sweeney, J. C., & Soutar, G. N. (2001). Consumer perceived value: The development of a multiple item scale. *Journal of Retailing*, *77*(2), 203–220.

Taiwan Tourism Bureau. (1993). *Study on the assessment of travel agencies' service quality and incentive system*. Taipei: Tourism Bureau.

Timothy, D. J., & Butler, R. W. (1995). Cross-border shopping—A North American perspective. *Annals of Tourism Research*, *22*(1), 16–34.

Tourism Bureau. (2002). *2001 annual survey reports on R.O.C. outbound travelers*. Taipei: Tourism Bureau.

Tourism Bureau. (2003). *2002 annual survey reports on R.O.C. outbound travelers*. Taipei: Tourism Bureau.

Tourism Bureau. (2004). *2003 annual survey reports on R.O.C. outbound travelers*. Taipei: Tourism Bureau.

Tourism Bureau. (2005). *2004 annual survey reports on R.O.C. outbound travelers*. Taipei: Tourism Bureau.

Travel Quality Assurance Association R.O.C. (2004). *The statistics of Taiwan GPT tourists' disputes cases from 1980-2004 [Online]*. Available: http://www.travel.org.tw/casestudy/ [May 21, 2004].

Tsai, M., Ryan, C., & Lockyer, T. (2002). Culture and evaluation of service quality—A study of the service quality gaps in a Taiwanese setting. In C.-C. Lu, K. Chon, H.-Q. Zhang, & Y.-J. Xie (Eds.), *Proceedings of the 8th annual conference on Asia Pacific tourism association* (pp. 754–763). Dalian, China: Dongbei University of Finance & Economics.

Villanova, P. (1992). A customer-based model for developing job performance criteria. *Human Resource Management Review*, *2*(1), 103–114.

Vogt, C. A., & Fesenmaier, D. R. (1995). Tourists and retailers' perceptions of services. *Annals of Tourism Research*, *22*(4), 763–780.

Wang, K.-C., Hsieh, A.-T., & Chen, W.-Y. (2002). Is the tour leader an effective endorser for group package tour brochures? *Tourism Management*, *23*(5), 489–498.

Wang, K.-C., Hsieh, A.-T., & Huan, T.-C. (2000). Critical service features in group package tour: An exploratory research. *Tourism Management*, *21*(2), 177–189.

Wang, K.-C., Hsieh, A.-T., Yeh, Y.-C., & Tsai, C.-W. (2004). Who is the decision maker: The parents or child in group package tour? *Tourism Management*, *25*(2), 183–194.

Wang, Y., & Sheldon, P. J. (1995). The sleeping dragon awakes: The outbound Chinese travel market. *Journal of Travel & Tourism Marketing*, *4*(4), 41–54.

Weber, R. P. (1990). *Basic content analysis*. Newbury Park, CA: Sage.

Wei, C., & Clark, A. (2002). Exploratory study of travel agency marketing in China. In C.-C. Lu, K. Chon, H.-Q. Zhang, & Y.-J. Xie (Eds.), *Proceedings of the 8th annual conference on Asia Pacific tourism association* (pp. 77–83). Dalian, China: Dongbei University of Finance & Economics.

Weiermair, K., & Fuchs, M. (1999). Measuring tourist judgment on service quality. *Annals of Tourism Research*, *26*(4), 1004–1021.

Wisner, J. D., & Corney, W. J. (1997). An empirical study of customer comment card quality and design characteristics. *International Journal of Contemporary Hospitality Management*, *9*(3), 110–115.

Yamamoto, D., & Gill, A. (1999). Emerging trend in Japanese package tourism. *Journal of Travel Research*, *38*(2), 134–143.

Yuksel, A., & Rimmington, M. (1998). Customer-satisfaction measurement. *Cornell Hotel and Restaurant Administration Quarterly*, *39*(6), 60–71.