

Running Head: RESTAURANT TIPS AND SERVICE

Restaurant Tips and Service Quality:
A Weak Relationship or Just Weak Measurement?

Michael Lynn

School of Hotel Administration

Cornell University

This research was supported by a grant from the Center for Hospitality Research at Cornell University.

Correspondence concerning this article should be addressed to Michael Lynn, School of Hotel Administration, Cornell University, Ithaca, NY 14853-6902. Electronic mail may be sent via Internet to WML3@Cornell.edu.

Abstract

Restaurant patrons were asked to rate their service using either a semantic differential scale or an unbounded write-in scale. The service indices derived from semantic differential and unbounded write-in scales had different distributions, but did not differ in their relationship to tip percentages. The non-significant service by scale-type interaction suggests that the weak service-tipping relationship in the existing literature is not due to response biases associated with semantic differential scales.

Restaurant Tips And Service Quality:
A Weak Relationship or Just Weak Measurement?

Restaurant tips are supposed to be an incentive/reward for service (Lynn & Graves, 1996). However, this supposed function of, and motivation for, tipping has been challenged by empirical research. A recent meta-analysis of 14 studies involving 2,645 dining parties at 21 different restaurants found that, on average, evaluations of service quality accounted for less than 2% of the variability in tips expressed as a percentage of the bill (Lynn, 2001). This weak relationship suggests that equity motivations are weak in the commercial exchanges between servers and their customers and that tipping does not really function as an incentive/reward for service.

Despite the research findings described above, there is still some doubt about the accuracy of the conclusion that service has only weak effects on tipping, because most of the existing research on this topic has relied upon customer's ratings of service using semantic differential scales. Customer satisfaction ratings using semantic differential scales are often negatively skewed (Peterson & Wilson, 1992) and the service ratings in the tipping literature are no exception (see Bodvarsson & Gibson, 1999; Lynn, 2000, 2001). The strong negative skewness produced by these scales may be due to artifacts such as positive response biases or ceiling effects. If so, the validity of the measures would be undermined because people with relatively moderate levels of true satisfaction would be grouped among those with higher levels of true satisfaction. Thus, the use of semantic differential ratings in the existing literature may have artificially weakened the

observed correlation between service quality and restaurant tips. This possibility is tested in the study reported below.

In this study, patrons departing four restaurants were asked to complete a brief questionnaire about their dining experience and tipping behavior. Those patrons agreeing to participate were randomly assigned to use either a semantic-differential scale or an unbounded write-in scale to rate the quality of the restaurants' service, food and atmosphere. The unbounded write-in scale, developed by Eric Marder (1997), produces ratings that are much less negatively skewed than are semantic-differential ratings. If negative skewness related measurement biases are responsible for the weak service-tipping relationship in the existing literature, then service ratings using the unbounded write-in scale should be more strongly related to tip size than are service ratings using the semantic differential scale in this study.

Method

Data Source and Procedure

The data for this study was collected at four independently owned restaurants in Ithaca, New York. Undergraduate research assistants stood outside the restaurants and asked departing patrons to complete a brief paper and pencil questionnaire. Three hundred ninety seven useable questionnaires (with missing data for some variables) were obtained with a response rate of approximately 70%.

Instrument and Manipulation

There were two versions of the questionnaire passed out to the restaurant patrons. Both versions asked the patrons to indicate how much they liked or disliked each of the following aspects of their dining experience – server appearance, server friendliness,

server attentiveness, server promptness, food appearance, food taste, food portion size, food value for the money, dining room lighting, dining room temperature, dining room noise level, and dining room crowd level. Both versions of the questionnaire also asked participants to indicate the sex of their server, the number of adults, teenagers and children in their party, their own sex, their check size, their tip amount, and (at two of the restaurants) how frequently they patronized that restaurant.

The two versions of the questionnaire differed only in the response scale participants were asked to use when rating their dining experiences. One version asked participants to use nine-point semantic differential scales whose endpoints were labeled “Dislike Very Much” and “Like Very Much” respectively. The other version asked participants to use unbounded write-in scales. Using this scale involved writing L’s or D’s or an N to express their feelings about each rated aspect of the dining experience. Participants were instructed to write as many L’s or D’s as they wanted next to each aspect of the dining experience to express how much they liked or disliked it – the more they like (disliked) it, the more L’s (D’s) they should write. The sum of the L’s for each aspect of the dining experience (or negative one times the sum of the D’s) constituted the score for that aspect. If participants neither liked or disliked some aspect of the dining experience, they were instructed to write an N next to it and this was scored as zero. These two versions of the questionnaire were randomly ordered before being passed out to restaurant patrons so that participants were randomly assigned to one of the two scale conditions.

Results

Data Reduction

Separate analyses of the semantic differential and unbounded write-in scales indicated that, for both scales, the four service ratings were highly correlated with one another (all r 's $> .60$) as were the four food (all r 's $> .57$) and room ratings (all r 's $> .33$). Therefore, the four ratings in each set were averaged to form two indices each of service, food, and room quality. For the semantic differential scales, the coefficient alphas for these three indices were .90, .88, and .78 respectively. For the unbounded write-in scales, the coefficient alphas for these indices were .92, .88, and .83 respectively.

Distributions of the Semantic Differential and Unbounded Write-In Scales

The semantic differential ratings produced a service index with a skewness of -1.60 , while the unbounded write-in scale produced a service index with a skewness of $.88$ (see Figures 1 and 2). This supports the assumption underlying the current study design that the unbounded write-in scale ratings would display less negative skewness than the semantic differential scale ratings.

Comparison of the Service-Tipping Relationships Produced by the Two Scales

The service indices based on unbounded write-in scale and semantic differential scale ratings were separately standardized and entered into a single multiple regression predicting percent tip from restaurant (dummy coded), service quality, type-of-scale, and the product of service-quality and type-of-scale. In this analysis, the service-quality by type-of-scale interaction was not significant ($B = -.13$, $t(389) = -0.27$, $p = .79$). This

interaction, as well as a comparison of zero-order correlations, indicates that the standardized service index based on unbounded write-in scale ratings was not more strongly related to percent tip ($r = .16$, $n = 200$, $p < .03$) than the standardized service index based on semantic differential scale ratings ($r = .19$, $n = 196$, $p < .01$). This failure to find a stronger service-tipping relationship when using the unbounded write-in scale as compared to the semantic differential scale is unlikely to be a Type II error. The test of this direction-specific interaction was not significant even with a sample size of 396 and a one-tailed alpha of .50. Under these conditions, even a small interaction effect (of size $r = .10$) in the population would be found 97.6 percent of the time (see UCLA Department of Statistics, 2002). Thus, it is unlikely that this study simply failed to detect an interaction effect of non-trivial size that actually exists in the population.

Other Variables' Relationships to Tipping

A multiple regression predicting percent tip from restaurant (dummy coded), bill size, server's sex, number of adults in the party, number of teens in the party, number of kids in the party, customer's sex, and service, food and room indices (standardized by scale type) produced a significant effect only for the service index ($B = .91$, $t(355) = 2.74$, $p < .007$). The number of teens in the dining party had a marginally significant negative effect ($B = -1.12$, $t(355) = -1.73$, $p = .09$), but none of the other variables was even marginally significantly related to percent tip (all t 's (355) < 1.01 , all p 's $> .10$). A similar regression model that included patronage frequency -- for the two restaurants at which this variable was collected -- indicated that patronage frequency was also unrelated to percent tip ($B = -.02$, $t(103) = -.25$, $p = .80$).

Discussion

This study extended previous research by finding that the service indices derived from unbounded write-in and semantic differential scales, despite having different distributions, did not differ in their relationship to tip percentages. This failure to find a significant service by scale-type interaction suggests that widespread use of negatively skewed, semantic differential scale ratings to test the relationship between service and tipping has not artificially reduced the size of the observed relationship in the existing literature. Thus, the results of this study further support the conclusion that tip percentages really are only weakly related to customers' perceptions of service quality.

Theoretical Importance

Evidence for a weak relationship between service and tipping is theoretically important for two reasons. First, it supports Oliver and Swan's (1989) finding that equity motivations are weak in commercial buyer-seller exchanges. This support is particularly significant because tipping is much closer to a social exchange than are more traditional commercial exchanges. That equity motivations have little impact on even this near-social, commercial exchange suggests that they are unlikely to have a strong effect on more traditional commercial exchanges. Second, a weak service-tipping relationship undermines economists' theories that tipping exists because it is the most efficient way of monitoring and rewarding server effort (Bodvarsson & Gibson, 1994). Tips can serve as an incentive to deliver good service only if they are perceived as being strongly related to service quality. The weak relationship observed in this study, and in the previous tipping literature, raises serious questions about servers' ability to see a connection between their service and the tips they receive (Lynn, 2001).

Practical Importance

Evidence for a weak relationship between service and tipping is also of practical importance. Restaurant managers rely upon tips to: (a) motivate servers to deliver good service, (b) measure server performance, and (c) identify dissatisfied customers (Lynn, 2001). All of these uses of tips assume that they are strongly related to customers' perceptions of service quality. A weak service-tipping relationship means that managers should look for other means of accomplishing these tasks.

References

Bodvarsson, O. & Gibson, W. (1994). Gratuities and customer appraisal of service: Evidence from Minnesota restaurants. Journal of Socio-Economics, *23*, 287-302.

Bodvarsson, O. & Gibson, W. (1999). An economic approach to tips and service quality: Results of a survey. The Social Science Journal, *36*, 137-147.

Lynn, M. (2000). The relationship between tipping and service quality: A comment on Bodvarsson and Gibson's article. The Social Science Journal, *37*, 131-135.

Lynn, M. (2001). Restaurant tipping and service quality: A tenuous relationship. Cornell H.R.A. Quarterly, *42*, 14-20.

Lynn, M. & Graves, J. (1996). Tipping: An incentive/reward for service? Hospitality Research Journal, *20*, 1-14.

Marder, E. (1997). The Laws of Choice: Predicting Consumer Behavior. New York: Free Press.

Oliver, R.L. & Swan, J. E. (1989). Consumer perceptions of interpersonal equity and satisfaction in transactions: A field study approach. Journal of Marketing, *53*, 21-35.

Peterson, R.A. & Wilson, W.R. (1992). Measuring customer satisfaction: Fact and artifact. Journal of the Academy of Marketing Science, *20*, 61-71.

UCLA Department of Statistics. (2002). Power Calculator. [On-line]. Available: <http://ebook.stat.ucla.edu/calculators/powercalc>.

Figure Captions

Figure 1. Distribution of scores for the service index (average of ratings for server's appearance, friendliness, attentiveness and promptness) based on semantic differential scale ratings. On this scale, 1 was labeled "dislike very much" and 9 was labeled "like very much."

Figure 2. Distribution of scores for the service index (average of ratings for server's appearance, friendliness, attentiveness and promptness) based on unbounded write-in scale ratings. On this scale, negative values mean the service was disliked and positive values mean the service was liked, with the magnitude of the value indicating the degree of liking or disliking.

FIGURE 1

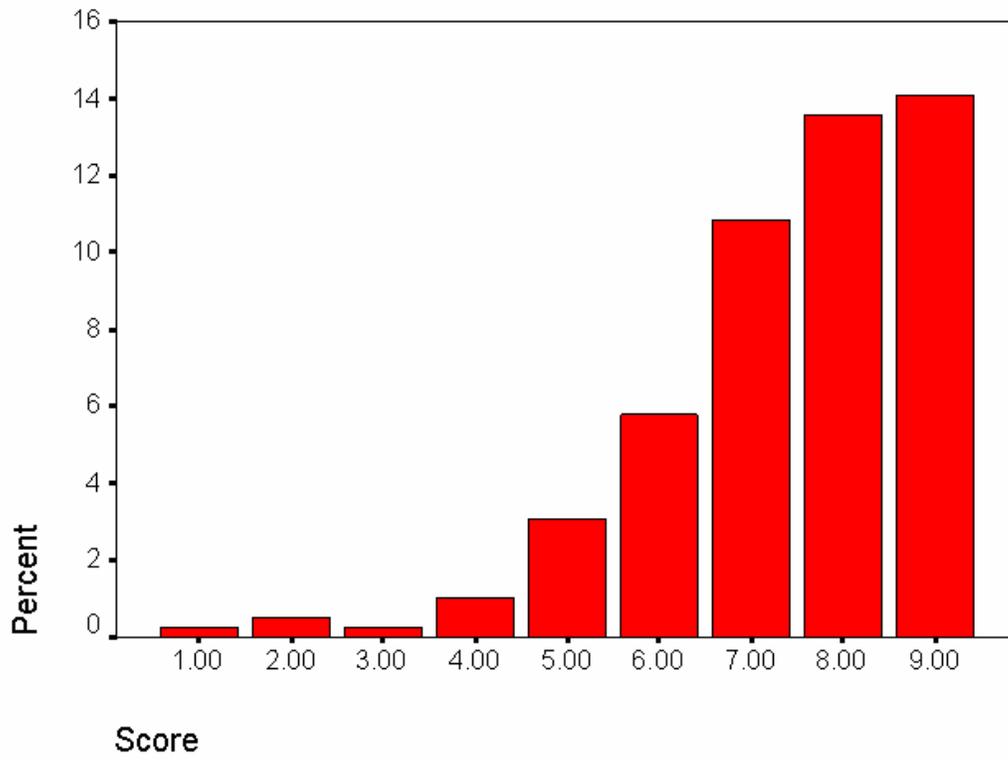


FIGURE 2

