Bjarnadottir, A., & Campbell, J. P. (2001). Development of a model of individual performance in customer service *Ergometrika*, 2, 2-25.

DEVELOPMENT OF A MODEL OF INDIVIDUAL PERFORMANCE IN CUSTOMER SERVICE

ASTA BJARNADOTTIR
Reykjavik University

JOHN P. CAMPBELL
University of Minnesota

ABSTRACT

The study deals with the development of a general model of individual performance in customer service. Principal components analyses and higher-order analyses were conducted based on similarity information for 400 critical service incidents generated by expert judges. A model of 10 performance dimensions was proposed, and model reliability demonstrated via retranslation.

INTRODUCTION

The prediction and control of individual job performance is among the most prominent applied problems in industrial and organizational psychology. An important prerequisite to prediction and control, however, is a good substantive description of the nature of the domain (Austin & Villanova, 1992; Campbell, 1990; Campbell, Dunnette, Lawler, & Weick, 1970). In the case of job performance, substantive description can take the form of a model, specifying the major dimensions of performance requirements for the relevant job or jobs. A "mid-range" model of job performance involves the specification of the substantive content of job performance at a level of generalization that is narrower than a general model of job performance, but broader than a job-specific model (Borman, 1991; Campbell, McCloy, Oppler & Sager, 1993; Fleishman & Quaintance, 1984).

With two thirds of the U.S. labor force engaged in private or public service sector employment, and predictions that this proportion will increase in the future (Daniels, 1993; Riddle, 1986), effective customer service has indeed become a critical component of individual job performance for a large portion of the work force. Increasing parity in technology and cost configurations in recent years, coupled with rising consumer expectations, have forced service organizations to focus more on competition through service quality (Desatnick, 1987). At the same time, however, frequent complaints about low and declining quality in services suggest that implementing a high-quality approach to service is easier said than done (Barnes & Glynn, 1994; Davidow, 1988; Grönroos, 1990; Koepp, 1988; Quinn & Gagnon, 1988). In response to the problem of uneven quality in services, many approaches have been suggested (Bowen & Schneider, 1988; Luthans & Davis, 1990; Schlesinger & Heskett, 1992; Schneider, 1990; Zeithaml, Parasuraman & Berry, 1992). Although diverse in other ways, the majority of these approaches identify the behavior of front-line service providers as a major determinant of service quality and customer satisfaction, a notion that is consistently supported by services marketing research into the determinants of customer satisfaction (Bitner, 1990; Crosby & Stevens, 1987). Therefore, the prediction and control of individual performance in customer service is currently a critical management problem. However, an empirically-based description of the behaviors involved in providing good service has not been developed.

The objective of the current research is to develop a general dimensional model of individual performance in customer service, using appropriate empirical data. A general model of customer service is intended to apply to all jobs that are primarily service jobs, and also to the service-portion of other jobs that might not normally be considered "service jobs." Therefore, the term "customer service role" is used here instead of "customer service job". A customer service role is enacted each time an employee of a service organization is engaged in interaction with, or independent work conducted directly on the behalf of, an internal or external customer or client, during the course of providing a service to said customer or client (Lovelock, 1985).

This type of interaction has been labelled a *service encounter* (Bitner, 1990; Czepiel, Solomon, Surprenant & Gutman, 1985).

A review of prior literature reveals various sources that are relevant to the development of a model of individual performance in customer service roles. Relevant models from the service marketing literature include the Mersha and Adlakha (1992) and Parasuraman, Zeithaml and Berry (1985, 1988) models of general service quality, the Bitner, Booms and Tetreault (1990) model of service behaviors, and the Saxe and Weitz (1982) model of customer oriented selling. Relevant models from the applied psychology literature include the Hogan, Lock and Brinkmeyer (1995) model of interpersonal behaviors at work, models from applied training work (Target stores, 1996; Wilcock, 1989), criterion models from selection validation research (Motowidlo & Carter, 1990; Paajanen & McLellan, 1993), and models from behavior modification studies (Crowell, Anderson, Abel & Sergio, 1988; Brown, Malott, Dillon & Keeps, 1980). This literature was used to develop a preliminary *a priori* model of individual performance in customer service, which is used at later stages in the study (see table 2).

Several types of empirical data can be used for job performance modeling. However, to avoid problems associated with job performance modeling based on data collected with structured instruments (Campbell, 1983; Landy & Farr, 1983; Murphy, 1989), it seems necessary to base an exploratory study such as this one primarily on naturalistic performance data, or data collected without the assistance of a pre-selected vocabulary. Critical incidents (Flanagan, 1954) were selected as the primary unit of analysis for this study. Critical incidents are naturalistic job performance data that have been widely used to develop models of job performance, and that have also been popular in service marketing research (e.g., Bitner et al., 1990; Bitner, Booms & Mohr, 1994; Hayes, 1992; Johnston, 1994).

The first research question of this study is wether the exploratory analysis of critical incidents from various customer service settings can support the development of a dimensional model with distinguishable performance components. It is expected that distinguishable, albeit correlated, components can be found in statistical analyses based on similarity data derived from critical incidents.

The second research question concerns the reliability of such a model, as an important aspect of the construct validity of a dimensional model. Based on previous research (Campbell, 1987; Hogan et al., 1995), it is expected that reliability, defined as interjudge agreement, can be shown to equal or exceed 60% on average (Fleiss, 1971).

METHOD

Participants and data collection

Critical incidents of service encounters were collected from 105 general consumers, who were approached in an introductory psychology class at a large university. Additional incidents were collected from 44 service job incumbents in a large retail organization. Incidents were collected in workshop settings using protocols from Flanagan's original (1954) procedure (cf. Bownas & Bernardin, 1988). Of the 105 consumer participants, 41% were shown the a priori model derived from earlier literature *before* they wrote any incidents, and asked to use it to aid their recall of the relevant incidents. Incidents written by this group are henceforth identified as the *with-model* incidents, and incidents collected from the remaining group are labelled *no-model* incidents.

Incidents collected were transcribed into a common format, to create a common flow of events and to remove redundant information. Incidents that originally described multiple acts by the employee in question, were split into two or more incidents.

Phase 1: Model development via principal components analyses

In order to conduct principal components analysis, similarity data on the critical incidents had to be generated. This was done using expert classification.

Sorting procedure for generating incident similarity data

Two samples of 200 incidents each were randomly chosen from the set of 490 *no-model* incidents (Sample 1) and the set of 453 *with-model* incidents (Sample 2), respectively. Twenty-one individuals, 11 women and 10 men, all with at least two years of academic work in I/O psychology, participated in the sorting task. They were unfamiliar with the a priori model. Judges were paid a nominal fee of $50 for their time. Ten judges classified the 200 *no-model* incidents, and ten judges classified the *with-model* incidents. Judges were asked to independently classify the incidents into homogeneous groups, using as many dimensions of their own creation as

needed (Bownas & Bernardin, 1988; Campbell, Dunnette, Arvey & Hellervik, 1973).

Principal components analysis of incident similarity data

Similarity data on the critical incidents was created as follows. First, a 200*200 matrix of "raw" incident similarity values was constructed based on the proportion of judges who classified the two incidents together into any one of their groups (Miller, 1969). Second, a *standardized mean inner product* (SMIP) was computed using the similarity profiles of all possible pairs of incidents, to create an indirect incident similarity matrix. The SMIP index has an absolute zero point and a range of 0.0 to +1.0, and can be interpreted as the degree to which each incidents' profile of co-occurrence with other incidents is similar to the profile for all other incidents in the sorting results. Two 200*200 matrices were created, one for the *no-model* sample of incidents and one for the *with-model* sample.

The 200*200 matrices were then submitted to principal components analysis with varimax rotation (Davison, 1992; Borman & Brush, 1993; Borman, Dunnette, & Hough, 1976). Fifteen initial factors were first extracted from each matrix, using diagonals as initial communality estimates. Following this, several subsets of factors from each set of initial factors were rotated according to the varimax criterion. Subsets were considered for interpretation as long as the last factor had an eigenvalue greater than one and explained more than 1% of the common variance. Solutions were selected for interpretation based on these criteria plus substantive interpretability.

Oblique primary factoring and higher-order factoring

To investigate higher-order relationships in the data, two oblique primary factor structures were derived, one for each sample of incidents, using a procedure described by Overall and Klett (1972). The last three components in the *with-model* solution had to be dropped before the primary factoring, as they did not have enough "highest loadings" associated with them. The similarity among the oblique primary factors was obtained via the calculation of factor cosines, which range from 0.0 to +1.0. A higher-order principal components analysis was then conducted, based on the two oblique primary factor cosine matrices. This analysis was conducted in two parts, separately for *no-model* and *with-model* data. Higher-order components were chosen for interpretation if they had eigenvalues higher than one.

At the conclusion of phase 1, a new dimensional model was developed, based primarily on the two principal components analysis solutions, but with reference to the a priori model.

Phase 2: Analyses of model reliability

The second phase of the study was concerned with the reliability of the revised model. The analysis was based on three new samples of critical incidents.

Incident sampling and retranslation procedure

Three samples of incidents, counting a total of 875 incidents, were retranslated. Sample 3 contained 200 *no-model* critical incidents, which had not been used in phase 1. Sample 4 contained the remaining 442 *with-model* and *no-model* incidents that occurred in the four most commonly represented settings seen in the study; retailing, restaurants, health care and hair salons. Sample 5 contained 233 incidents written by service job incumbents. Table 1 provides an overview of the five samples of incidents used in this study.

Three expert judges, who had not worked on the previous sorting task, carried out the retranslation task. Judges met the same criteria as before, although non-familiarity with the a priori model was not required. Each incident in each of the retranslation samples was independently retranslated (i.e., classified into dimensions of the revised model) by all 3 judges, using the set of behavioral dimensions identified in the revised model and their definitions as guidelines.

Table 1. Overview of the samples of critical incidents used in the study.

Sample	Number of Incidents	Type of incidents	Number of judges	Task presented to judges
Model development samples				
1	200	No-model consumer incidents	10	Dimension identification via sorting task
2	200	With-model consumer incidents	11	Dimension identification via sorting task
Retranslation samples				
3	*200*	*No-model consumer incidents not in 1 or 2*	*3*	*Retranslation, with final model*
4	*442*	*Consumer incidents, four most common settings*	*3*	*Retranslation, with final model*
5	*233*	*Incumbent incidents, retail setting only*	*3*	*Retranslation, with final model*

Overall model reliability

The overall reliability of the dimensional model, operationalized as the level of interjudge agreement in incident re-classification, was estimated in two ways, average interjudge agreement and expected probability of agreement corrected for chance.

Average interjudge agreement in retranslation was defined as the average percentage of judges who agreed in their classification, across the entire set of incidents used. In the present study, the input for this index is limited to the values of 0.0%, 66.7% and 100.0%. The cutoff point, at which any given incident can be deemed to be "reliably classified" was set at 66.7%, or two-thirds, following prior research (Bownas & Bernardin, 1988; Campbell, 1987; Fogli, Hulin, & Blood, 1971; Landy, Farr, Saal & Freytag, 1976).

Interjudge agreement was also estimated using a generalized version of Cohen's kappa statistic, with correction for chance agreement (Cohen, 1960; Fleiss, 1971). Kappa shows the probability of agreement to be expected if a randomly selected stimulus, in this case an incident, were to be selected from the set and classified independently by two randomly selected judges (Fleiss, 1971). The correction for chance agreement was based on the marginal probabilities associated with the various categories.

Dimension reliability

Model reliability can also be conceptualized at the level of categories used, or in this case, at the level of dimensions. This can identify model dimensions that could benefit from improved clarity in definition. The following analyses are based on samples 3, 4 and 5 combined.

Average interjudge agreement in incident retranslation was calculated separately for each dimension of the model. This analysis necessarily excludes all incidents not "reliably classified" into any one dimension, resulting in the restricted range of 66.7%-100% for this comparison. This index shows the relative reliabilities of different dimensions, or their tendency to contain incidents that are unanimously classified.

The distinctiveness of the different dimensions was also assessed with an "overlap" index, which measures the number of common attributes (incidents) for a pair of dimensions, relative to the total number of attributes associated with the pair combined. A high value indicates confusion by the judges as to the meaning of two dimensions. This index was derived for all possible pairs of dimensions, using the following formula: *Incident* $overlap_{lk} = n_c/(n_c + n_n)$ (Fleishman & Quaintance, 1984).

RESULTS

Results of phase 1: Model development

Incidents collected and sorting task results

The 105 consumer volunteers wrote a total of 743 incidents, or 7.1 on average per person. Service job incumbents wrote 246 incidents, or 5.6 on average. The incidents collected from consumers occurred in various service settings, but the most common ones were retailing (21%), hotels and restaurants (22%), health care (10%), and personal services (9%). Fifty-three incidents had to be dropped because they did not conform to the required content or format. The splitting up of incidents that originally described multiple acts by the service person in question resulted in a 35% increase in overall number of incidents in the consumer sample (from 699 to 943), and a 21.5% increase in the incumbent sample (from 237 to 288). Examples of edited incidents can be seen in table 8.

The average numbers of dimensions generated by the 21 judges who carried out the initial classification task on samples 1 and 2 were 11.5 (*no-model* sample), and 12.3 dimensions (*with-model* sample) respectively. The overall number of dimensions proposed by the judges was 250. Only seven new dimensions appeared after the point at which judges had classified 75% of the incidents.

Principal components analyses

For the *no-model* incidents, the most differentiated rotated solution that met all three conditions of eigenvalues greater than one, variance explained greater than 1%, and substantive interpretability, was a solution with nine components. This solution accounted for 89% of the total variance in the matrix, with the first five components accounting for 71% of the total variance, or 80% of the common variance. For the *with-model* incidents, a solution of 12 components was the most differentiated solution that met the three conditions set, although the last three components were fairly small. This solution accounted for 93.9% of the total variance in the matrix, with the first six components accounting for 82% of the total variance, or 87% of the common variance.

The components identified in these two solutions are shown in columns 2 and 3 of table 2, along with the percentage of total common variance accounted for. The ordering of components in table 2 is based on the a priori model. Table 2 reveals highly similar structures resulting from the two principal components analyses. The principal components accounting for the largest amounts of variance (over 10% each in both samples), were those defined by the following types of content: (1) Friendly behavior and courtesy; (2) Responsiveness and follow-through; (3) Behaviors that involve doing more than the minimum, (4) Providing information; and finally; (5) Redoing service or compensating for service failures.

As Table 2 shows, all components identified in the two principal components solutions interpreted can be mapped onto the a priori solution, although a one-to-one correspondence is only observed for five of the a priori components.

Table 2. Dimensions derived from two principal components analyses of incident similarity data, organized by dimensions of the a priori model.

A priori dimensions	**Components based on no-model incidents** [a]	**Components based on with-model incidents** [a]
(1) Responsiveness and attentiveness toward customers	Follow-through (15%): Following up on service already in progress, completing the transaction, and giving customers one's full attention as long as necessary. Keeping promises. Response (7%): Responding quickly to new customers, not making customers wait.	Response and follow-through (14%): Responding quickly to new customers, following up on service already in progress, completing the transaction, and giving customers one's full attention as long as necessary. Timely (1%): Completing service in the time-frame promised.
(2) Socially engaging customers and being polite	Friendliness and courtesy (20%): Being friendly towards customers, showing an interest in customers as individuals and talking with them on a personal level. Demonstrating courtesy and respect for customers, avoiding negative facial expressions and tones of voice, as well as inconsiderate or blunt remarks and responses. Respect (5%): Treating all customers with respect and courtesy, and refraining from demeaning treatment and/or discrimination.	Friendly (16%): Being friendly towards customers, showing an interest in customers as individuals and talking with them on a personal level. Courtesy (15%): Demonstrating courtesy and respect for customers, avoiding negative facial expressions and tones of voice, as well as inconsiderate or blunt remarks and responses.
(3) Creating trust between organization and customers	Extra mile (12%): Demonstrating a willingness to go beyond basic job requirements to solve customers' problems. Being inventive, flexible with rules, and generous about sharing organizational resources and spending one's own time to help customers.	Extra mile (18%): Demonstrating a willingness to go beyond basic job requirements to solve customers' problems. Being inventive, flexible with rules, and generous about sharing organizational resources and spending one's own time to help customers. Empathy (4%): Displaying genuine concern and sympathy towards customers

		in distress. Customer-oriented selling (1%): Putting customer interests ahead of making a sale.
(4) Receiving, eliciting and using customer input	Listening (6%): Listening to customers and asking questions to diagnose customer needs. Using the customer's input, discussing their options and honoring their wishes, including wishes to be left alone.	Listening (3%): Listening to customers and asking questions to diagnose customer needs. Using the customer's input, discussing their options and honoring their wishes, including wishes to be left alone.
(5) Communicating information to meet customer needs	Information (16%): Providing customers with full and accurate information, explaining service-related issues, clarifying expectations and answering customers' questions. Giving referrals when unable to complete service.	Information (11%): Providing customers with full and accurate information, explaining service-related issues, clarifying expectations and answering customers' questions. Giving referrals when unable to complete service.
(6) Accurate and reliable processing of routine service transactions	Processing (3%): Performing core job tasks and processing routine service-related transactions without error.	Processing (2%): Performing core job tasks and processing routine service-related transactions without error.
(7) Managing and preventing conflict and customer dissatisfaction	Redress (17%): Accepting responsibility for problems, regardless of who is at fault. Re-doing or altering service as needed, apologizing and compensating customers for service failures.	Redress (13%): Accepting responsibility for problems, regardless of who is at fault. Re-doing or altering service as needed, apologizing and compensating customers for service failures.
(8) Appearance and presentation		Appearance (1%): Displaying cleanliness and job-appropriate appearance.

a Note. *Percentage of total common variance accounted for by each component is shown in parentheses*

Oblique primary factors and higher-order factors

The two matrices of cosines for the oblique primary factors derived from *no-model* incidents and *with-model* incidents, respectively, are shown in tables 3 and 4. The cosines can be interpreted as depicting average or aggregate similarities of the incidents defining each pair of factors, on a scale from 0.0 to 1.0.

*Table 3. Similarity matrix for the oblique primary factors identified in the **no-model** incident sample (sample 1).* [a]

	Friendly & courtesy	**Re-dress**	**Infor-mation**	**Follow-through**	**Extra mile**	**Re-sponse**	**Liste-ning**	**Re-spect**
Redress	.30							
Information	.38	.16						
Follow-through	.44	.34	.50					
Extra mile	.37	.39	.40	.49				
Response	.27	.23	.20	.63	.19			
Listening	.44	.22	.45	.56	.23	.27		
Respect	.69	.43	.32	.48	.39	.38	.42	
Processing	.41	.33	.56	.88	.53	.52	.61	.50

[a] *Note. Entries in the table are on the scale of 0.0 to +1.0.*

*Table 4. Similarity matrix for oblique primary factors identified in the **with-model** incident sample (sample 2).* [a]

	Extra mile	**Friendly**	**Court-esy**	**Response & follow-through**	**Re-dress**	**Infor-mation**	**Emp-athy**	**Liste-ning**
Friendly	.17							
Courtesy	.31	.59						
Response & follow-through	.71	.17	.40					
Redress	.52	.16	.46	.42				
Information	.58	.19	.30	.57	.25			
Empathy	.55	.44	.41	.43	.50	.35		
Listening	.34	.18	.24	.38	.30	.37	.41	
Processing	.42	.10	.27	.45	.33	.45	.37	.42

[a] *Note. Entries in the table are on the scale of 0.0 to +1.0.*

The results of higher-order principal components analyses based on the similarity matrices for oblique primary factors are shown in tables 5 and 6.

For the *no-model* data, the first higher-order component identified was defined by strong loadings of the factors *Follow-through, Processing, Listening, Information* and *Response*. This factor was labeled *Routine task behaviors.* The second higher-order component was defined by high loadings by the factors *Respect*, *Redress*, *Friendliness & courtesy* and a lower loading by *Extra mile*. The latter factor represents service behavior that tends to be less routine, and was labeled *Discretionary behaviors*. These two higher-order components explain 61% of the total variance in the *no-model* primary factor similarity matrix.

In the *with-model* set of primary factors, two higher-order components were also identified. The first component was interpreted as a task-oriented cluster, defined by high loadings of the factors *Extra mile*, *Response & follow-through, Information*, and *Processing* This component was labeled *Task behaviors.* The second component was interpreted as a social interaction cluster, defined by the primary factors labeled *Friendliness*, *Courtesy* and *Empathy* and was labeled *Social behaviors*. These two higher order components explain 60% of the variance in the with-model primary factor similarity matrix.

*Table 5. Higher-order components identified in oblique primary factors derived from **no-model** incident similarity data (sample 1).* [a]

Oblique primary factor	**Higher-order components**	
	Routine task behaviors	**Discretionary behavior**
Follow-through	.87	.30
Processing	.87	.31
Listening	.68	.24
Information	.65	.21
Response	.65	.13
Courtesy	.33	.78
Redress	.06	.76
Friendly and courtesy	.31	.72
Extra mile	.35	.58

a Note. *Entries in the table are factor loadings.*

Table 6. Higher-order components identified in oblique primary factors derived from with-model

Oblique primary factor	Higher-order components	
	Task behaviors	**Social behavior**
Extra mile	.82	.19
Response & follow-through	.80	.18
Information	.74	.10
Processing	.71	.06
Listening	.59	.18
Redress	.53	.40
Friendly	.01	.89
Courtesy	.25	.82
Empathy	.53	.57

a Note. *Entries in the table are factor loadings.*

The Final Model

The final model proposed contains 10 dimensions. The model conforms closely to the principal components analysis solutions, and seven of eight dimensions in the a priori model reappear in some form as well. Table 7 shows the final model of individual performance in customer service roles, including labels and definitions. Dimension definitions are written in positive terms, with the exception of the *Courtesy* dimension, which had been defined almost entirely by negative incidents. Table 8 contains examples of incidents that loaded highly on the principal component contributing most directly to each of the 10 final model dimensions.

Table 7. A final model of individual performance in customer service roles.

(1) Response

Responding quickly to new customers as they enter facility or establish contact in other ways. Acknowledging customers promptly and offering help before being asked.

(2) Follow-through

Giving undivided and sustained attention to customer needs for as long as it takes to complete service. Following up on existing customers and service already in progress, and working to complete the service transaction, also when customer is not present.

(3) Friendly

Being friendly towards customers, showing an interest in them as individuals and talking with them on a personal level.

(4) Courtesy

Demonstrating common courtesy and respect for all customers by avoiding negative facial expressions and tones of voice, inconsiderate or blunt remarks and responses, and by refraining from discrimination.

(5) Extra mile

Offering more than expected and demonstrating a willingness to go beyond basic job requirements to solve customers' problems. Being inventive, flexible with rules, and generous about sharing organizational resources and spending one's own time to help customers.

(6) Empathy

Displaying genuine concern and sincere empathy towards customers in distress and/or customers whose problems cannot be easily remedied.

(7) Listening

Listening to customers and asking questions to accurately diagnose their needs. Using customer input and engaging customers in a collaborative discussion about their options and the service to be done. Honoring customers' wishes, including wishes to be left alone.

(8) Information

Providing customers with full and accurate information, tailored to their characteristics and needs. Explaining service-related facts, clarifying expectations, providing guidance and answering customers' questions. Giving honest opinions and suggestions, and avoiding deception. Offering referrals when unable to complete service.

(9) Processing

Performing core job tasks and processing routine service-related transactions thoroughly and without error. Doing things right the first time.

(10) Redress

Accepting responsibility for service-related problems, regardless of who is at fault. Re-doing or altering service when appropriate, and readily apologizing and compensating customers for service failures.

Table 8. Examples of incidents loading highly on each dimension in the final model. [a]

(1) Response

"I came in for a hair cut. She didn't make me wait too long before she got started on my hair." (Effective, .91)

(2) Follow-through

"We took our computer in to get more memory installed and because it wasn't working right. They told us it would be done on a certain day. We got there and it wasn't." (Ineffective, .66)

(3) Friendly

"I was at my dentist's office. They made me feel appreciated by asking me about myself, like how school is going or how my family is doing." (Effective, .95)

(4) Courtesy

"As she handed me my bag I wasn't sure if my receipt was in the bag. I asked if she had put the receipt in my bag. She said very rudely (in a tone of voice like I was dumb): "Yeah . . I DID!"." (Ineffective, .92)

(5) Extra mile

"A friend of mine was doing poorly in this class. The teacher agreed to meet with my friend before school every day for the week prior to the mid-semester exam." (Effective, .95)

(6) Empathy

"I had gotten in a car accident and contacted my insurance agent. I didn't know what to do in the situation. This person was very concerned about me. She asked if everything was OK and how the car accident had happened." (Effective, .84)

(7) Listening

"I asked for a trim. I asked for a little layering to bring out some curl and she cut way too much. She just assumed she knew what I wanted and I thought she did." (Ineffective, .76)

(8) Information

"I came in to see what I needed to do to get retainers for my teeth. He was very clear and concise, giving me info and relating to me his suggestions about what I needed to do." (Effective, .98)

(9) Processing

"I went to the bank to deposit my paychecks and then asked the teller to please give me a balance. She read my balance out loud, although from my understanding, you are supposed to write a balance." (Ineffective, .73)

(10) Redress

"I ordered a meal that I did not find appetizing (burnt steak, cold fries). The waiter brought me a new platter cooked just right and did not charge us for our drinks and appetizers." (Effective, .99)

[a] Note. *Evaluation of the behavior depicted and its loading on the relevant principal component are shown in parentheses.*

Results of phase 2: Model reliability

The goal of this part of the study was to assess the reliability of the dimensional system, via various indices of interjudge agreement.

Overall model reliability

Table 9 shows the average interjudge agreement in retranslation, separately for the three retranslation samples, as well as overall. The overall average interjudge agreement in retranslation is 77%, with a 95% confidence interval ranging from 75% to 79%. Given the nature of the data input, this can be considered a fairly high level of agreement

(Campbell, 1987; Hogan et al., 1995). The results confirm the expectation that average interjudge agreement would equal or exceed 60% on average.

Table 9. Overall model reliability: Average interjudge agreement in retranslation.

Retranslation sample	**Average interjudge**	**95% confidence interval**	
	agreement (%)	**Lower**	**Upper**
Sample 3: 200 consumer incidents, all settings	78.8	74.9	82.8
Sample 4: 442 consumer incidents, four settings	80.3	77.8	82.8
Sample 5: 233 incumbent incidents, retail only	69.8	65.3	74.3
Total	*77.2*	*75.2*	*79.2*

Table 10 shows the point estimate for generalized kappa corrected for chance, for the three retranslation samples as well as overall. For the overall set of 875 incidents this index equals .59, with a 95% confidence interval ranging from .58 to .60. In other words, the expected probability of agreement, if two randomly chosen judges were to independently classify a randomly chosen stimulus from the original set, is about 59%.

Table 10. Overall model reliability: Expected probability of agreement between two judges, corrected for chance.

Retranslation sample	**Kappa**	**95% confidence interval**	
	agreement (%)	**Lower**	**Upper**
Sample 3: 200 consumer incidents, all settings	.61	.58	.64
Sample 4: 442 consumer incidents, four settings	.63	.61	.65
Sample 5: 233 incumbent incidents, retail only	.50	.47	.53
Total	*.59*	*.58*	*.60*

Dimension reliability

The analysis of dimension reliability is intended to suggest which dimensions of the revised model are clearly defined and distinctive in content, and which dimensions would most likely benefit from a second revision.

Table 11 shows the average interjudge agreement, contrasted across the 10 dimensions of the model, with 95% confidence intervals. The values in table 11 are restricted in range, as all incidents that were not reliably classified (66.7% or 100% agreement) to a dimension had to be excluded. Table 11 suggests that the dimensions *Response, Friendliness, Information*, and *Redress* are the most reliable ones, but that the dimensions *Follow-through, Extra mile* and *Empathy* may be less reliable.

Table 11. Average interjudge agreement in retranslation, contrasted across model dimensions, samples 3, 4 and 5 combined.

Model dimension	Average interjudge agreement (%)	95% confidence interval		Number of incidents
		Lower	Upper	
Response	88.9	85.1	92.7	69
Follow-through	81.3	77.5	85.0	80
Friendly	89.4	86.0	92.9	82
Courtesy	85.9	82.9	89.0	116
Extra mile	81.1	78.2	84.1	122
Empathy	81.1	74.8	87.4	30
Listening	83.0	77.9	88.0	45
Information	87.0	84.3	90.7	104
Processing	83.3	77.6	89.1	36
Redress	88.2	85.1	91.2	107
Total	*85.4*	*84.2*	*86.5*	*791*

Table 12 shows that the amount of overlap among all possible pairs of dimensions, in terms of common attributes, is fairly low, or only 4.5%, (SD 2.8%). A high level of overlap between a pair of dimensions suggests that the retranslation judges had difficulty distinguishing between the two. The highest overlap is between the dimensions *Follow-through* and *Extra mile* (15%), followed by the pairs *Response* and *Follow-through* (11%), and *Extra mile* and *Redress* (9%).

Table 12. Dimension reliability: Incident overlap for all possible pairs of model dimensions, samples 3, 4 and 5 combined.

	Response		**Friendly**		**Extra mile**		**Listening**		**Processing**
	Follow-through		**Courtesy**		**Empathy**		**Information**		
Follow-thr.	.11								
Friendly	.03	.03							
Courtesy	.06	.07	.08						
Extra mile	.03	.15	.03	.03					
Empathy	.01	.03	.05	.04	.07				
Listening	.01	.02	.02	.05	.03	.03			
Information	.02	.07	.01	.05	.04	.04	.06		
Processing	.02	.07	.01	.02	.08	.04	.07	.06	
Redress	.01	.05	.02	.07	.09	.04	.02	.03	.06

Based on the findings shown in tables 11 and 12 it seems that the dimensions *Follow-through* and *Extra mile* would be most likely to gain in reliability and distinctiveness by a revised definition.

REFERENCES

Austin, J.T. & Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology, 77*, 836-874.

Barnes, J.G. & Glynn, W.J. (1994). Beyond technology: The consumer wants service, do employees agree? A cross-cultural exploration. In T.A. Swartz, D.E. Bowen & S.W. Brown (Eds.), *Advances in services marketing and management* (Vol. 3, pp. 219-239). Greenwich, CT: JAI Press.

Bitner, M.J. (1990, April). Evaluating service encounters: The effects of physical surroundings and employee responses. *Journal of Marketing, 54*, 69-82.

Bitner, M.J., Booms, B.H. & Mohr, L.A. (1994, October). Critical service encounters: The employee's viewpoint. *Journal of Marketing, 58*, 95-106.

Bitner, M.J., Booms, B.H. & Tetreault, M.S. (1990, January). The service encounter: Diagnosing favorable and unfavorable incidents. *Journal of Marketing, 54*, 71-84.

Borman, W.C. (1991). Job behavior, performance and effectiveness. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 271-326). Palo Alto, CA: Consulting Psychologists Press.

Borman, W.C. & Brush, D.H. (1993). More progress toward a taxonomy of managerial performance requirements. *Human Performance, 6*, 1-21.

Borman, W.C., Dunnette, M.D. & Hough, L.M. (1976). *Development of behaviorally based rating scales for evaluating the performance of U.S. Navy recruiters*. (NPRDC Tech. Rep. no. 76-31).

Bowen, D.E. & Schneider, B. (1988). Services marketing and management: Implications for organizational behavior. *Research in Organizational Behavior, 10*, 43-80.

Bownas, D.A. & Bernardin, H.J. (1988). Critical incident technique. In S. Gael (Ed.), *The job analysis handbook for business, industry and government* (Vol. 2, pp. 1120-1137). New York: John Wiley.

Brown, M.G., Malott, R.W., Dillon, M.J., & Keeps, E.J. (1980). Improving customer service in a large department store through the use of training and feedback. *Journal of Organizational Behavior Management, 2*, 251-265.

Campbell, J.P. (1983). Some possible implications of "modeling" for the conceptualization of measurement. In F. Landy, S. Zedeck & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 277-298). Hillsdale, NJ: Erlbaum.

Campbell, J. P. (Ed.) (1987). *Improving selection, classification, and utilization of Army enlisted personnel: Annual report, 1985 fiscal year* (AIR Tech. Rep. no. 746). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J.P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 687-732). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J.P., Dunnette, M.D., Arvey, R.D. & Hellervik, L.V. (1973). The development and evaluation of behaviorally based rating scales. *Journal of Applied Psychology, 57*, 15-22.

Campbell, J.P., Dunnette, M.D., Lawler, E.E. III. & Weick, K.E. (1970). *Managerial behavior, performance, and effectiveness.* New York: McGraw-Hill.

Campbell, J.P., McCloy, R.A., Oppler, S.H. & Sager, C.E. (1993). A theory of performance. In N. Schmitt & W. Borman (Eds.), *Personnel Selection in Organizations* (pp. 35-70). San Francisco: Jossey-Bass.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Crosby, L.A. & Stephens, N. (1987). Effects of relationship marketing on satisfaction, retention, and prices in the life insurance industry. *Journal of Marketing Research, 24*, 404-411.

Crowell, C.R., Anderson, D.C., Abel, D.M. & Sergio, J.P. (1988). Task clarification, performance feedback and social praise: Procedures for improving the customer service of bank tellers. *Journal of Applied Behavior Analysis, 21*, 65-71.

Czepiel, J.A., Solomon, M.R., Surprenant, C.F. & Gutman, E.G. (1985). Service encounters: An overview. In J.A. Czepiel, M.R. Solomon & C.F. Surprenant (Eds.), *The service encounter: Managing employee/customer interaction in service businesses* (pp. 3-15). Lexington, MA: Lexington.

Daniels, P.W. (1993). *Service industries in the world economy*. Oxford, UK: Blackwell.

Davidow, W.H. (1988). The coming service crisis. In C.H. Lovelock (Ed.), *Managing services: Marketing, operations and human resources* (pp. 17-21). Englewood Cliffs, NJ: Prentice Hall.

Davison, M.L. (1992). *Multidimensional scaling*. Malabar, FL: Krieger.

Desatnick, R.L. (1987). *Managing to keep the customer.* San Francisco: Jossey-Bass

Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327-357.

Fleishman, E.A. & Quaintance, M.K. (1984). *Taxonomies of human performance*. New York: Academic Press.

Fleiss, J.H. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*, 378-382.

Fogli, L., Hulin, C.L. & Blood, M.R. (1971). Development of first-level behavioral job criteria. *Journal of Applied Psychology, 55*, 3-8.

Grönroos, C. (1990). *Service management and marketing: Managing the moments of truth in service competition*. Lexington, MA: Lexington.

Hayes, B.E. (1992). *Measuring customer satisfaction: Development and use of questionnaires*. Milwaukee, WI: ASQC Quality Press.

Hogan, J., Lock, J. & Brinkmeyer, K. (1995). Interpersonal skills required at work. Unpublished paper, Hogan Assessment Systems, Tulsa, OK.

Johnston, R. (1994). Service failure and recovery: Impact, attributes and process. In T.A. Swartz, D.E. Bowen & S.W. Brown (Eds.), *Advances in services marketing and management,* (Vol. 3, pp. 211-228). Greenwich, CT: JAI Press.

Koepp, S. (1988). Why is service so bad? Pul-eeze! Will somebody help me? In C.H. Lovelock (Ed.), *Managing services: Marketing, operations and human resources* (pp. 208-215). Englewood Cliffs, NJ: Prentice Hall.

Landy, F.J. & Farr, J.L. (1983). *The measurement of work performance: Methods, theory, and applications.* New York: Academic Press.

Landy, F.L., Farr, J.L., Saal, F.E. & Freytag, W.R. (1976). Behaviorally anchored scales for rating the performance of police officers. *Journal of Applied Psychology, 61*, 750-758.

Lovelock, C.H. (1985). Developing and managing the customer service function in the service sector. In J.A. Czepiel, M.R. Solomon & C.F. Surprenant (Eds.), *The service encounter: Managing employee/customer interaction in service businesses* (pp. 265-280). New York: Lexington Books.

Luthans, F. & Davis, T.R. (1990). Applying behavioral management techniques in service organizations. In D.E. Bowen, R.B. Chase & T.G. Cummings (Eds.), *Service management effectiveness: Balancing strategy, organization and human resources, operations and marketing* (pp. 177-209). San Francisco: Jossey-Bass.

Mersha, T. & Adlakha, V. (1992). Attributes of service quality: The consumers' perspective. *International Journal of Service Industry Management, 3* (3), 34-45.

Miller, G.A. (1969). A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology, 6*, 169-191.

Motowidlo, S.J. & Carter, G.W. (1990). *Development and validation of selection procedures to assess customer contact skills* (PDRI Tech. Rep. no. 199). Minneapolis, MN: Personnel Decisions Research Institutes.

Murphy, K.R. (1989). Dimensions of job performance. In R.F. Dillon & J.W. Pellegrino (Eds.), *Testing: Theoretical and applied perspectives* (pp. 218-247). New York: Praeger.

Overall, J.E. & Klett, C.J. (1972). *Applied Multivariate Analysis.* New York: McGraw-Hill.

Paajanen, G.E. & McLellan, R.A. (1993). *Programmatic development of customer service behavior criteria.* Paper presented at the eighth annual conference of the Society of Industrial and Organizational Psychology, San Francisco, California.

Parasuraman, A., Zeithaml, V.A. & Berry, L.L. (1985, Fall). A conceptual model of service quality and its implications for future research. *Journal of Marketing, 49*, 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*, 12-40.

Quinn, J.B. & Gagnon, C.E. (1988). Will services follow manufacturing in decline? In C.H. Lovelock (Ed.), *Managing services: Marketing, operations and human resources* (pp. 6-16). Englewood Cliffs, NJ: Prentice Hall.

Riddle, D.I. (1986). *Service-led growth: The role of the service sector in world development*. New York: Praeger.

Saxe, R. & Weitz, B.A. (1982). The SOCO scale: A measure of the customer orientation of salespeople. *Journal of Marketing Research, 19*, 343-351.

Schlesinger, L.A. & Heskett, J.L. (1992). De-industrializing the service sector: A new model for service firms. In T.A. Swartz, D.E. Bowen & S.W. Brown (Eds.), *Advances in services marketing and management* (Vol. 1, pp. 159-176). Greenwich, CT: JAI Press.

Schneider, B. (1990). The climate for service: An application of the climate construct. In B. Schneider (Ed.), *Organizational climate and culture* (pp. 383-412). San Francisco: Jossey Bass.

Target Stores (1996). *Speed and service at the checkouts*. Minneapolis, MN: Author.

Weitz, J. (1961). Criteria for criteria. *American Psychologist, 16*, 228-231.

Wilcock, K.D. (1989, November). Customer service behavior spelled out. *Training and Development Journal*, 79-82.

Zeithaml, V.A., Parasuraman, A. & Berry, L.L. (1992). Strategic positioning on the dimensions of service quality. In T.A. Swartz, D.E. Bowen & S.W. Brown (Eds.), *Advances in Services Marketing and Management (Vol. 1, pp. 207-228). Greenwich, CT: JAI Press.*

Editor's Note: Pagination in this article may be different from the citation source due to formatting and layout conversions