END-USER SATISFACTION AND DESIGN FEATURES OF PUBLIC AGENCIES

RICHARD F. CALLAHAN
University of Southern California
G. RONALD GILBERT
Florida International University

This study relates end-user satisfaction to three design features of public agencies that provide services. The research connects the research discussion on public participation in administrative processes with a core consideration of public administration: the design features of public organizations. The study seeks to move from the descriptive literature to an empirically grounded survey methodology that examines end-user satisfaction across varied levels of government. Based on a sample of 2,816 end users of 17 public-sector organizations, the study tests for associations between organizational performance features and service satisfaction. The findings correlate user satisfaction with three design characteristics of public agencies: agency dependence on user satisfaction for future funding, a clearly identifiable end-user focus by the agency, and the ability of the user to exercise choice in her or his future use of the agency’s services. These findings provide a methodology for survey of public preferences that connects agency performance with public-agency design.

Keywords: customer service; public organizations; politics of structural choice; organizational performance; customer surveys

Connecting public preferences to government organizations has been an enduring challenge in that “public administration has struggled for more than 30 years with how to bring the public citizens and citizen groups into the administrative process” (Thomas, 1999, p. 83). Empirical research needs to probe the complexities involved in the public’s relationship with government agencies, including effective structuring of public participation (Moynihan, 2003, p. 164). Public-management research connecting the effects of organizational structure with public-agency performance has rarely examined public satisfaction (Boyne, 2003, pp. 383-384). As Hood and Peters (2004) indicate, there is need to move beyond the “uncritical and universal adoption of poorly grounded recipes for institutional design” (p. 278) through carefully structured measurement and hypothesis testing. Indeed, there is need for empirical research that links service effectiveness with public-agency design features that are associated with current administrative reform movements. The research design of this article tests the normative assumptions that market features in public-agency contexts contribute to increased consumer satisfaction with service delivery.

RECENT FRAMEWORKS OF ADMINISTRATIVE REFORM

The current American presidential administration’s management agenda stresses the reform of federal agencies. The proposals promise more customer-focused, results-oriented,
and market-driven approaches in the design of government and assume that, as a result, “citizens will recognize improved service, and performance and citizen satisfaction will increase” (U.S. Office of Management and Budget, 2002, p. 14). The 1990s call for reinventing government emphasized a shift of focus from rules-driven organizational designs to more flexible and agile results-driven ones (Osborne & Gaebler, 1992, pp. 25, 166). Related to reinvention, but with a markedly international arena, The New Public Management (NPM) proponents offer arguments on how to organize public institutions and delivery systems, suggesting “authoritative answers to the what-to-do questions of administration” (Hood & Jackson, 1991, p. 12). The evolution of NPM heightens the need for “evidence-based learning” (i.e., to move from descriptive mapping of key traits; Hood & Peters, 2004, pp. 268, 278) to more rigorous empirical analysis of the effects of administrative reform (Barzelay, 2001, p. 161; Hood & Jackson, 1991, p. 178; R. Moe, 2003, p. 22). The findings from such empirical studies are vital to testing the assumptions about how to improve the business of government (Kelly & Swindell, 2003, p. 95).

CONNECTING THE PUBLIC

The NPM management model draws on the private emphasis on customer orientation and the value of markets along with “the acceptance of the idea that government organizations are inefficient” (Barzelay, 2001, p. 11). NPM proponents advocate adoption of private-sector management principles within the public sector, assuming competition and the efficiency of markets are good for service delivery in government (Terry, 1998, p. 195). The research design of this article tests the belief that market exposure improves end-user satisfaction with service delivery while recognizing the difficulties that exist in identifying a public agency’s “customer” (Swiss, 1992, p. 358).

Measurement of public satisfaction relates the public-agency to one part of the organizational environment—the intended service recipient. The coalignment of internal production processes with the external environment is an essential feature for understanding organizational performance and explaining organizations (Thompson, 1967). Individual satisfaction with bureaucratic encounters has been researched at the federal level (Serra, 1995, p. 178), local level (Thomas & Melkers, 1999), across jurisdictions (Swindell & Kelly, 2002, p. 62), as well as in the emergence of e-government (Thomas & Streib, 2003). Within the reinvention framework of the 1990s and operating assumptions of subsequent reform efforts, public satisfaction remains a fundamental measure of performance (Hall, 2002, p. 23).

Of particular note in the measurement of satisfaction has been the creation of the American Customer Satisfaction Index (ACSI) and the European Customer Service Index. These two customer satisfaction indices function as intangible economic indicators and are used to monitor the financial viability of companies, industries, and international trade unions (Anderson & Fornell, 2000; Eklöf & Westlund, 2002; Fornell, 2001). Today, the ACSI is used in annual analyses of customer service quality in 35 separate industries, 190 companies, and government agencies. The ACSI has been expanded in its focus to include government agencies within the United States. An application of the ACSI to New York City services has been used to discern the city services that drive overall satisfaction of residents (Van Ryzin, Muzzio, Immerwahr, Gulick, & Martinez, 2004, p. 338).

Critiques of the market model in public administration have been based on constitutional issues (R. Moe & Gilmour, 1995), equity concerns (Fountain, 2001, p. 56), the uniquely public aspects of government (DeLeon & Denhardt, 2000), and an argument that the role of
public agencies goes well beyond that of simply being service providers (Kirlin, 1996). Another concern related to reliance on satisfaction surveys is to get at the “nagging issue of democratic accountability” that faces the entrepreneurial model (Terry, 1998, p. 197). Critics of NPM juxtapose the term citizen with the term customer (Denhardt & Denhardt, 2003) as well as characterize reinvention as a retreat from legal accountability as established in the constitutional foundation of American democracy (R. Moe, 2003, p. ix).

The classical challenge outlined in American political theory is to facilitate public participation as the foundation of legitimate democracy (Redford, 1969, pp. 21, 197) while “achieving the optimal ‘mix’ of democracy and bureaucracy, under complicated and continually changing circumstances” (Waldo, 1977, p. 7). Contemporary public administration theory contends that public participation improves decision-making quality and responsiveness (Denhardt & Denhardt, 2003, p. 93) and is essential to building effective states (The World Bank, 1997).

Public agencies serve various types of end users (Hyde, 1991). Public consumers who use government services have been classified into three distinct groups (Gilbert, Nicholls, & Roslow, 1998): first, direct buyers who are the equivalent of customers operating in an open market for a government service; second, clients who operate in a limited market with few sources from which to rely on to receive a publicly provided service; and third, captives who operate in a closed market with only one source from which to receive a publicly provided service.

Surveys of end users found a strong relationship between service satisfaction and the degree of discretion available to the public customer in the selection of a service provider—higher satisfaction correlated with increased discretionary powers by the public consumer (Gilbert et al., 1998, pp. 22-23). The direct-buyer types were significantly more satisfied with the service provided them by government than were client- or captive-type end users. This research finding suggests a potential link between market-oriented design features that position the public as direct buyers, clients, or captives and end-user satisfaction.

**RESEARCH HYPOTHESES**

Additional review of the research literature on NPM and market-oriented reform of public agencies suggests the following four gaps around which the four related hypotheses posed in this research undertaking are aligned: (a) a general need for empirical research on doctrinal claims associated with reform movements such as the NPM, drawing on interdisciplinary approaches and comparative analysis for more rigorous, empirically based research (Barzelay, 2001, p. 161); (b) a need for empirical research with market–like surveys from outside the public agency to gain insight about public agency effectiveness—there is need for surveys that move away from perceptual measures of employees in the workplace (e.g., Brewer & Selden, 2000, p. 696) or of administrators assessing their agencies’ progress in meeting goals (e.g., Burke & Wright, 2002, pp. 10-11); (c) a need for empirical research that links public satisfaction explicitly with specific design features that cut across varied levels of public agencies, moving beyond a focus at one level of government such as municipal (e.g., Streib & Poister, 1999, p. 111) or federal (e.g., Hall, 2002, p. 23); and (d) a need for empirical research that examines the relationship between end-user discretion in choosing public services or products available to them, design features of public agencies, and end-user satisfaction (Kelly & Swindell, 2003; Thomas & Melkers, 1999).
As a result of these four gaps and the previous review of administrative reform and NPM literature, the following formal hypotheses are posed:

**Hypothesis 1:** Service will be rated higher by end users when they pay directly for services provided.

**Hypothesis 2:** Service will be rated higher by end users of government agencies that are dependent on users for their revenues.

**Hypothesis 3:** Service will be rated higher by users if an agency is customer focused by design.

**Hypothesis 4:** Service will be rated higher in government agencies where their public end users can exercise choice as to whether to use the services again or to go somewhere else for them.

The research hypotheses are intended to examine empirically the role of the user in payment of service, dependence of an agency on user satisfaction, users’ choice, and end-user focus by design. Combined, the hypotheses also shed added light on the need for empirical research on doctrinal claims associated with NPM and administrative reform. The independent variables from which end-user satisfaction is gauged in this study include four statements: user payment for service; dependence of the agency on satisfied end users; the extent to which an agency is deliberately designed for end-user service; and the opportunity for users to choose services. The working hypothesis of this article holds that when end-user focused design features (independent variables) are in place in the organization, user satisfaction with services and products (dependent variables) will be higher than when the design features do not exist.

**STUDY SAMPLE**

This study included only government agencies having clearly defined end users with whom the agencies’ service providers have direct contact. Not all government agencies directly serve end users. Therefore, many public organizations such as regulatory agencies and the like were not included in this study because they do not serve direct end users as customers.

The study sample consisted of 17 public agencies and included 2,816 public-agency end users. The study sample included public agencies at all levels of government: Six agencies were federal (N = 884), six were state (N = 1,320), and five were local (N = 812). Although the types of agencies are noted in this report, the specific service ratings of each agency have been withheld to guard against the possibility that findings for any one location of an agency would be inappropriately used to generalize about an entire agency’s effectiveness. Although the agencies included in the survey represented three levels of government, no controls were conducted to measure the effect of level of government on service-satisfaction outcomes. Testing for the effects of the level of the government requires a better matching of agencies selected for each of the three levels of government than that which was attained in the sample gathered in this study. However, such an assessment is strongly recommended for future research in this area of inquiry.

The federal agencies included in this study consist of one U.S. Immigration and Naturalization district office (N = 189), two U.S. Office of Personnel Management training centers (N = 74), a Veteran’s Administration hospital clinic (N = 113), a U.S. Social Security local branch office (N = 208), and a U.S. Postal Service branch office (N = 100).

The state agencies consist of a state university library (N = 196), two state university cafeterias (N = 376), an intercity rail transportation system (N = 400), a state university public
safety department (N = 200), and a department of motor vehicles local branch office (N = 148).

The local agencies include a municipal building permit office (N = 112), a city-run mass transit rail system (N = 274), a municipal information and payment desk (N = 93), an airport (N = 159), and a county library branch (N = 174).

The agencies in this study were selected as available data sets—based on the research team’s access to these agencies. Once an agency was selected for this research undertaking, the end users who comprised the pool of candidates from each agency were chosen through a randomized process. All respondents were queried immediately after each completed a service episode at an agency. Of the 17 public agencies included in the sample, 15 were from south Florida (N = 2,427), and two (those from the U.S. Office of Personnel Management) were from the Rocky Mountain region (N = 74) of the United States.

Because differences in service satisfaction based on age, gender, and ethnicity/race have been noted in previous studies published by one of the authors, information about the customer respondents on these three demographic characteristics was recorded at the time they completed the survey. The demographic characteristics of the respondents included age classed as 1 = 29 years or younger (N = 1,157) and 2 = 30 and older (N = 1,362); gender as 1 = male (N = 1,396) and 2 = female (N = 1,368); and ethnicity/race as 1 = Asian, Black, Hispanic, and other (N = 1,509) and 2 = Caucasian (N = 782). Table 1 identifies the number and percentage of respondents in each of the three demographic categories measured in this survey. The differences in sample sizes between categories reported in Table 1 are because of the inability of the researchers to gain age, gender, education, and ethnicity/race at all the sites where the data were collected. Because all but 74 of this sample were derived from the southeastern region of the United States, the ethnic/racial composition of the study sample consists of a larger percentage of Hispanic and Black than would be found in populations in some other regions. The respondents’ demographics are similar to the actual demographics of the Miami-Dade and south Broward County geographic populations, where most of the data were collected. Although the study does not include demographic data regarding the profile of end users at each of the agencies surveyed, the overall demographics are similar to the demographics of those who live in this urban area.

The characteristics of the study sample are reported in Table 1. Each of the three categories included only two groups, and consistent statistical procedures were applied in this analysis.

**ASSESSMENT INSTRUMENTS**

Data for this study were obtained through two separate independent instruments: the first instrument gauged the design features of the public agencies sampled, and the second instrument identified the level of end-user satisfaction with the products and services provided. The two instruments include ratings of the extent to which an agency has the design features associated with reform and a standardized end-user satisfaction survey tool. In the data-gathering phase of this study, the assessment of the design features of the agencies and the assessment of end-user satisfaction ratings were identified separately as well as by different research teams.

The instrument schedule used to assess the design features of the agencies was tightly structured, leaving little opportunity for discretion on the part of the raters when classifying agency features captured in this study. Research literature in decision making supports this
approach, finding little difference between the ratings of novices and experts when using highly structured instruments within routine-task environments (McIntyre & Currim, 1982; Spence & Brucks, 1997). As noted below, the specific statements employed in the instrument were straightforward with specific criteria from which expert opinion would lend little advantage over the decision of novices. The specific statements and dichotomized scales included in the instrument were as follows:

1. Direct User Involvement in Pay for Service (V36)
   a. Low (no money or a minimal fee is paid by the user)
   b. High (a substantial portion of the fee for services is directly paid for by the user)
   Basic query: “Does the customer personally pay for services ‘out-of-pocket’ or are services provided through other funding sources?”

2. Agency’s Dependence on User Satisfaction for Future Funding (V38)
   a. Low (very low and low)
   b. High (moderate, high, and very high)
   Basic query: “Does the agency’s future funding depend on end-user satisfaction, or is it based on other types of organizational allocations that are not market driven?”

3. Organization’s Degree of End-User Focus (designed to treat the user like a “customer”) (V41)
   a. Low (very low and low)
   b. High (moderate, high, and very high)
   Basic query: “Is the system designed to make the end-user’s experience from ‘start to finish’ successful?”

4. Degree to Which the End User Has Choice to Decide to Return to the Agency or to Go Somewhere Else (V42)
   a. Low (very low, low, and moderate)
   b. High (high and very high)
   Basic query: “Does the end user have a little or a lot of discretion when choosing a public agency or government supplier?”

The characteristics associated with 2, 3, and 4, above, were dichotomized based on median scores. However, in future studies, where larger samples of agencies and respondents may be possible, it is recommended that more levels of categories be included in the scales.

Independent of the end-user satisfaction survey, a panel of trained observers classified each of the public agencies included in the sample as either low or high in specific design features. The panel of trained observers was not involved in the administration of the customer

<table>
<thead>
<tr>
<th>TABLE 1: Characteristics of the Study Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic Characteristics</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Younger than 30</td>
</tr>
<tr>
<td>30 and older</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Ethnicity/race</td>
</tr>
<tr>
<td>Non-Caucasian</td>
</tr>
<tr>
<td>Caucasian</td>
</tr>
</tbody>
</table>
satisfaction survey used in the study to assess end-user satisfaction, working independently from those who solicited satisfaction ratings of end users in the study sample. The trained observers were six student scholars (the average age was 29), studying organizational behavior. They were trained in making independent judgments about the agencies by one of the authors of this study prior to making their assessments. Working in two subgroups of three individuals, each subgroup made independent on-site visits to the organizations where customer satisfaction survey data were gathered from samples of each agency’s end users. These subgroups then met and compared their independent observations about the agencies’ design features with one another as a panel. The panel then established a consensus rating for each of the agency’s four design features based on the rating form developed by the research team. Because of geographic distance, it was not possible for the rater teams to make an on-site visit to one agency. Instead, the panel interviewed one of the authors, who was highly familiar with the organization, and they also conducted a telephone interview with the agency’s on-site program manager to rate the organization on each of the four design features.

Assessments of three of the four design features (i.e., V36, V38, and V42) were based on objective data readily available from agency records and community records. However, assessment of V41 required more discretion on the part of the trained observers in deciding the degree to which the agency focused on user satisfaction. To aid the panel, the following standards were set forth: (a) signage to help the customer know what to do, (b) ease of entry, and (c) a hospitable environment (décor, seating, spatial features) for the customer from start to finish during the service episode.

Table 2 identifies the agencies in the study sample and the ratings received from the panel of experts on the four design features.

The design features classified in this study cannot be generalized to the overall agency but are site specific. For example, a Social Security branch office with 208 respondents was rated Low on V36 (“User pays directly for service”); Low on V38 (“Agency’s funding depends on

<table>
<thead>
<tr>
<th>Agency</th>
<th>Sample Size</th>
<th>Variable 36, User Payson Satisfied Customers</th>
<th>Variable 38, Agency Dependence</th>
<th>Variable 41, Customer Focus</th>
<th>Variable 42, Customer Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. immigration</td>
<td>189</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>University library</td>
<td>196</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>College cafeteria</td>
<td>169</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>City building permits</td>
<td>112</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>College cafeteria</td>
<td>207</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>University public safety</td>
<td>200</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>U.S. OPM training center</td>
<td>52</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>County tri rail</td>
<td>400</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Municipal metro rail</td>
<td>274</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>U.S. post office branch</td>
<td>100</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Municipal help desk</td>
<td>93</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>State motor vehicles branch</td>
<td>148</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>County regional airport</td>
<td>159</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>County library</td>
<td>174</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>VA hospital</td>
<td>113</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>U.S. Social Security branch</td>
<td>208</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>U.S. OPM training center</td>
<td>22</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

OPM = Office of Personnel Management.
user satisfaction at time of service"); Low on V41 ("End-user focused"); and Low on V42
("User has choice to return for additional services or to go somewhere else for them").
Although these ratings would not be deemed appropriate when applied to the overall service
quality of the Social Service Administration (SSA) as a whole, they were deemed appropri-
ate to the agency settings in which direct service to end users was observed. In a more com-
prehensive study, the customer satisfaction index ratings of general satisfaction with the SSA
were found to be relatively high among the general public. The design features found in this
particular agency setting are not representative of the SSA.

SATISFACTION RATING INSTRUMENT

A widely used instrument termed by others as a Customer Satisfaction Survey was used
for this purpose (Gilbert et al., 1998, pp. 241-243; Gilbert, Roslow, & Nicholls, 1997, pp. 23-
24). The instrument has been validated in both cross-cultural and cross-industry settings.
The instrument consists of 17 statements based on well-recognized service quality features.
The instrument includes two empirically derived independent factors to gauge end-user sat-
isfaction. The factors are Satisfaction With Personal Service (courtesy, timeliness, being
treated as a valued customer, easy to get help) and Satisfaction With the Service Setting (con-
venient hours, place neat and clean, personal safety and security). The survey tool is espe-
cially useful in the measurement of user perceptions of service quality immediately follow-
ing a service episode when the user has the service encounter fresh in mind.

The sampling procedure employed was a systematic probability sampling—one of the
most widely used sampling techniques. The procedure produced a representative sample in
an economically efficient manner and within a short time period. At each agency, surveyors
used a skip-interval approach to select government agency end users. This approach ensured
reasonable randomness among the eligible pool of candidates while making possible a
known and equal probability for any person in the population to be selected into the sample.

The survey instrument asked the respondent to indicate the extent to which he or she dis-
agrees or agrees with each of the 17 statements (1-17) using a Likert-type scale (1 = strongly
disagree; 5 = strongly agree). The two measures were originally identified by the factor anal-
ysis statistical procedure using principal components and varimax rotation. Thus, the mea-
sures have zero correlation with one another.

Factor analysis is especially important when measuring customer satisfaction across
organizations or industries when standardized measures are required. The customer satisfac-
tion literature has reasonably established that standardized measures are appropriate for
external reporting, but customized measures are preferred to reflect specific aspects of a
given organizational entity (McColl-Kennedy & Scheneider, 2000). Therefore, this study
employed the two standardized measures rather than specific customized measures that
would be unique to each specific government setting but not comparable across the
multiorganizational sample used in this investigation.

The overall satisfaction with both setting and personal service was also correlated with
Variable 18, which is a criterion statement pertaining to the rater’s overall satisfaction with
the product and service provided through the agency. This statement measured the conver-
gent validity of the two factors identified by them. That is, a high score on a factor is strongly
correlated with the end-user’s overall satisfaction with the service or product received. This
type of validity measurement has been established as an acceptable method in several studies
in the customer-satisfaction measurement literature (Cronin & Taylor, 1992; Parasuraman, Zeithaml, & Berry, 1985).

In addition to the criterion statement, three other statements were added to the instrument to identify the respondents’ age (1 = younger than 30, 2 = 30 and older), gender (1 = male, 2 = female), and ethnicity/race (1 = non-Caucasian, 2 = Caucasian). This allowed the researchers to control for demographic factors when testing for the association between design features and user satisfaction.

TESTING FOR USE IN PUBLIC SECTOR AGENCIES

Although the customer satisfaction survey originally sampled consumers from some public-sector agencies, the two factors were derived from a database consisting primarily of private-sector organization consumers. Because of this, a need existed to retest the measures in terms of their suitability within the public-sector alone to assure that the measures were suitable for government agencies. This concern about suitability is well documented in research literature where factors identified in one study have not been replicable in another (Gorsuch, 1974). Although this task was laborious, it could not be assumed that factors identified elsewhere were suitable measures to apply in a subsequent study unless the next study requires the use of the same variables and true random sampling from the same population (Gorsuch, 1974, p. 182). These criteria delineated by Gorsuch for suitability could not be met in this study, suggesting the need to retest the measures for public-sector application.

The empirical methods used in this study to identify end-user satisfaction measures in an exclusive public-sector sample were the same as those reported in the original primarily private-sector study from which the two end-user satisfaction factors were derived. Factor analysis is partly founded on the principle of parsimony, with some variation to be expected from the factors previously identified in earlier studies (Nunnally, 1967). Factor analysis procedures were employed with principal components and varimax rotational procedures using the pairwise method when encountering missing variables. Two decision rules were used to identify variables that load on a factor: One, all had to have a factor loading of at least .50; and two, none could have a split loading with any other factor above .35. Researchers typically consider variables that, at a minimum, have a loading of at least .30 to be considered to be part of a factor (Russett, 1967). More conservatively, some researchers may use .35, .40, or even higher levels as the cutoff (Aron & Aron, 1994, pp. 512-515). The more conservative .50 was used in this study. Any variable so qualified as part of a factor also had to satisfy the test of scale reliability based on the Cronbach alpha score of at least .60. The application of these decision rules provides added confidence that the customer-satisfaction measures employed in this study are, indeed, applicable to public-sector organizations, per se.

Table A1 (see appendix) reveals the similarity between the original factors that were derived from a primarily private-sector sample using the customer satisfaction survey instrument (Gilbert et al., 1997, p. 23) and those factors derived from the public sample used in this study. This similarity provides empirical support to demonstrate that the two factors derived from the public sample are equivalent to those derived in previous studies.

The use of these two factors provides greater insight about customer satisfaction than would be possible to infer from Variable 18 alone. In both studies, the two factors were highly correlated with Variable 18—the criterion validity measure that identified each respondent’s overall assessment of the product and service quality that he or she received.
The ratings of the design features (low vs. high agency characteristics for V36, 38, 41, 42) of each agency were added to the data records attained from the customer satisfaction survey along with data gathered on Variable 18 (convergent validity measure used for internal validation of constructs) and the respondents’ age, gender, and ethnicity/race. Thus, each of the 2,816 records in this study sample included data pertaining to each respondent’s ratings about the service each received from a public agency using the customer satisfaction survey, the high or low design features attributed to the agency rated, and the end user’s age, gender, and ethnicity/race. These steps lead to one data set for each of the 2,816 records that were used for data analysis in this study to test the assumptions of reinvention and reform against actual service satisfaction by end users.

**SATISFACTION RATINGS COMPARED WITH AGENCY DESIGN FEATURES**

Presented in Table 3 are the end-user satisfaction ratings on the two measures (dependent variables), *Satisfaction With Personal Services* and *Satisfaction With the Service System*, with the high and low ratings of the design features of public-service delivery systems in the study sample. The independent variables are the agency features.

**STUDY FINDINGS**

The results of MANOVAs on *Satisfaction With Personal Service* and *Satisfaction With the Service Setting* measures and t tests for each of the four design features were significant (*p* < .001).

Respondents revealed more satisfaction with both personal service and service setting when the agency was more dependent on user satisfaction for future funding (high on V38), the agency’s focus was higher on user satisfaction (high on V41), and the degree to which the user has the choice to return or go somewhere else was higher (high on V42). However, the opposite held for V36. Respondents were more satisfied with both personal service and service setting when the user paid little or no fees for services (low on V36). Because of the large sample sizes that comprised this study, $\eta^2$ statistics were also reported to identify the effect of size on design features. Typically, $\eta^2$ values of .01 are considered as small, .06 as medium, and .14 as large (Green & Salkind, 2003, p. 162). $\eta^2$ statistics revealed small effects associated with V36 on both measures and with the service setting satisfaction measure associated with V38. Medium to large effects were identified with both satisfaction measures for all other design features. Thus, even in smaller samples, the medium to large effects found in this study would have been significant.

When controlling for the effects of differences in respondents’ age, gender, and ethnicity/race through the application of the MANCOVA procedure as illustrated by the Wilks’s $\Lambda$ analysis, no differences in respondents’ satisfaction levels were found based on V36, pay for service (Wilks’s $\Lambda = .998_{(2,1984)}$, $F = 2.16$, $p < .115$). When controlling for age, gender, and
The analysis supports three out of the four hypotheses associated with this research effort. No effect was found when users pay directly for services (V36) and when they do not. However, significant differences were found when the agency depends on the user for its revenue (V38); the agency, itself, is customer-focused in the way it does its business (V41); and these differences are also evident when the user can exercise choice in deciding to return to the agency or go somewhere else to get the service he or she needs or expects (V42).

### TABLE 3: Relationships Between Organizational Design Features and End-User Satisfaction

<table>
<thead>
<tr>
<th>Measure</th>
<th>Sample Size</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>df</th>
<th>η²</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable 36: User Pay for Service⁵</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with personal service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>721</td>
<td>3.64</td>
<td>.865</td>
<td>5.73</td>
<td>2499</td>
<td>.013</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>High</td>
<td>1780</td>
<td>3.41</td>
<td>.925</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with service setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>733</td>
<td>3.75</td>
<td>.766</td>
<td>3.36</td>
<td>2687</td>
<td>.004</td>
<td>.001</td>
</tr>
<tr>
<td>High</td>
<td>1956</td>
<td>3.62</td>
<td>.883</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable 38: Agency’s Dependence on User Satisfaction for Future Funding⁷</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with personal service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1567</td>
<td>3.32</td>
<td>.932</td>
<td>−11.42</td>
<td>2499</td>
<td>.05</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>High</td>
<td>934</td>
<td>3.74</td>
<td>.819</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with service setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1750</td>
<td>3.57</td>
<td>.870</td>
<td>−7.29</td>
<td>2687</td>
<td>.02</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>High</td>
<td>939</td>
<td>3.82</td>
<td>.800</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable 41: Agency’s Focus on User Satisfaction⁴</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with personal service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1260</td>
<td>3.18</td>
<td>.918</td>
<td>−17.06</td>
<td>2499</td>
<td>.11</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>High</td>
<td>1241</td>
<td>3.77</td>
<td>.809</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with service setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1548</td>
<td>3.50</td>
<td>.862</td>
<td>−11.11</td>
<td>2687</td>
<td>.04</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>High</td>
<td>1141</td>
<td>3.87</td>
<td>.798</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable 42: Degree to Which the User Has Choice to Decide to Return to the Agency or Go Somewhere Else⁶</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with personal service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1785</td>
<td>3.31</td>
<td>.907</td>
<td>−14.53</td>
<td>2499</td>
<td>.08</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>High</td>
<td>716</td>
<td>3.88</td>
<td>.801</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with service setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1962</td>
<td>3.53</td>
<td>.858</td>
<td>−12.70</td>
<td>2687</td>
<td>.06</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>High</td>
<td>727</td>
<td>3.99</td>
<td>.749</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Based on a 5-point Likert-type scale from 1 (strongly disagree) to 5 (strongly agree).

b. Wilks’s Λ = .981 (2,385) F = 23.43, p < .001.

c. Wilks’s Λ = .936 (2,385) F = 81.49, p < .001.

d. Wilks’s Λ = .904 (2,385) F = 126.44, p < .001.

e. Wilks’s Λ = .903 (2,385) F = 127.93, p < .001.

ethnicity/race, the other three design features remained significant (V38, Wilks’s Λ = .998 (2,1984) F = 7.68, p < .001; V41, Wilks’s Λ = .962 (2,1984) F = 37.72, p < .001; and V42, Wilks’s Λ = .906 (2,1984) F = 97.80, p < .001).

The analysis supports three out of the four hypotheses associated with this research effort. No effect was found when users pay directly for services (V36) and when they do not. However, significant differences were found when the agency depends on the user for its revenue (V38); the agency, itself, is customer-focused in the way it does its business (V41); and these differences are also evident when the user can exercise choice in deciding to return to the agency or go somewhere else to get the service he or she needs or expects (V42).
Institutional debates addressing design of public agencies are often beyond the reach or control of public managers (Light, 1997; March & Olsen, 1983). Furthermore, the interest group politics that contribute to the design of public agencies often result in political compromises that may reduce the performance effectiveness of organizations (T. Moe, 1989). The reality of political design at the institutional level creates a sense of immutable features that public managers inherit rather than shape. To the extent that managers can manage upward and outward (Moore, 1995), the findings from this study identify design features that public managers can seek to incorporate in structuring their systems.

This study empirically tests the assumptions of reform, where adoption of market features of service delivery within public organizations is encouraged and end-user satisfaction is assumed to be an indicator of organizational effectiveness. The application of customer satisfaction measures in the public sector is not always appropriate; thus, these measures should be used with caution and restraint. At the same time, end-user satisfaction can serve as one useful measure of organizational performance in the needed move to a parsimonious set of measures (Rainey & Steinbauer, 1999, p. 28).

The research methodology employed in this study is limited by the methods used to sample public agencies included in the database and by the types of agencies surveyed. Yet, in spite of these limitations, the research findings shed light on the application of customer-satisfaction ratings and provide insight about the relative effectiveness of the design features of public agencies.

This empirical approach suggests the possibility of transitioning from the rhetoric of reinvention and other reforms to the actual undertaking of testable research questions pertaining to them. Moreover, the methodology may provide an approach for testing a wide range of performance questions across various public agencies and levels of government.

There are limits to the findings as the agencies selected for the study do not represent the full range of agencies or roles performed by public agencies. Also, many of the end users selected enjoyed greater flexibility and choice than is often possible in the public sector. Therefore, they were not fully representative of all end users who receive services and products from government organizations, and the findings need to be viewed accordingly. In addition, more work is needed in the identification and operational measurement of the design features of public agencies. This study employed four such measures, but this research thrust would benefit measurably in the future from the inclusion of more independent variables and more concrete means of measuring the degree to which they exist.

In this study, the independent variables were presented in dichotomous form (i.e., low and high) and with a moderate measure being variously included in either the high or low groupings. In future research, a larger sample of government agencies and respondents would make it possible for more sophisticated classifications of the characteristics of public agencies that are of interest in this field of study. In future research, it might also be helpful to systematically match similar types of federal, state, and local agencies to compare user satisfaction by government level.

Despite these limits, the findings empirically connect three features of public-agency design with end-user outcomes. The research design narrows the search for parsimony in variables that connect agency performance with public-agency design. Across a variety of agencies at the federal, state, and local level, a wide range of respondents reacted favorably to public agencies characterized by delivery systems that incorporate local planning, end-user choice, and an end-user satisfaction focus. The findings imply the importance of these
features for end-user satisfaction and have implications for the design of delivery systems for government services.

The survey findings identify the end-user aspect of reform with specific features in public-agency design. For the types of agencies surveyed, this suggests an empirical link between the theory of the intervention (agency design) and the theory of the outcome (end-user satisfaction). However, end users are not the only stakeholders for public agencies. Research in best practices in public agencies has found successful public managers serve a wide range of stakeholders beyond end users (Moore, 1995). The question of measuring performance through satisfaction of a wider range of stakeholders beyond direct users is not addressed by this study. But, for the piece of performance that measures end-user satisfaction, the research provides support for linkages between agency design and end-user satisfaction.

This study does not support or reject pay-for-services arguments that are critically important in a wide range of public policy debates and in the introduction of market-like choices among end-user options. The study sample was not designed specifically to capture all areas (e.g., school vouchers) where that is a pressing public-policy issue. Yet, the methodology introduced in this study might serve to provide a platform from which user satisfaction with various forms of pay-for-services and alternative forms of service delivery might be measured and their effectiveness from the eye of the user gauged.

More generally, in identifying measurable agency features that are associated with end-user satisfaction of government agencies, the findings provide an empirical link between the organization and a piece of the external environment. As has been demonstrated in business, the case may be made that the level of service quality provided by government agencies contributes to the overall quality of life in America.

The study sheds light on the utility of acquiring end-user data. When public managers have the discretion to affect agency design features, a survey of end users can prove useful. The study also offers policy makers some empirical basis for the creation of design features of public agencies that are well received by the end users. The research provides a methodology to connect accountability in a democracy to the market focus suggested by NPM and reinvention reform. The measurement of individual satisfaction with a public agency’s service delivery can be a possible mechanism for increasing administrative and political accountability. Rather than waiting for election cycles, public satisfaction can be measured in real-time fashion. Not to overstate the case, but empirically linking public agency performance to end-user satisfaction has the potential to provide for the meaningful participation and authentic accountability that critics of market-based reform decry as lost. The findings suggest the compatibility of a focus on customer service and responsiveness of public agencies.

The findings do not imply that the agencies that were identified with higher satisfaction ratings are more effective overall than are the other agencies in the study sample. End-user satisfaction is but one metric in evaluating public-agency performance. Research that measures public-agency performance in fiscal and management capacity, as well as evaluation research on a variety of metrics, needs to be linked to the end-user satisfaction research offered here to provide more comprehensive comparative analyses.

The potential to empirically test reform doctrinal claims is a step forward. The linking of end-user perceptions with other measures of organizational performance broadens the empirical platform from which academicians and practitioners may gain insight about the design of public agencies. Developing a theory about how and why the link between users who depend on funding and a focus on user satisfaction with users who have other choices
get better ratings would entail additional research that not only replicates the current findings but also moves to connect additional, nonperceptual measures of performance to the end-user perceptual measures.

To proponents of reform, the findings offer cautious support for some important doctrinal claims on the efficacy of market-like forces in public management. For opponents of reinvention-inspired reform, the findings offer a potential to honor a commitment to administrative and political accountability as well as public participation through empirical measures of end-user satisfaction. Research on end-user satisfaction offers an empirical approach to bringing together reformers focused on market theory and critics focused on democratic theory. Shared concern for the public provides a foundation for future research to empirically address efficiency, accountability, and participation in responsible self-governance.

APPENDIX

Statements Used to Derive Measures Pertaining to Characteristics of Public Service Delivery Systems

The specific statements and scales included in the schedule to measure characteristics of public service delivery systems were as follows:
Variable 36. Direct User Involvement:
1. No money is paid for services by user
2. A minimal fee is paid by the user
3. A substantial portion of the fee for services is paid for by the user

Variable 38. Agency’s Dependence on User Satisfaction for Future Funding
1. Very low
2. Low
3. Moderate
4. High
5. Very high

Variable 41. Degree of Customer Focus (Designed to Treat the User Like a Customer) of Organization
1. Very low
2. Low
3. Moderate
4. High
5. Very high

Variable 42. Degree to Which the Customer Has Choice to Return to the Agency for Future Service if the Service They Receive Is Not Satisfactory
1. Very low
2. Low
3. Moderate
4. High
5. Very high

TABLE A1: Comparison of Factor Loadings Identified in the Original Private-Sector Dominant Study
With the Public-Sector Sample Used in This Study

<table>
<thead>
<tr>
<th>Variable</th>
<th>Public-Agency Factors</th>
<th>Original Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2002 Factor 1</td>
<td>Factor 2</td>
</tr>
<tr>
<td>1 Provider courtesy</td>
<td>.802</td>
<td>.762</td>
</tr>
<tr>
<td>2 Timely service</td>
<td>.756</td>
<td>.781</td>
</tr>
<tr>
<td>4 Easy to get help</td>
<td>.360</td>
<td>.444</td>
</tr>
<tr>
<td>6 Neat and clean place</td>
<td>.654</td>
<td>.605</td>
</tr>
<tr>
<td>8 Easy access to service</td>
<td>.768</td>
<td>.740</td>
</tr>
<tr>
<td>10 Security within the organization</td>
<td>.748</td>
<td>.751</td>
</tr>
<tr>
<td>12 Prompt help</td>
<td>.760</td>
<td>.700</td>
</tr>
<tr>
<td>14 Fair treatment</td>
<td>.732</td>
<td>.372</td>
</tr>
<tr>
<td>16 Helpful personnel</td>
<td>.806</td>
<td>.769</td>
</tr>
<tr>
<td>18 Overall, product and service quality</td>
<td>.848</td>
<td>.819</td>
</tr>
<tr>
<td>% of variance</td>
<td>54.58</td>
<td>6.39</td>
</tr>
<tr>
<td>Alphas</td>
<td>.93</td>
<td>.72</td>
</tr>
<tr>
<td>Pearson correlation Factors 1 and 2 with</td>
<td>.81</td>
<td>.54</td>
</tr>
<tr>
<td>Variable 18 (criterion variable)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Shaded variables indicate the items met the preestablished criteria for interrelatedness and independence necessary to be included in one of the two factors that comprise this study. The decisional rules used to identify variables that load on a factor is that each must have a loading of at least .50 without being split loaded on another factor above .35.

a. The Cronbach Alpha scale reliability was less when this was included than when deleted, thus it was deleted.
b. Variable 18 was used as the criterion measure from which the other factors could be validated. It was omitted as a variable in the factor analysis procedure.

REFERENCES


Richard F. Callahan is associate dean and director of state capital and leadership programs for the University of Southern California (USC). Since 1998, he has directed three graduate degree programs in Sacramento as well as directing five leadership training programs for national, state, county, and local government officials and nonprofit health directors. He teaches graduate management classes at USC and has been published in the Public Administration Review and Journal of Public Administration Research and Theory. Recently, he served on a team advising the newly elected governor’s cabinet appointees for a state in Mexico. He has his doctorate in public administration from USC. His e-mail is rcallaha@usc.edu

G. Ronald Gilbert is an associate professor of management in the College of Business Administration at Florida International University. There he serves as faculty director of the MBA program for Miami-Dade county public managers. He took his doctorate from the University of Southern California in public administration. He has recently published in journals such as the European Journal of Marketing, Omega, The International Journal of Management Science, Public Personnel Management, Journal of Services Marketing, Journal of Applied Social Psychology, The Public Manager, and the Journal of Consumer Marketing. He has developed a variety of assessment tools that are used by organizations in the United States and abroad. He consults, trains, and conducts research for organizations in the Americas, Europe, and Asia. His Web site is www.Gilbertems.com