2004

A Methodology for Business Value-Driven Website Evaluation: A Data Envelopment Analysis Approach

Jungpil Hahn
Purdue University, jphahn@mgmt.purdue.edu

Robert J. Kauffman
University of Minnesota, RKauffman@csom.umn.edu

A Methodology for Business Value-Driven Website Evaluation: A Data Envelopment Analysis Approach

Jungpil Hahn
Purdue University
jphahn@mgmt.purdue.edu

Robert J. Kauffman
University of Minnesota
rkauffman@csom.umn.edu

ABSTRACT

Managers at e-commerce firms are in need of proven methods for ongoing website evaluation. However, current approaches to website evaluation are not perfectly suited to the task at hand. This paper proposes a new business value-driven approach to website evaluation, which is theoretically grounded in the *economic theory of production*. We view online shopping as an economic production process in which customers are using various functionalities of an e-commerce website in order to complete a purchase transaction. This view enables us to formulate a novel perspective on website performance — the ability to transform inputs (i.e., use of website functionalities) into outputs (i.e., completed purchase transactions). We propose two DEA-based metrics, *InefficiencyBreadth* and *UnitInefficiency* that help identify website functionalities that are potentially ineffective.

Keywords

B2C e-commerce, website evaluation, business value, website effectiveness metrics, data envelopment analysis

INTRODUCTION

Web usability is a critical success factor in Internet-based selling. Consequently, one of the most pressing questions on the minds of e-commerce managers is whether the design of their online storefronts is effective, and if not, which aspects require attention. Answers to such questions allow managers to prioritize (re)design projects to maximize return on investment of the firm's development initiatives. However, due to the lack of proven methods for e-commerce website evaluation, gaining insights into the effectiveness of e-commerce websites is not a simple matter.

Current approaches to e-commerce website evaluation fall into three categories (Ivory and Hearst, 2001) – user testing (e.g., Spool, Scanlon, Schroeder, Synder and DeAngelo, 1999), inspection (e.g., Nielsen and Mack, 1994), and inquiry (e.g., Schubert and Selz, 1999). Even though these approaches have been successfully applied in the evaluation of user interfaces for traditional IS and software applications, they are not perfectly suited for web-based e-commerce applications. First, websites are frequently updated and redesigned, which makes the costs of recruiting test users, experts or survey respondents for the evaluation of each redesign excessive. Second, users of web-based applications are most often *customers*, which is untypical of traditional IS applications developed for use by employees within a firm. As a result, greater constraints are placed on what a designer/developer must do to create a desirable setting for system use by a user/customer since end-user training is not an option. Also, customers display a greater level of heterogeneity of human-computer interaction than organizational users. This makes it difficult to assume that a large enough set of usability problems will be detected with a limited number of subjects in usability studies. Finally, current methods for website evaluation focus mainly on usability (i.e., task success / error rates, task completion time and user satisfaction) without much concern for *business value*, which is more critical for commercial websites.

The purpose of this paper is to propose a business value-driven approach to e-commerce website evaluation. The proposed method is grounded in the *economic theory of production* and thus explicitly puts business value to the foreground. In addition, our proposed method makes extensive use of actual customer-website interaction data using web server logs. Such a new approach to data collection has the potential to resolve some of the aforementioned problems of current usability evaluation methods: (1) continuous collection of web server logs can enable on-going website evaluation; (2) data is collected for *all* customers making it possible to effectively cope with heterogeneity in consumer behavior; and (3) linking the customer-website interaction data to actual purchase transactions allows for an explicit consideration of business value.

The paper proceeds as follows. We first present our conceptualization of online shopping as an economic production process and briefly review the relevant foundational concepts in production economics and present the empirical method for website evaluation based on *data envelopment analysis* (DEA). We propose two new metrics to help identify website functionalities that are less than effective. The value of the proposed method is illustrated by applying it to the evaluation of a real-world e-commerce website.

ONLINE SHOPPING AS ECONOMIC PRODUCTION

Early research in service operations recognized the importance of the customer's involvement in the service production and delivery process as a source for increasing a service firm's productivity (Chase, 1978, Lovelock and Young, 1979). The notion of *customer co-production* is especially relevant when the service encounter involves the use of self-service technologies (SSTs) because customers are essentially performing the tasks that a paid employee must otherwise execute. With SSTs, since customers are the actual producers of the service good, the customers' efficiency and productivity at utilizing the SST become critical success factors for service firms (Zeithaml, Parasuraman and Berry, 1990). E-commerce websites, especially transactional web-based applications for Internet-based selling, can be viewed as SSTs (Meuter, Ostrom, Roundtree and Bitner, 2000). Customer-website interaction during online shopping can thus be conceptualized as an *economic production process* in which the customer performs a purchase transaction by utilizing various functionalities provided by the e-commerce website.

Viewing an e-commerce website as a service production environment enables us to start thinking about the evaluation of website performance: the ability to transform inputs to outputs. In the context of online shopping, the *inputs* consist of the customers' use of the various functionalities provided by the e-commerce website, which represent the effort put forth by the customer in filling their virtual shopping carts (e.g., number of product page views, extent of navigation through product listings, and references to help pages). The *outputs* are the contents of the purchase transaction, which can be represented by the number of items (or the dollar amount of items) in the shopping cart at checkout. Other factors may influence the online shopping production process, including the customer's level of experience with the website, the quality and speed of the customer's Internet connection, and so forth. Borrowing the formalism from production economics, we model the online shopping service production process as the following *production model* or as its inverse (or "dual") the *cost model*:

$$\mathbf{y} = f(\mathbf{x}, \mathbf{s}, \boldsymbol{\varepsilon}^{\mathbf{output}}) \text{ or } \mathbf{x} = g(\mathbf{y}, \mathbf{s}, \boldsymbol{\varepsilon}^{\mathbf{input}})$$

$$y_r = f(x_i, s_k, \varepsilon_r^{output}) \text{ or } x_i = g(y_r, s_k, \varepsilon_i^{input}),$$

where

$f(\cdot)$ = production function that translates inputs into outputs,
$g(\cdot)$ = cost function that translates outputs into inputs,
$\mathbf{y}$ = vector of r outputs (y_r) resulting from the production; $r > 0, y_r \geq 0$,
$\mathbf{x}$ = vector of i inputs (x_i) used in the production process; $i > 0, x_i \geq 0, \mathbf{x} \neq \mathbf{0}$,
$\mathbf{s}$ = vector of k environmental variables (s_k) influencing the production process, and,
$\boldsymbol{\varepsilon}^{\text{output}}$ = vector of r deviations from the production frontier (ε_r^{output}); $r > 0, \varepsilon_r^{output} \geq 0$.
$\boldsymbol{\varepsilon}^{\text{input}}$ = vector of i deviation from the production frontier (ε_i^{input}); $i > 0, \varepsilon_i^{input} \geq 0$

In production economics, a *production process* describes the technical means by which inputs (e.g., materials and labor) are converted into outputs (e.g., goods or services). This relationship is represented by the *production function*, which articulates the maximum level of outputs that can be produced for a given level of inputs, i.e., the *efficient* (or "best practice") *frontier*. Deviations from the frontier reflect inefficiencies in production (Aigner and Chu, 1968).

The distinction between the *output-oriented production model* and *input-oriented cost model* is important. It provides a basis for interpreting the inefficiencies in production. Inefficiencies in the production model relate to *slack output* (ε_r^{output}): more outputs could have been produced with the same amount of inputs. Inefficiencies in the cost model relate to *excess input* (ε_i^{input}): the same amount of outputs could have been produced with less input. The distinction between the production and cost models is also important because it provides us with flexibility to model different shopping behaviors (e.g., goal-directed purchasing vs. hedonic shopping). The input-oriented cost model, which seeks to minimize input given a level of output would be more appropriate for evaluating websites where goal-directed purchasing is more typical (e.g., grocery). The output-oriented production model, which maximizes output given an input level, is more appropriate in the evaluation of websites where hedonic purchasing is more common (e.g., apparel).

VALUE-DRIVEN WEBSITE EVALUATION: A DEA APPROACH

With the above conceptualization of online shopping as a service production process, we may utilize methods from production economics—frontier estimation methods—in evaluating website performance. Of the various analysis methods available, we selected *data envelopment analysis* (DEA), a non-parametric method for production frontier estimation. DEA does not assume a specific functional form for the production function. It requires just a few assumptions (e.g., a monotonically increasing, convex input-output relationship). The parametric formulation for stochastic frontier estimation and the non-parametric formulation of DEA have been shown in prior research to yield very similar results (Banker, Datar and Kemerer, 1991).

Data Envelopment Analysis and Website Performance Evaluation

In DEA, the unit of analysis is the *decision-making unit* (DMU). This represents a production unit. A DMU may be defined narrowly as an individual or as broadly as a firm, an industry, or even as an economy. DEA estimates the relative efficiencies of DMUs from observed input and output measures. The productivity of a DMU is evaluated by comparing it against a hypothetical DMU that is constructed as a convex combination of other DMUs in the dataset. We use the *input-oriented BCC model* (Banker, Charnes and Cooper, 1984) to estimate the efficiencies of the online purchase transactions in the evaluation of the effectiveness of e-commerce websites.

The efficiency h_{j_0} of DMU j_0, characterized on the basis of the consumption of inputs x_{ij_0} and production of outputs y_{rj_0}, is assessed by solving the following linear program:

$$\begin{aligned}
&\text{Min } h_{j_0} \\
&\text{subject to:} \\
&h_{j_0} x_{ij_0} = \sum_{j=1}^{J} x_{ij}\lambda_j + \varepsilon_{ij}^{input}, \quad \forall i \\
&y_{rj_0} = \sum_{j=1}^{J} y_{rj}\lambda_j - \varepsilon_{rj}^{output}, \quad \forall r \\
&\lambda_j \geq 0, \quad \forall j \\
&\sum_{j=1}^{J} \lambda_j = 1
\end{aligned}$$

Taken together, the specification of the constraints is such that the production possibilities set conforms to the axioms of production in terms of regularity, monotonicity, convexity and minimum extrapolation (Banker et al., 1984).

The DEA program is run iteratively for all DMUs (j=1…J) to yield efficiency scores h_j^*. A DMU j is said to be *fully efficient* if the optimal solution h_j^* to its linear program yields $h_j^* = 1$ with no slack (i.e., $\varepsilon_{ij}^{input} = 0$ and $\varepsilon_{rj}^{output} = 0, \forall i, r$). All other DMUs with $0 < h_j^* < 1$ are said to be *inefficient* (i.e., $\varepsilon_{ij}^{input} > 0 \text{ or } \varepsilon_{rj}^{output} > 0, \exists i, r$).

Identifying Inefficient Website Functionalities

Recall that an important managerial concern is to understand not only how the e-commerce website is performing, but more importantly, which areas of the website are not effective so as to identify areas for improvement. Efficiency estimation via DEA gives an overall view of website performance—whether or not the website is effective. If the website is ineffective, we can derive additional metrics from the DEA results to gain insights into why the website was ineffective. We next propose two metrics.

InefficiencyBreadth. The *InefficiencyBreadth* of website functionality represents how widespread inefficiencies due to the particular website functionality are. Since input i in the online shopping production model is conceptualized as the customer's use of website functionality i, all non-zero ε_{ij}^{input} represent excess input in the use of website functionality i that resulted in the inefficiency in the production of output r. If we define the set $D_i = \{ j \in J \mid \varepsilon_{ij}^{input} > 0 \}$ (i.e., all DMUs where inefficiency in the use of website functionality i was observed) and $n_i = |D_i|$ (i.e., the cardinality of D_i, the number of elements in set D_i), the proportion of n_i with respect to the total number of DMUs (J) represents the scope of inefficiency due to functionality i:

$$InefficiencyBreadth_i = \frac{n_i}{J}$$

UnitInefficiency. The *UnitInefficiency* of website functionality on output represents how much the inefficiencies due to the particular website functionality are with respect to a unit of output. The magnitude of the inefficiencies can be represented via the values of ε_{ij}^{input} for all inefficient DMUs. However, since DMUs may have differing levels of outputs, we must normalize the magnitude of inefficiency by level of output to define:

$$UnitInefficiency_{irj} = \frac{\varepsilon_{ij}^{input}}{y_{rj}}$$

EMPIRICAL APPLICATION

Data for this study were collected at an online grocery retailer which operates in a metropolitan area in the upper Midwest, where it is the only online service within its regional market. Clickstream data were collected directly from an online grocer's web servers. Typical data pre-processing procedures for using webserver logs were used to extract navigation path sequences for visitors from the clickstream data. The navigation sessions were combined to identify purchase transactions. Then website usage metrics were observed for each transaction. The data span two weeks from June 23 to July 5, 2001. Our analysis focuses on 5,383 actual completed purchasing transactions from 4,941 customers. Given the heterogeneity of online shopping behaviors between customers, rather than having each customer as a unit of analysis (i.e., a DMU), we use each completed purchase transaction as a DMU to ensure a basis for comparison.

DEA Model

A DEA analysis is only as good as the initial selection of input and output variables. The inputs must represent the resources consumed by the DMUs and the outputs the production of the DMUs. In online shopping, *inputs* consist of customers' use of various website functionalities and the *output* consists of a checkout of a basket of products (see Table 1).

VARIABLE		MEASURE	DESCRIPTION
Inputs	x_1	*Products*	# product page views
	x_2	*Lists*	# product lists views
	x_3	*Personal*	# personal list views
	x_4	*Order History*	# orders history page views
	x_5	*Search*	# search conducted
	x_6	*Promotion*	# promotional page views
	x_7	*Recipe*	# recipe page views
	x_8	*Checkout*	# checkout pages
	x_9	*Help*	# help page views
Output	y_1	*Basket Size*	# product types at checkout

Table 1. Input and Output Variables

Overall DEA Results

Figure 1 shows the aggregate efficiency scores of all DMUs (J = 5,383) against the respective output of each observation (i.e., *Basket Size* or number of items purchased). The horizontal axis represents the efficiency scores of the online shopping transactions ($0 < h_j^* \leq 1$), whereas the output level (i.e., number of items in the cart at checkout) is represented on the vertical axis. The efficient transactions lie on (or near) the right edge of the graph ($h_j^* \approx 1$). Visual inspection gives a summary of overall website efficiency. The plot shows significant variability of efficiency scores at all output levels, suggesting that the website may not be entirely effective.

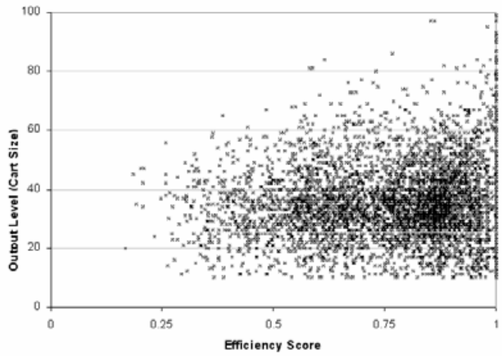


Figure 1. DEA Scores by Output Level

Inefficiency Results by Website Functionality

To gain insights into the causes of the observed overall website inefficiency, we analyzed the inefficiencies by website functionality with the inefficiency metrics proposed earlier. Figure 2 shows a radar chart with *InefficiencyBreadth* (*IB*; dashes) and average *UnitInefficiency* (*UI*; solid line) sorted by decreasing order of total *UnitInefficiency* (clockwise starting from the north). The results show that website functionalities Product Information, Search and Personal List were most problematic in incurring inefficiencies at the e-tailer's website. Inefficiencies due to website functionalities Order History, Help, Checkout and Recipe were quite insignificant. For example, we see that the website functionality of Product Information was the area in which inefficiencies were not only the most wide spread but also the most severe (*IB* =42.2%; *UI* =0.34). On the other hand, inefficiencies due to Order History were neither widespread nor serious (*IB* =2.4%; *UI* =0.13). We also observe interesting results for website functionalities, Promotion and Product List. Inefficiencies due to Promotion were widespread, but the degree of inefficiency was low (*IB* =39.4%; *UI* =0.18). Meanwhile, the scope of inefficiencies due to Product List was narrow but the degree of inefficiency was substantial (*IB* =6.67%; *UI* =0.39).

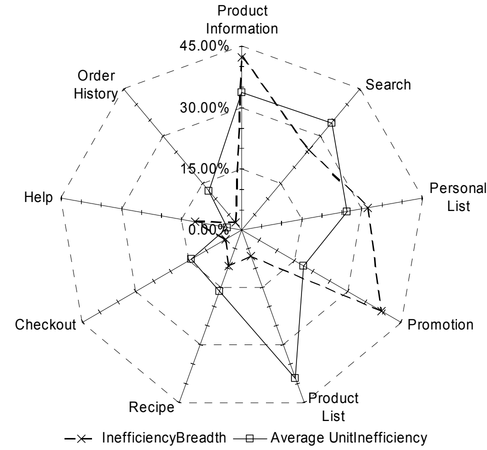
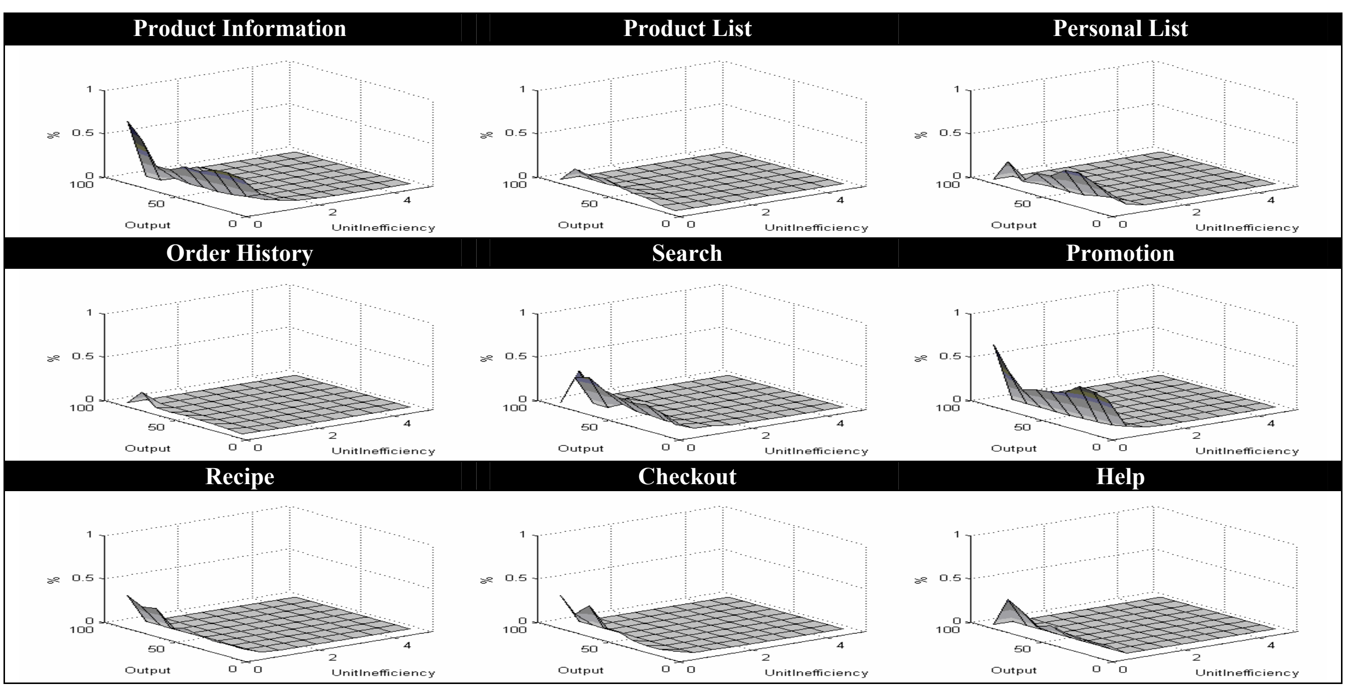


Figure 2. InefficiencyBreadth and UnitInefficiency Metrics by Website Functionality

Even though these analyses provide insights into the effectiveness of the e-commerce website, the results need to be interpreted with caution. They only focus on <u>average</u> tendencies. Given the potential heterogeneity in consumer online purchasing behaviors, we need to delve deeper into consumer behaviors that may impact website effectiveness. For example, customers that conduct high-volume transactions may exhibit different purchasing and website-interaction behaviors. We explore these issues using the metrics proposed herein.



Note: The height of the surface shows the proportion of transactions in which a particular *UnitInefficiency* value was observed for a particular output volume.

Figure 3. Distribution of UnitInefficiency Values by Output Volume and Functionality

Figure 3 shows the distributions of *UnitInefficiency* values by output volume for each of the website functionalities. Several results are noteworthy. The distribution of *UnitInefficiency* values is skewed with most observations in the lower range. *UnitInefficiency* seems to follow an exponential or half-normal distribution rather than a normal distribution. Second, we reconfirm some of the insights generated previously. On average, *UnitInefficiency* was most salient for website functionalities Product Information, Promotion, Search and Personal List. The results also indicate that higher volume transactions seem to be relatively more likely to incur inefficiencies than lower volume ones, regardless of the website functionality. This suggests that the website may be geared toward supporting smaller carts. This last finding is an interesting result when we consider the nature of the product being sold by the online grocer. The overall design strategy of the current website is typical in Internet-based selling. What the results may be suggesting is that even though such a design strategy may be effective for e-tailers where the typical number of products being purchased is small (i.e., cart size of 1 to 5 items), a different overall design strategy may be required for grocery shopping where the number of different items being purchased is larger (i.e., cart size of 40+ items).

5. CONCLUSION AND DISCUSSIONS

Evaluating the effectiveness of e-commerce website design is an important, yet complex problem for e-commerce retailers. Their success hinges largely on the ability to provide a high-quality website. So e-commerce retailers need to constantly monitor the effectiveness of their web-based storefronts. However, current methods do not offer practical means for a solution to this problem. We proposed an innovative methodology for e-commerce website evaluation based on the conceptualization of online shopping as economic production.

By modeling online shopping as an economic production process and using evaluative methods for frontier analysis, we defined and estimated a value-driven model for website effectiveness. We also demonstrated the value of our method by applying it to the evaluation of a real-world e-commerce website. Through the application, it was possible to gain a deeper understanding of which website functionalities were potentially problematic. It was also possible to discover unexpected knowledge related to the potential inappropriateness of the overall design strategy of the e-tailer's website. Although we do not have conclusive results with respect to this last point, such knowledge discovery provides a useful starting point for delving deeper into these issues.

The proposed website evaluation methodology provides significant benefits over current widely used methods. The empirical insights generated could not have been uncovered using the traditional methods of user testing, inspection or inquiry. One of the major advantages of the proposed methodology is that firms can make use of observable customer actions for *all* users/customers at a given website. In fact, the problem of scalability is a major concern with the previous evaluation methods. For example, with user testing, deciding on the adequate number of subjects to test for a representative picture of website usability problems is still in debate (Bevan, Barnum, Cockton, Nielsen, Spool and Wixon, 2003). Also, it is difficult for usability experts to be able to identify all usability problems that may arise for the wide variety different users that may be customers at the website due to bounded rationality. We are not arguing, however, that traditional testing, inquiry and inspection don't provide value. Instead, these methods have their own specific complementary strengths (especially before a site is launched) and should be employed in conjunction with the proposed method.

REFERENCES

1. Aigner, D. J. and Chu, S. F. (1968) On estimating the industry production function, *American Economic Review*, 58, 4, 826-839.
2. Banker, R. D., Charnes, A. and Cooper, W. W. (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science*, 30, 9, 1078-1092.
3. Banker, R. D., Datar, S., M. and Kemerer, C. F. (1991) A model to evaluate variables impacting the productivity of software maintenance projects, *Management Science*, 37, 1, 1-18.
4. Bevan, N., Barnum, C., Cockton, G., Nielsen, J., Spool, J. M. and Wixon, D. (2003) The "Magic Number 5": Is it enough for web testing? *Proceedings of the CHI 2003*, Ft. Lauderdale, FL.
5. Chase, R. B. (1978) Where does the customer fit in a service operation?" *Harvard Bus. Rev.* 56, 6, 138-139.
6. Ivory, M. Y. and Hearst, M. A. (2001) The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys*, 33, 4, 470-516.
7. Lovelock, C. H. and Young, R. F. (1979) Look to consumers to increase productivity. *Harvard Business Review*, 57, 3, 168-178.
8. Meuter, M. L., Ostrom, A. L., Roundtree, R. I. and Bitner, M. J. (2000) Self-service technologies: understanding customer satisfaction with technology based service encounters." *Journal of Marketing*, 64, 3, 50-64.
9. Nielsen, J. and Mack, R. L. (1994) *Usability Inspection Methods.* Wiley & Sons.
10. Schubert, P. and Selz, D. (1999) Web assessment: measuring the effectiveness of electronic commerce sites going beyond traditional marketing paradigms, *Proceedings of the 32nd HICSS*, Maui, HI.
11. Spool, J. M., Scanlon, T., Schroeder, W., Synder, C. and DeAngelo, T. (1999) *Web Site Usability: A Designer's Guide*. Morgan Kaufmann.
12. Zeithaml, V. A., Parasuraman, A. and Berry, L. L. (1990) *Delivering Quality Service: Balancing Customer Perceptions and Expectations*. Free Press