Categorization of Brazilian Internet Users and Its Impacts on the Use of Electronic Government Services

Marcelo Henrique De Araujo and Nicolau Reinhard

School of Economics, Business and Accounting, Business Department
University of Sao Paulo, Av. Prof. Luciano Gualberto, 908 05508-900, Sao Paulo, Brazil
{marcelo.haraujo,reinhard}@usp.br

Abstract. This paper explores the micro-data from the ICT Households Survey in order to categorize the Brazilian Internet users according to the diversity of activities undertaken by these users on-line and assess the propensity of these Internet user groups to use e-gov services. The Amartya Sen's Capability Approach was adopted as theoretical framework for its consideration of people's freedom to decide on their use of the available resources and their competencies for these decisions, leading to the use of e-government services. This paper uses a positivistic approach a descriptive and exploratory analysis of secondary data (micro-data) from the 2007, 2009 and 2011 editions of the ICT Household survey.

Keywords: Capability Approach, Electronic Government Services, Internet use patterns.

1   Introduction

In Brazil, e-government (e-gov) initiatives have become popular among the citizens due to the diversity of electronic services available to the population, such as electronic voting system, filing an income tax return report, scheduling medical appointments, enrollment of students in public schools, among others [1,2]. However, the success of these initiatives (e-gov) depends on the access and predisposition to use the ICT resources in the interaction between the government and the civil society.

Public policies for digital inclusion in the country encourage both the individual access (through tax reduction on computer equipment, promoting the purchase of computers and access to the Internet) and the collective access, which includes the offer and availability of Internet access in public access centers (for example, Public Telecenters and Cybercafés)[3, 4, 5].

Through its Center of Studies on Information and Communication Technologies (CETIC.br), the Brazilian Internet Steering Committee (CGI.br) conducts surveys in order to generate statistics about the use of Information and Communication Technologies (ICTs). For the purposes of this paper, the data used were taken from their ICT Household Survey, a survey conducted annually and whose data enable the generation of indicators about the availability and use of the Internet by the Brazilian population.

This research is based on the following objectives: (i) Categorize the Internet users based on their Internet use (for example, personal communications, entertainment, business applications, etc.) and (ii) Explore the relationships of their predisposition to use the e-gov services with the abovementioned categories.   The analysis is based on the ICT Household Survey micro-data for the years 2007, 2009 and 2011. The Capability Approach [6] was adopted as a theoretical framework, for its consideration of people's freedom to decide on their use of the available resources and their competencies for these decisions, leading to the use of e-government services.

Although the theme had already been explored by [7], the contribution of this paper is the characterization of the different groups of Internet users based on their usage of the Internet - and assessing the propensity to use the e-gov in each one of these groups.

The paper is structured as follows: Section 2 discusses the theoretical principles that guide this research. In Section 3, the reference model and the adopted methodological strategy are presented. Section 4 discusses the main results and findings of this research. Section 5 presents final remarks.

2   Literature Review

2.1   The Capability Approach

The Capability approach [6] was developed in order to offer an objective basis for interpersonal perceptions of welfare [8,9] thus overcoming limitations of the utilitarian approach, in which this type of comparison - essential for the welfare economy - would not be possible.

According to the Capability Approach [6], simply providing a commodity or resource does not directly imply a direct increase in the welfare of a community, due to the multiple manners this resource can be used (capability) and the results obtained by the effective use of this commodity (functioning). Therefore, the assessment of a resource (commodity) and its use depend on a set of cultural, social and knowledge factors, named conversion factors. Thus, from the access conditions to a certain commodity and the referred conversion factors, individuals recognize a set of resources, which will be assessed based on the possibility of access and perception of use. Consequently, reducing this initial set to its capability [6] [10], that is, a smaller set of resources that they will take into consideration to make their choices of use. Based on these capabilities, the individuals decide how they will use these resources in order to achieve their functioning, that is, the result of the effective use of these resources, which, in a last analysis, will lead to their utility, for example, exercise of rights, welfare [10].

2.2   ICT Household Survey

The ICT Household survey, performed annually since 2005 by the Brazilian Internet Steering Committee (CGI.br), aims at mapping the availability of ICT resources by the Brazilian population [11] (ICT Household Survey 2012).

The survey uses methodological standards developed by the OECD (Organization for Economic Co-operation and Development), Eurostat and by the Observatory for the Information Society in Latin America and the Caribbean (OSILAC), and Economic Commission for Latin America and the Caribbean (ECLAC), thus allowing international comparisons. The sampling plan is based on the national Census and Household Sample Survey.

3   Research Methodology and Reference Model

The paper's reference model, presented in Figure 1, is based on the Capability Approach [6].
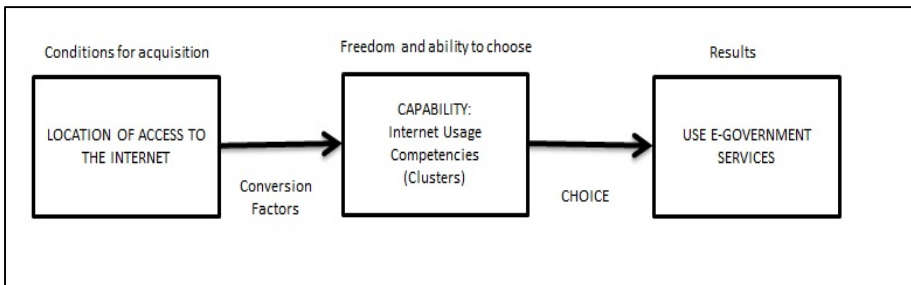
Fig. 1. Reference Model: Capability Approach (Source: adapted from [7], [9])

The objective of this model is to demonstrate the factors influencing the individuals' decisions to use electronic government services. The model consists of three main constructs: (i) conditions to obtain (or access) the resource; (ii) freedom and ability to choose (capabilities); (iii) results (functioning).

The survey's answer to the question "the most common or favorite access location to the internet" was used to operationalize the first construct. As a proxy to users' capabilities, the study constructs factors based on the survey's questions regarding the frequent usage of other Internet services: communication, leisure and entertainment activities, information search and on-line services. This study improves on previous analyses performed by [7], where capabilities were represented directly by the usage of e-mail, e-finance, e-commerce services. The last construct synthesizes the target phenomenon of this research, that is, the use of e-government services. For that end, this concept was operationalized through a binary primitive, indicating whether the Internet users made use of electronic government services in the last 12 months.

This study aims also to categorize Internet users, in order to understand the factors that drive or restrict the use of e-gov services. Differently from other models that are based on users' beliefs and attitudes (for example, perception of utility and facility), this paper is based on observed user behavior (actual usage of Internet services) and explores the mediation chain showed by the Internet users' choices and actions, attributes that are directly measured by the ICT Household Survey.

This research uses a positivistic approach a descriptive and exploratory analysis of secondary data (micro-data) from the 2007, 2009 and 2011 editions of the ICT Household Survey, Table 1 presents the survey's variables used in the paper.

Table 1. Variables

Variables for this study	ICT Survey questions and Categories
Social Class	AB (combined), C, and DE (combined);
Internet Access Location	Most common access location (Home, Work, At someone else's home, Telecenter, Cybercafe, Other place).
Use of Electronic Government Services	Yes/No
Internet Use for Communication	Several
Internet Use for Information Search and On-line Services	Several
Internet Use for Leisure and Entertainment	Several
Internet Use for Financial Services	Several
Internet Use for Education and Training	Several

The first variable presented in Table 1 is a construct derived from respondents' demographic attributes; due to the small sizes of classes A and D, the data were combined in three groups: Class AB (upper class); Class C (middle class); Class DE (lower class).

The variable "Use of Electronic Government Services" is a dichotomous measure that maps the Internet users that used any of the e-gov services in the last 12 months.

The other (binary) variables represent the different Internet uses, represented by a set of binary metric that explores several activities connected to communication (e.g. e-mail, instant messages, participation in networking websites), information search and on-line services (e.g. search for health information), leisure and entertainment (e.g. on-line games, virtual reality environments), financial services (consult and transactions via internet banking), and education and training activities (e.g. school activities, on-line courses, etc.).

Statistical analysis was performed through multivariate statistical techniques (Binary Factor Analysis, Logistic Regression and Clusters Analysis), using SPSS and Stata tools.

Binary Factor Analysis and the Cluster Analysis were used for the construction of user categories.

Due to the great amount of activities explored in each one of the Internet use patterns, the Binary Factor Analysis technique was applied in order to reduce the problem dimension. Cluster Analysis, using the factorial scores of the latent factors was then applied to identify homogenous Internet user groups regarding the multiple activities performed on the Internet. These clusters were used as proxies for the capabilities of the reference model (Figure 1).

Binary logistic regression technique was used to assess the relationship of the clusters with the users' preferred Internet access locations and also to establish the relationship of the clusters with the use of e-gov services.

Binary logistic regression is a technique that aims at studying the relation between the independent, metric and non-metric variables, and a binary dependent variable [12]. In order to perform the logistic regression, a stepwise procedure was used (with 95% of confidence level). Since the study is not focused on creating predictive models, it was decided to analyze only the odds ratio measures of the statistically significant independent variables.

4   Analysis and Discussion of Results

In accordance with the objectives of this study, the Internet users were categorized based on their uses of the Internet. The ICT Household Survey lists a number of categories of Internet uses: (i) communication (for example, e-mail, networking websites); (ii) information search and on-line services; (iii) leisure and entertainment (for example, on-line games, virtual reality); (iv) financial services (consult and transact via internet banking); (v) education and training activities (for example, on-line courses).

Due to the great number of variables in each one of these categories, Binary Factor Analysis statistical technique [13] was used in order to reduce these variables to a smaller number of latent factors. This analysis was applied to each one of the five abovementioned categories, for each year of the survey (2007, 2009 and 2011) Eigenvalues equal to or higher than 1 were used a criterion for factors retention. Furthermore, where necessary, a Varimax orthogonal rotation was applied, yielding one single significant factor for each category (except for the leisure and entertainment activities category with two factors). The analysis resulted in the following single factors: (a) communication; (b) information search and on-line services; (c) financial services; and (d) education and training.

For the leisure and entertainment activities, the first latent factor gathered all the variables related to the activities with high degree of interaction and collaboration among Internet users, such as participating in virtual reality environments, on-line games, etc. The second factor represented the non-interactive activities (for example, listening to on-line radios, reading news and magazines, etc.). Therefore, the first factor was called "interactive and collaborative activities," and the second one was called "non-interactive activities."

The standardized factor scores derived from the factor analysis were then used as input for the cluster analysis. In order to define the number of clusters, hierarchical clustering was initially applied, defining the "between groups", and the "squared Euclidean distance" as clustering and distance method, respectively. The results suggested the formation of three clusters. Then, the non-hierarchical clustering was applied through k-means algorithm, in order to create the three clusters, whose final positioning is showed in tables 2, 3 and 4.

Table 2. Final Cluster Centers (2007)

	Cluster		
	1	2	3
Communication	0.10	-1.08	0.48
Information Search and On-line Services	-0.28	-0.98	0.68
Interactive and Collaborative Activities (Leisure and Entertainment)	0.82	-0.92	-0.06
Non-interactive Activities (Leisure and Entertainment)	-0.59	-0.82	0.79
Financial Services	-0.07	-0.65	0.37
Training and Education	-0.16	-0.89	0.55

Table 2, shows the differences in profiles of the three user groups: cluster 2 has a negative value for all different Internet use patterns, indicating that this cluster gathers Internet users that make very little use of the analyzed Internet uses. Therefore, these Internet users are characterized as sporadic users. On the other hand, the data on cluster 3 present positive sign for almost all Internet use activities (except for leisure and entertainment interactive and collaborative activities), indicating that this group has a more Advanced profile when compared to the previous cluster. And, finally, cluster 1 has an intermediate profile (regarding the Internet use patterns) compared to the abovementioned clusters, with a higher leisure and entertainment component.

Table 3. Final Cluster Centers (2009)

	Cluster		
	1	2	3
Communication	-0.47	0.80	0.31
Information Search and On-line Services	-0.69	0.96	0.56
Interactive and Collaborative Activities (Leisure and Entertainment)	0.09	0.55	-0.44
Non-interactive Activities (Leisure and Entertainment)	-0.70	0.61	0.77
Financial Services	-0.32	1.72	-0.36
Training and Education	-0.55	0.83	0.42

The analysis of the 2009 data (Table 3) and 2011 (Table 4) shows a cluster formation similar to 2007 (only with clusters in different order).

The Cluster Analysis for each of the three years resulted in a set of three clusters with similar composition between the years. The observation of the factor loadings led the authors to name them as "Advanced use", "Intermediate use" and "Sporadic use".

Table 4. Final Cluster Centers (2011)

	Cluster		
	1	2	3
Communication	1.03	-0.57	0.31
Information Search and On-line Services	0.80	-0.69	0.70
Interactive and Collaborative Activities (Leisure and Entertainment)	1.22	-0.08	-0.78
Non-interactive Activities (Leisure and Entertainment)	0.33	-0.62	0.92
Financial Services	0.14	-0.44	0.70
Training and Education	0.74	-0.57	0.52

Table 5. Electronic government use by Internet user clusters (percentage)

	Advanced use			Intermediate use			Sporadic use		
	2007	2009	2011	2007	2009	2011	2007	2009	2011
Use of e-gov	56.4	93.7	82.7	83.7	80.3	89.4	88.8	54.3	50.7
Non-use of e-gov	43.6	6.3	17.3	16.3	19.7	10.6	11.2	45.7	49.3

Table 5 shows the percentage distribution of the e-gov services users in each one of the proposed Internet user clusters. The analysis of the table 5 data shows that the e-gov services are widely used by the Internet users with the Advanced and intermediate use profile. On the other hand, in the Internet user group of sporadic use, the use of e-gov services has decreased. Therefore, the Table 5 data shows that the proportion of Internet use the e-gov services is much higher among those with higher proficiency in the Internet use (Advanced and intermediate users groups).

Table 6. Internet Access Location by Internet user clusters (percentage)

	Advanced use			Intermediate use			Sporadic use		
	2007	2009	2011	2007	2009	2011	2007	2009	2011
Home	25.5	56.3	75.6	39.0	52.4	67.8	43.8	35.9	54.9
Work	13.6	28.6	10.6	25.1	18.7	22.8	25.8	12.3	13.3
School	4.0	0.6	1.0	3.3	3.1	0.9	2.7	2.7	1.3
Someone else's home	12.7	5.4	3.6	7.8	6.7	2.2	6.2	12.5	9.3
Telecenter	1.6	0.0	0.4	1.2	0.7	0.1	1.1	1.3	0.5
Cybercafé	40.9	8.8	7.2	22.9	18.1	4.9	20.2	34.5	18.5
Other Location	1.7	0.2	1.5	0.7	0.4	1.2	0.2	0.8	2.2

The Table 6 data shows the Internet access location distribution for each group. The data clearly show an increase of the Internet use preferably at home in all groups. However, the Internet use at home is much higher for the advanced use group (75.6% in 2011) and intermediate group (67.8% in 2011) than the sporadic use group (54.9% in 2011). Table 6 also shows that Cybercafés are, in percentage terms, the second most important access location for this (underprivileged) last group.

Table 7. Internet Access Location by Internet user clusters (percentage)

	Advanced use			**Intermediate use**			**Sporadic use**		
	2007	2009	2011	2007	2009	2011	2007	2009	2011
Class AB	29.5	64.0	52.5	45.6	45.7	64.3	53.0	32.0	36.4
Class C	54.8	33.5	44.0	47.4	47.7	33.5	38.0	56.1	54.2
Class DE	15.8	2.5	3.5	7.0	6.5	2.2	9.0	11.8	9.4

Table 7 presents the Internet user groups in relation to the socioeconomic status of their members, showing that the three groups are composed mainly of class AB members (upper class) and class C (middle class). The proportion of Internet users in class DE (lower class), although small in all clusters, is greater in the Internet user group of sporadic use.

4.1 Use of Electronic Government Services

The paper's objective was to assess the survey' s respondents decision to use e-government services, based on their preferred Internet access location, mediated by their patterns of Internet usage patterns (use groups).

For this purpose binary logistic regression was used, yielding the odds ratios, that measure how much, maintaining the other conditions constant, each one of the explanatory variables increases the chance of occurrence (or probability) of the studied phenomenon (e-government usage). Tables 8, 9 and 10 show the odds ratio values (of statistically significant attributes at 95% of reliability) of different Internet access locations according to the probability of the individual belonging to each one of the analyzed Internet groups.

Table 8. Contribution of the Internet access location to the Advanced Internet user group

	C			**DE**		
	2007	2009	2011	2007	2009	2011
At Home and Work	0.43	3.34	2.24	0.28		
At someone else's home						
Cybercafé						
Telecenter						

Table 9. Contribution of the Internet access location to the intermediate Internet user group

	C			DE		
	2007	**2009**	**2011**	**2007**	**2009**	**2011**
At Home and Work	1.55		2.25			5.50
At someone else's home		0.51				
Cybercafé		0.48				
Telecenter		0.24				

Table 10. Contribution of the Internet access location to the sporadic Internet user group

	C			DE		
	2007	**2009**	**2011**	**2007**	**2009**	**2011**
At Home and Work	2.11	0.33	0.36	2.70	0.43	0.35
At someone else's home						
Cybercafé						
Telecenter			4.71			

The data in Tables 8, 9 and 10 show that the preferred Internet use at home or at work (individual access) has a positive impact on the chance of belonging to the Internet user group of Advanced and intermediate use. On the other hand, Table 10 showed that Internet use at home or at work implies the reduction in the probability of belonging to the Internet sporadic use group. Therefore, it demonstrates that the individual access (home or work) encourages the development of capabilities.

Table 11. Contribution of different Internet user groups to the decision of using e-gov services (odds ratio)

	C			DE		
Groups (Clusters)	**2007**	**2009**	**2011**	**2007**	**2009**	**2011**
Advanced Use	0.17	6.80	2.93	0.31	7.15	4.23
Intermediate Use	3.71	2.50	4.60	3.20	3.37	10.93
Sporadic Use	4	0.24	0.19		0.23	0.12

The data in Table 11 shows that belonging to the proposed Internet user groups (capability measure) positively impacts the use of electronic government services. Odds ratio values were much higher among the members of the intermediate and advanced use groups, respectively. The propensity to use the e-gov services was shown to be greater among the users of the intermediate use group, compared to the advanced Internet users. An analysis of the data in Tables 3 and 4 emphasizes that the main difference between these two groups is the use intensity of interactive and collaborative activities of leisure and entertainment (for example, on-line games, virtual environment). Therefore, such competence could benefit the use of e-gov services in the intermediate Internet user group. On the other hand, the data showed odds ratio lower than 1 for Internet users of sporadic use, indicating that belonging to this group implies the reduction in the potential use of e-gov services. In other words, the data emphasizes that the lower competence Internet users tend not to use e-gov services.

5   Conclusions

The objective of this research was to categorize the Brazilian Internet users according to the diversity of activities undertaken by these users on-line and assess the propensity of these Internet user groups to use e-gov services. The results of this research showed the formation of three different groups of Internet users, called (i) Advanced use, (ii) Intermediate use, and (iii) Sporadic use. The first group gathers the Internet users with most competence in Internet use, that is, those who make use of all or almost all analyzed Internet use patterns. The second group consists of Internet users that perform a lower number of activities on the Internet, but that are characterized by the most extensive use of interactive and collaborative activities of leisure and entertainment, such as participating in virtual reality environments and on-line games. Finally, the last group comprises the Internet users with a more mundane (sporadic) use of the Internet, limiting themselves to the performance of few activities on the Internet (when compared to the previous groups).

Regarding the propensity to use e-gov services, the results showed that the Internet users belonging to the advanced and intermediate use groups were more likely to use the e-gov services. On the other hand, the Internet users of the sporadic use group were less likely to use the e-gov services. These results strengthen the notion that the development of Internet use competence has a positive impact on the propensity to use e-gov services. The results also demonstrated that the Internet user group of intermediate use presented a higher tendency to use e-gov services than the Internet user group of advanced use. Such tendency is possibly related to the extensive use of interactive and collaborative activities of leisure and entertainment performed by this user.

In summary, the paper showed the users' preferences for internet access location and the capabilities that favor their use of e-government services.

These findings indicate the appropriateness of enacted public policies that have addressed internet access problems through varied initiatives: at the regulatory and fiscal levels, connectivity rates, especially mobile have dropped significantly due to increased competition among service providers (the creation of a state-owned nationwide broadband network, mandatory users' free mobility between providers), reduction on taxes on entry-level computers, tablets and communication services. There are large direct public investments in connectivity to schools, free wi-fi access in public spaces, with satellite connection being provided to remote rural areas (Amazon Region). Schools with increased connectivity are then able to promote computer literacy. More affordable access to devices and internet connection contributed to the widespread use of communication services, especially social network services, on mobiles (mostly smartphones but also tablets). As a response, e-gov services are increasingly being offered to social class D and E citizens for mobile devices.

References

1. Avgerou, C., Ganzaroli, A., Poulymenakou, A., Reinhard, N.: Interpreting the trustworthiness of government mediated by Information and Communication Technology: Lessons from electronic voting in Brazil. Information Technology for Development 15, 133–148 (2009)

2. Reinhard, N., Moya, R.W.: The Adoption of Electronic Filing For Individual Income Tax Returns In Brazilian. In: International Conference on E-Business (ICEB), Proceedings of ICEB 2012, Beijing (2002)
3. Lemos, R., Martini, P.: LAN Houses: A new wave of Digital Inclusion in Brazil. Information Technologies & International Development 6, 31–35 (2010)
4. Mori, C.K., Assumpção, R.O.: Brazilian Digital Inclusion Public Policy: achievements and challenges. The Journal of Community Informatics 3(3), 1–6 (2007)
5. Madon, S., Reinhard, N., Roode, D., Walsham, G.: Digital inclusion projects in developing countries: Processes of institutionalization. Information Technology for Development 15, 319–340 (2009)
6. Sen, A.K.: Development as freedom. Oxford University Press, Oxford (1999)
7. Araujo, M.H., Reinhard, N.: Factors Influencing the Use of Electronic Government Services in Brazil. In: Joint Proceedings of Ongoing Research of IFIP EGOV 2013 and IFIP ePart, IFIP EGOV 2013, pp. 140–149. Koblenz (2013)
8. Prendergast, R.: The concept of freedom and its relation to economic development – a critical appreciation of the work of Amartya Sen. Cambridge Journal of Economics 29(6), 1145–1170 (2005)
9. Robeyns, I.: The Capability Approach: a theoretical survey. Journal of Human Development 6(1), 93–117 (2002)
10. Reinhard, N.: The Challenges of Universal Access: Models and Management – an invitation research. In: CGI. Survey the Use of Information and Communication Technologies in Brazil 2005-2009, pp. 189–198. CGI, São Paulo (2010)
11. CGI. Survey on the use of Information and Communication Technologies in Brazil: ICT Households and Enterprises 2012. Brazilian Internet Steering Committee, Sao Paulo (2013)
12. Hair, J.F., Black, W.C., Babin, J.B., Anderson, R.E., Tatham, R.L.: Multivariate Data Analysis. Pearson Prentice Hall, New Jersey (2006)
13. Bartholomew, D., Steele, F., Moustaki, I., Galbraith, J.: The Analysis and Interpretation of Multivariate Data for Social Scientists. Chapman and Hall/CRC Press, London (2002)