Government Information Quarterly 26 (2009) 193–202

Contents lists available at ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

Evaluating municipal websites: A methodological comparison of three think-aloud variants

Maaike J. van den Haak *, Menno D.T. de Jong, Peter Jan Schellens

Vrije Universiteit, Faculty of Arts, Netherlands
University of Twente, Faculty of Behavioural Sciences, Netherlands
Radboud University Nijmegen, Centre for Language Studies, Netherlands

ARTICLE INFO

Available online 29 October 2008

Keywords:
E-Government
Municipal websites
Usability testing
Think-aloud protocols

ABSTRACT

Usability methods have received relatively little methodological attention within the field of E-Government. This paper aims to address this gap by reporting on a usability test of the municipal website of Deventer (the Netherlands), carried out by means of three variants of the think-aloud method (concurrent/retrospective think-aloud protocols and constructive interaction). These three methods had proved successful in a previous evaluation of a different municipal website, yet we decided to replicate our study in order to investigate whether the three methods would reveal different results when applied to another municipal website with a different information architecture. The results of our study showed that, as in the previous municipal website evaluation, the three evaluation methods were largely comparable in terms of output. Nevertheless, we did find a number of differences between the present and previous municipal website evaluation regarding the workings of the three methods—differences that could be explained by the different information architectures of the municipal websites tested. This suggests that the three evaluation methods might indeed work differently depending on the nature of the website that is being evaluated, and calls for more research into the effect of task type on the validity of evaluation methods.

1. Introduction

Most municipalities and government institutions have their own space on the web, allowing their citizens to find information and, increasingly, to engage in all sorts of personalized E-Government services (Pieterson, Ebbers & Van Dijk, 2007). Citizens may, for instance, order copies of brochures, report changes in their address, or renew their vehicle registration, and the list of possibilities is likely to grow.

As the online activities of municipalities increased, so has the research output on this particular area (Heeks & Bailur, 2007). Studies have been published on municipal websites from countries as diverse as Norway (Halland & Saeth, 2004), Switzerland (Schedler & Summermatter, 2007), New Zealand (Cullen & Houghton, 2000), and Kenya (Kaaya, 2004). Some reports have even investigated the online activities of municipalities worldwide (Choudrie, Ghinea & Weerakkody, 2004; Holzer et al., 2006). The topics addressed in these studies vary widely from the accessibility of municipal websites (Potter, 2003; Shi, 2007; Jaeger, 2006) to textual content (Eschenfelder, 2004; Eschenfelder & Miller, 2007) to government–citizen interaction (Chadwick & May, 2003; Welch & Fulla, 2005; Griffin & Halpin, 2005).

Another major concern within the literature on E-Government is website usability. Numerous studies within E-Government journals, including GIQ, report on municipal website evaluations performed by means of usability methods like heuristics (Cullen & Houghton, 2000; Choudrie et al., 2004), scenario evaluation (Halland & Saeth, 2004; De Jong & Lentz, 2006b), interviews (Marcella, Baxter & Moore, 2003), and think-aloud protocols (Marcella et al., 2003; Jaeger, 2006). In describing the results of these website evaluations, however, most studies focus on the merits and drawbacks of the websites rather than on the working of the usability methods. As such, it seems that much is known about the ways in which municipal websites could be improved, but only little is known about the drawbacks and benefits of using a particular method for a particular municipal website. Since the validity of results revealed by usability methods is largely dependent on the validity of the methods themselves, more research into the exact working of usability methods within the field of E-Government seems highly desirable (see also Heeks and Bailur, 2007).

As a first step towards addressing this lack of attention for usability methodology within the E-Government area, we have recently evaluated a municipal website by means of three research methods: concurrent think-aloud protocols (CTA protocols), retrospective think-aloud protocols (RTA protocols), and constructive interaction (Van den Haak, De Jong & Schellens, 2007). The CTA protocols are perhaps the most common method of the three. They involve potential users who work with a particular test object while constantly verbalizing their

* Corresponding author. Vrije Universiteit, Faculty of Arts, Netherlands.
E-mail address: mj.vanden.haak@let.vu.nl (M.J. van den Haak).

0740-624X/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.giq.2007.11.003

thoughts. The RTA protocols are a variant of the CTA protocols, involving participants who silently work with a particular test object and then verbalize their thoughts afterwards, often on the basis of a video recording of their performance. Constructive interaction, finally, is a method which involves two participants rather than one. They work together and verbalize their thoughts by interacting with each other. The practical value of the three methods is that researchers cannot only *observe* participants while working with a particular test object, but can also *listen* to them, primarily with a view to either uncovering people's mental processes or, in the case of usability evaluation, detecting user problems.

All three methods have long been accepted as useful research methods and have been applied in various fields including psychology (e.g. Taylor & Dionne, 2000), nursing (e.g. Funkesson, Anbäcken & Ek, 2007), and reading and writing research (Schellings, Aarnoutse & Van Leeuwe, 2006; Wong, 2005). Particularly the CTA and RTA protocols have been extensively discussed in research contributions, with Ericsson and Simon (1993) as standard theoretical framework. Within the context of usability testing, Nielsen (1993) is an often-cited practical handbook. Van Someren, Barnard and Sandberg (1994) also offer practical advice for the entire process of collecting and analyzing think-aloud protocols, as do Rubin (1994) and Dumas and Redish (1999) for the broader context of usability testing. Current research into the think-aloud methods has focused on, for instance, the effect of personality traits on people's ability to think aloud (Schneider & Reichl, 2006) and the notion of reactivity. This notion refers to the fact that when asked to perform tasks *and* think aloud at the same time, participants might perform these tasks differently and might experience difficulty in verbalizing, as a result of their combined cognitive workload being too high. As such, reactivity might affect the working of the concurrent think-aloud method. The extent to which this happens has been and continues to be a much investigated topic (Russo et al., 1989; Ericsson and Simon, 1993; Van den Haak, De Jong & Schellens, 2003; Van den Haak, De Jong & Schellens, 2004; Alavi, 2005).

The results of our study involving the three usability methods (Van den Haak et al., 2007) suggested that each of the methods was equally useful for evaluating municipal websites: the CTA protocols, RTA protocols, and constructive interaction were all comparable in terms of quantity and relevance of problems detected. Each of the methods was equally capable of detecting the main output of the other two methods. We did, however, find some differences between the three methods. The participants in the RTA method, for instance, experienced more observable problems and were less successful in completing their tasks than the participants in the CI method, while the CI participants performed their tasks faster than the CTA participants.

While these findings are interesting, we felt that it would be good to conduct a second municipal website evaluation using the same three usability methods but a different municipal website. The main reason for this replication of our previous study is that even within one country there are many municipal websites (De Jong and Lentz, 2006a) and these may vary greatly in terms of information architecture. The municipal website of our previous evaluation (Van den Haak et al., 2007), for instance, contains large pieces of information on every web page, and a list of links from which users, once they have read the information on the page, can make deliberate selections. Such an information architecture involves substantial

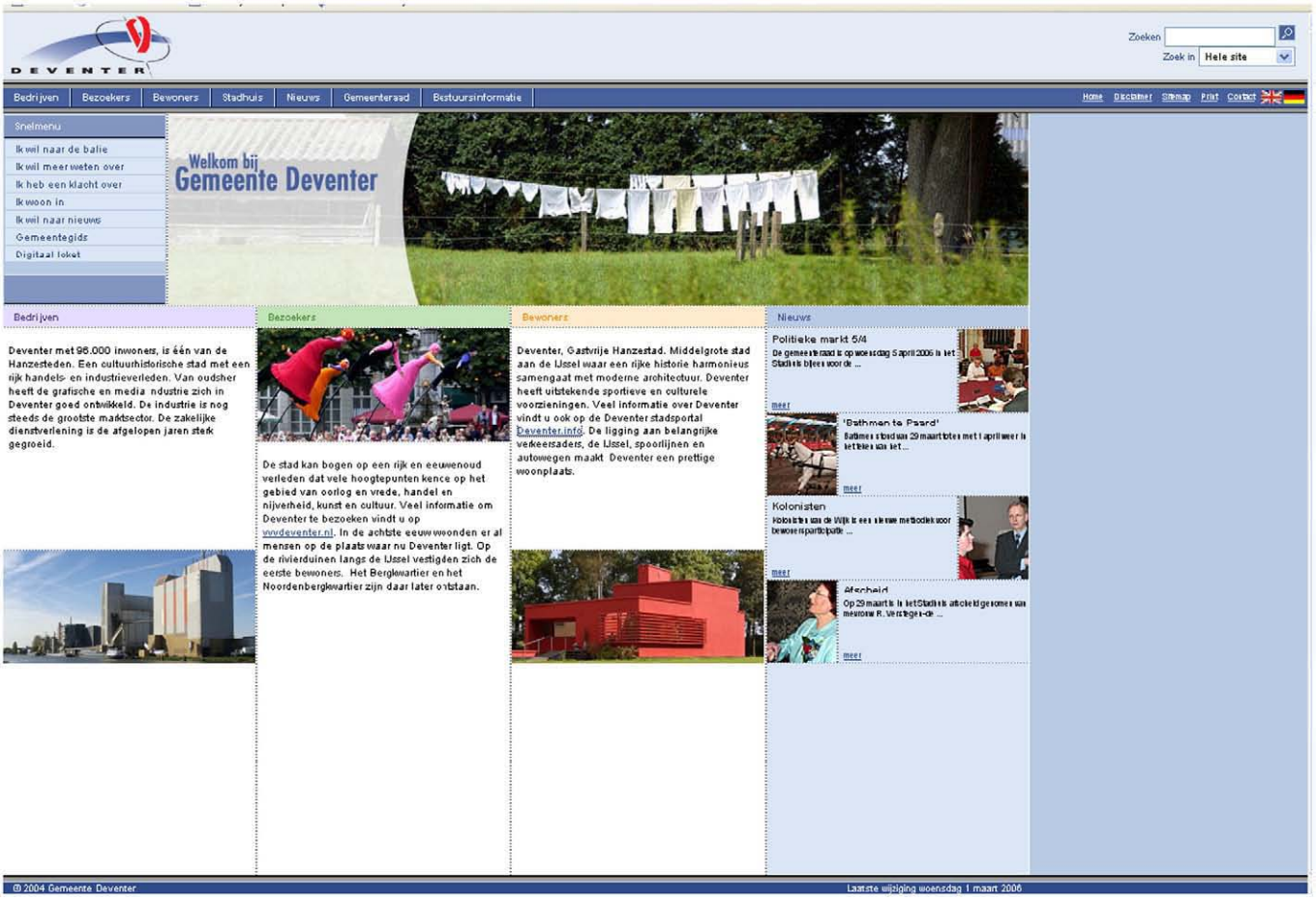

Fig. 1. Home page of the Deventer website.