Investigating Respondent Multitasking in Web Surveys using Paradata

Anže Sendelbah (anze.sendelbah@fdv.uni-lj.si) ab

Vasja Vehovar (vasja.vehovar@fdv.uni-lj.si) a

Ana Slavec (ana.slavec@fdv.uni-lj.si) a

Andraž Petrovčič (andraz.petrovcic@fdv.uni-lj.si) a

a Faculty of Social Sciences, University of Ljubljana

Kardeljeva ploščad 5

1000 Ljubljana

Slovenia

b Corresponding author

Investigating respondent multitasking in web surveys using paradata

ABSTRACT

Computers play an important role in everyday multitasking. Within this context, we focus on respondent multitasking (RM) in web surveys. RM occurs when users engage in other activities while responding to a web survey questionnaire. The conceptual framework is built on existing literature on multitasking, integrating knowledge from both cognitive psychology and survey methodology. Our main contribution is a new approach for measuring RM in web surveys, which involves an innovative use of the different types of paradata defined as non-reactive electronic tracks concerning respondents' process of answering the web questionnaire. In addition to using questionnaire page completion time as a measure of RM, we introduce 'focus-out' events that indicate when respondents have left the window containing the web questionnaire (e.g., to chat, email, browse) and then returned. The approach was tested in an empirical study using a web survey on a student sample (n=267). The results indicate that 60% of respondents have multitasked at least once. In addition, they reveal that item nonresponse as an indicator of response quality is associated with RM, while non-differentiation is not. Although this study confirms that a paradata-based approach is a feasible means of measuring RM, future research on this topic is warranted.

1. Introduction

The concept of multitasking refers to sequential or concurrent combinations of activities. In this paper, we focus on *respondent multitasking* (RM) where responding to a questionnaire is the *primary activity. Secondary activities* or distractions are any other activities performed by respondents at any moment between the start and the end of this primary activity.

Despite concerns expressed in previous survey methodology research (e.g., Holbrook, Green, & Krosnick, 2003; Lavrakas, Shuttles, Steeh, & Fienberg, 2007; Lynn & Kaminska, 2012) that respondents' secondary activities can affect their response process and the quality of the responses given, the problem has rarely been addressed in the literature. This is particularly true for the technical and methodological aspects of measuring RM, which have typically been investigated using self-reports: by asking respondents at the end of a questionnaire whether or not they were engaged in any secondary activities while responding (e.g., Kennedy, 2010; Lavrakas, Tompson, Benford, & Fleury, 2010; Zwarun & Hall, 2014). However, researchers have shown that self-reports about multitasking behaviour (e.g., Iqbal & Horvitz, 2007; Lottridge, Marschner, Wang, Romanovsky, & Nass, 2012) are not always reliable as people tend to over- or under-estimate the amount of time spent on activities. For example, some respondents might not report any secondary activities due to social desirability bias or they might simply fail to account for all instances of secondary activities. Different authors have highlighted the need for more studies that use electronic tracking to measure media and/or multitasking behaviour in order to compare and validate self-reported and non-reactive RM data (e.g., Greenberg et al., 2005; Möller, Kranz, Schmid, Roalter, & Diewald, 2013; Wallis, 2010).

Therefore, the present study aims to fill the gap in the literature on RM in web surveys by both conceptually and empirically exploring the potential of *paradata* for measuring RM. Following the approach of Couper (2000b), the concept of paradata refers to non-reactive tracks concerning the process of surveying, such as interviewer observations, contact attempt records, and electronic traces. Within the context of web surveys, we predominantly talk about *direct paradata*, which is generated by respondent interaction with the survey instrument and is automatically collected at the respondent level by the survey platform (Callegaro, Manfreda, & Vehovar, 2015). In the context of research on RM, paradata can provide more reliable results than self-reports. Such data can also decrease the respondents' burden as there is no need to include additional questions about RM at the end of questionnaires.

This study introduces a theoretically informed elaboration of new RM indicators that can describe and indicate possible RM behaviours. Specifically, based on the integration of an overview of existing interdisciplinary research on multitasking, media multitasking, RM, and response quality (RQ), we have developed a methodological framework for investigating RM in web surveys using paradata. Drawing on this framework as well as the available paradata, we have developed RM indicators related to two different types of navigational paradata events: long response times and switches away from the browser window or tab that contains the web questionnaire. Although such a methodological framework represents a first attempt at an integrative conceptualisation of RM in web surveys, its development might be relevant on at least two levels. On one hand, we believe research on this multifaceted and complex behaviour requires a thorough theoretical discussion and conceptualisation. On the other hand, a broader understanding of RM and RQ will allow us to properly discuss the limitations of our paradata-based approach.

In addition, the empirical part of this paper presents a proof of concept study on the feasibility of using a paradata-based approach for measuring RM in web surveys by employing it on a survey of incoming and outgoing exchange students at the University of Ljubljana during the 2012/2013 academic year. While the main focus of the empirical study was finding support for proof of concept, we also use our survey data to illustrate the descriptive characteristics related to the occurrence of the multitasking. Accordingly, in addition to investigating the relationship of RM with RQ, we also report the prevalence rates of RM in our web survey.

The structure of this paper is as follows. In Section 2, we first present the concept of multitasking, mainly based on threaded cognition theory. This is followed by an overview of empirical evidence on the prevalence of multitasking on personal computers in general and of RM in particular. We then expand on the RM taxonomy. In Section 3, we provide an overview of the theoretical and empirical literature dealing with the relationship between RM and RQ. Next, we introduce our paradata-based approach in Section 4. After presenting the research questions (Section 5), methods (Section 6) and results of the empirical study (Section 7), we provide a discussion of our findings along with identifying the advantages and limitations of our approach (Section 8).

2. Background on multitasking, media multitasking and respondent multitasking

2.1. Multitasking

Multitasking research has a long history in the cognitive sciences (Meyer & Kieras, 1997). The literature presents different, often incompatible or even contrasting, definitions of this phenomenon, as well as various theories on how our cognitive system copes with multitasking situations. Recently, several authors have proposed different cognitive architectures to explain our ability to multitask, including Executive-Process/Interactive Control (EPIC) (Meyer & Kieras, 1997), Multiple Resource Model (Wickens, 2008), and Threaded Cognition (Salvucci & Taatgen, 2011). Following the studies of Wang, David, Srivastava, Powers, Brady, D'Angelo and Moreland (2012), van Cauwenberge, Schaap and van Roy (2014) and Courage, Bahktiar, Fitzpatrick, Kenny and Brandeau (2015), it is suggested that the latter presents the most applicable and holistic conceptualisation of multitasking and the role of our cognitive system, as the other theories are usually only concerned with our cognitive ability to multitask.

Salvucci and Taatgen (2011) define multitasking on a continuum where one extreme represents *concurrent multitasking*, where tasks are performed simultaneously or with very short interruptions, while the other extreme represents *sequential multitasking*, where the time between *task switches* can be expressed in seconds, minutes or even hours. A core component of the threaded cognition theory is the adoption of ACT-R (Adaptive Control of Thought-Rational) cognitive architecture (Anderson, 2007) to explain the role and limitations of the human processing resources involved in multitasking behaviour. ACT-R architecture describes the interaction between different cognitive resources and their limitations in terms of processing different complex processes.

Salvucci and Taatgen (2011) use this architecture to model different activities and to predict conflicts that can arise during multitasking. For example, each resource can process a limited amount of different activities. If two activities require the same cognitive resource, they cannot access it at the same time and so one of the activities is temporarily suspended. Such conflicts may cause delays in execution, while the recall of suspended activities requires additional cognitive processing. Moreover, longer suspension periods and/or more complex activities increase the amount of cognitive resources required when switching between activities.

Other authors generally agree that our capability to multitask is limited by the capabilities of our cognitive system (e.g., Meyer & Kieras, 1997; Wickens, 2008). They also agree that multitasking is typically associated with longer completion

times and an inferior performance quality in goal-oriented activities. In addition to numerous experiments on multitasking in cognitive laboratories (Meyer & Kieras, 1997), this has also been confirmed in applied research. For example, studies have reported that an increased tendency to multitask with various media content in everyday life is associated with worse GPA scores or other measures of academic success (e.g., Fried, 2008; Junco & Cotten, 2012; Kirschner & Karpinski, 2010). However, empirical findings also show that multitasking does not always impair performance or that it can even improve performance under certain conditions (e.g., Adler & Benbunan-Fich, 2012; Ie, Haller, Langer, & Courvoisier, 2012; Tran, Carrillo, & Subrahmanyam, 2013). Moreover, different types of activities can have different implications for performance (e.g., Kennedy, 2010; Wood et al., 2012).

In addition to considering whether activities require the same cognitive resources, Salvucci and Taatgen (2011) discuss other factors that influence our ability to multitask. Increased knowledge of specific activities can mitigate the negative effects in multitasking situations that involve such activities. In the survey research context, for instance, a person who is used to responding to questionnaires might be more capable of combining such an activity with other activities. Another difference between individuals can arise from differences in cognitive system capabilities due to age, visual acuity, motor capabilities and/or other reasons (Salvucci & Taatgen, 2011).

Relevant theoretical insight into multitasking in the modern media environment is also offered by communication theories (see Yeykelis, Cummings, & Reeves (2014) for an overview). Compared to cognitive science literature, this research stream outlines the role of motivational systems in the different stages of multitasking behaviour. For example, Wang and Tchernev (2012) have applied the uses and gratification theory to explain how implicit emotional needs and subsequent emotional gratifications (such as feeling entertained or relaxed) drive habitual multitasking behaviour. Another important insight is offered by the Limited Capacity Model of Motivated Mediated Message Processing (LC4MP), which states that the allocation of cognitive resources partially depends on an individual's motivation (Lang, Shin, Bradley, Wang, Lee, & Potter, 2005).

To conclude, threaded cognition and other relevant theories result in several important lessons for research on RM. First, we should be especially careful with secondary activities that require the same cognitive resources as the primary activity. Such secondary activities are likely to increase response times and the (perceived) difficulty of a questionnaire. Moreover, we should be aware of

individual differences that can influence RM, including cognitive abilities and motivation.

2.2. Prevalence of computer-related media multitasking

Multitasking has recently become a topic of interest in many disciplines, including communication studies, human factors, organisational behaviour, information science, and human-computer interaction studies (Spink, Cole, & Waller, 2008). Arguably, the most prominent strain of multitasking research is focused on media multitasking (MM), a form of multitasking that involves various media activities such as consumption of media content, creation of media content and communication via electronic devices.

Since this study is focused on web surveys on personal computers, we are especially interested in the role of such devices in everyday MM behaviour. In this context, by personal computers (PC) we understand either desktop computers or laptops. Correspondingly, PC tasks are tasks completed on a PC that require the active attention of a person. The PC plays a key role in everyday MM, especially in combination with other PC tasks, listening to music, and/or activities on other media devices such as television or smartphones (e.g., Carrier, Cheever, Rosen, Benitez, & Chang, 2009; Google, 2012; Rideout, Foehr, & Roberts, 2010; Voorveld & van der Goot, 2013).

Several studies have focused solely on measuring sequential combinations of PC activities using non-reactive electronic tracking methods. Research suggests that an average PC-task switch happens every three minutes or even more often (Judd, 2013; Kraushaar & Novak, 2010; Mark, Voida, & Cardello, 2012; Mark, Wang, & Niiya, 2014; Yeykelis, Cummings, & Reeves, 2014; Zhang, Sun, Chai, & Aghajan, 2015).

In general, research indicates that PC-driven MM is most prevalent among the younger generations (e.g., Carrier et al., 2009; Voorveld & van der Goot, 2013; Zhang et al., 2015). For instance, almost two-thirds of US 8- to 18-year-olds multitask other media while using the computer and/or do multiple things at the same time on the computer (Foehr, 2006; Rideout et al., 2010). In other words, within this population PC tasks are performed exclusively only about a third of the time, while about a third of the time they are combined with other PC tasks and are combined with non-PC tasks the remaining third of the time (Foehr, 2006). Besides age, other factors are discussed in relation to the prevalence and forms of MM behaviour. Gender, race, different individual traits, ownership and access to technology, user interface design, and situational aspects are often underlined as potentially important drivers of MM (e.g., Jeong et al., 2005; Meng & McDonald,

2009). However, empirical research on MM factors is still relatively scarce and that which does exist provides inconclusive findings.

The presented studies on MM suggest that web surveys could face strong competition for a respondent's attention in what Petrič, Petrovčič, and Vehovar (2011) refer to as the complex media environment. We may thus reasonably assume that for many respondents, participating in a web survey could be just another media activity that is commonly combined with other activities.

2.3. Prevalence of respondent multitasking

RM has been most thoroughly investigated in telephone surveys, either via respondent self-reports (Kennedy, 2010; Lavrakas et al., 2010; Lynn & Kaminska, 2013) or via interviewer observations (Pew Research Center, 2006).

Lavrakas et al. (2010) analysed nine different mobile phone surveys of varying lengths (n=2,103; a random digit dial sample of mobile phone numbers in the USA). In total, 51% of respondents reported engaging in some form of RM and 16% of respondents were assessed by the main author of the study as being at least moderately distracted. On the other hand, interviewers at the Pew Research Center (2006) indirectly observed that 20% of respondents in a phone sample engaged in RM, although only 8% were marked as somewhat or very distracted. Their 11-minute survey was also conducted on a random digit sample representative of the US population (n=1,503). The other referenced studies on telephone surveys do not report the total prevalence rates of RM.

Overall, PC activities, video games and television are the most commonly mentioned secondary activities in these telephone studies, followed by eating, drinking or preparing a meal, driving (in) a car, working and household chores (Kennedy, 2010; Lavrakas et al., 2010; also see Pew Research Center, 2006; Lynn & Kaminska, 2013).

In contrast with telephone surveys, less is known about RM in web surveys. In fact, the very recent paper by Zwarun and Hall (2014) is, to the best of our knowledge, the only published study on RM in the context of web surveys. Zwarun and Hall's (2014) results show that almost 30% of the 5,853 respondents recruited from an international online panel self-reported engaging in some form of RM during a 30-minute web questionnaire. The most common secondary activities were related to the use of (other) electronic devices, followed by background music and activities that required one to physically leave the PC. Further, the authors also suggested that younger respondents were significantly more likely to engage in secondary activities. For example, 52% of respondents in the youngest age group (18- to 24-year-olds) engaged in some form of secondary activity on electronic

devices, while the share in the oldest age group (65-years-old and above) was only 17%. Comparable trends can be observed for other forms of secondary activity.

2.4. Taxonomy of respondent multitasking in web surveys

To study RM, Zwarun and Hall (2014, p. 239) introduced a taxonomy that differentiates between three types of web survey RM: concurrent secondary *environmental activities* (which do not require respondents to "task-switch by shifting their primary attention from the survey", but have "the potential to take up cognitive processing"); sequential secondary *non-media activities* (which require task-switching to non-media activities); and sequential secondary *media activities* (which require task-switching to media activities).

More specifically, environmental activities in Zwarun and Hall's (2014) study include background music, background video and background conversation. Non-media activities include leaving the computer and direct conversation. Media activities include leaving the browser screen the survey is on to perform another task on the computer, tablet, or phone; talking on the phone or participating in a voice chat; reading or writing a text, social network update, e-mail, etc.; and hearing a noise from a notification.
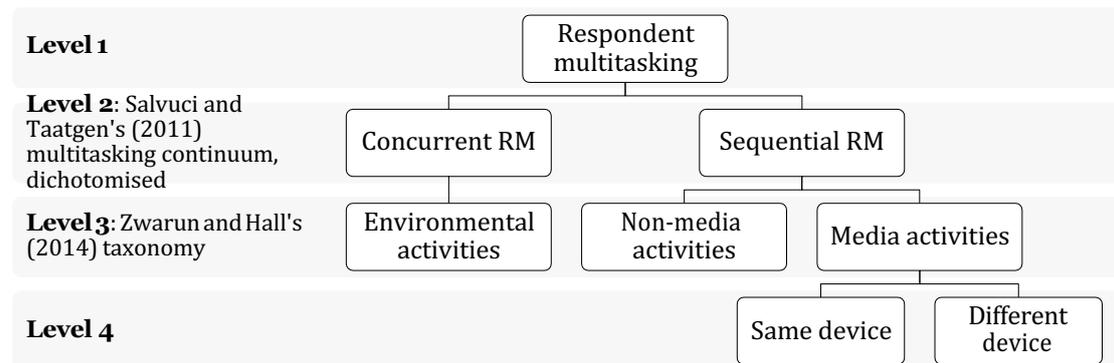

Figure 1. Taxonomy of RM in web surveys.

Drawing on a conceptual review of multitasking and media multitasking, Zwarun and Hall's (2014) RM taxonomy could be further detailed into two directions, thereby providing a more exhaustive classification of RM. Following Salvucci and Taatgen's (2011) multitasking continuum, environmental activities can be associated with concurrent multitasking, while non-media and media activities belong to the sequential multitasking dimension (see Figure 1). In addition, media activities can be divided into two subtypes: *same-device media activities* refer to media activities using the same device that is used for responding to a questionnaire, while *different device media activities* refer to media activities on other devices.

3. Conceptualising the relationship between respondent multitasking and response quality

3.1. Conceptual links between respondent multitasking and response quality

In this study, response quality (RQ) refers to those aspects of data quality that are influenced by respondents' individual differences in perception, understanding, and interaction with a questionnaire. Specifically, RQ encompasses break-offs (drop-outs), item nonresponse, and measurement errors occurring due to suboptimal response processes. Ganassali (2008) overviews the most commonly researched errors in this context, which include "the filling-up rate" (i.e., the percentage of non-substantive answers given by respondents), incomplete responses (e.g., brief, insufficient answers to open-ended questions), non-differentiation of responses (e.g., giving identical or nearly identical responses on rating scales for all items in matrix questions), and similar.

Drawing on the most widely adopted information processing model of the response process (Tourangeau, Rips, & Rasinski, 2000) [1], which includes four stages (i.e., comprehension, retrieval, judgment and response) each containing several cognitive processes, scholars have explored and elaborated the possible connections between RM and RQ (e.g., Holbrook et al., 2003; Lavrakas et al., 2010; Zwarun & Hall, 2014). Kennedy (2010) proposes that multitasking can actually distract all four stages of the response process: 1) comprehension might suffer due to the inability to completely process the entire question; 2) respondents might employ less demanding retrieval strategies; 3) at the judgment stage, respondents might not fully contemplate all possible considerations; and 4) in the final stage, respondents might fail to optimally map their judgment onto the required response format due to their involvement in secondary activities. A similar relationship between RM and the response process has been proposed by Lynn and Kaminska (2013), who point out that responding to a questionnaire is a cognitively demanding task and that an insufficient response process can negatively influence RQ.

However, we should bear in mind that this relationship between RM and the response process is also influenced by different underlying factors. The questionnaire-dependent factors of RQ relate to formal and content-related questionnaire characteristics, such as the general structure and length of the questionnaire, question wording, interactivity, and response format (Ganassali,

---

[1] For an informative overview of other models, see Berzelak (2014).

2008). The respondent-dependent factors of RQ include the respondent's motivation, ability and opportunity (e.g., Bosnjak & Tuten, 2001; Couper, 2000a; Krosnick, 1991; Tourangeau et al., 2000). Moreover, as suggested in Section 2.1, individual differences in terms of ability and the motivation to cope with multitasking situations are also an important factor, as well as the nature of the secondary activities (i.e., if they require the same cognitive resources as the response process).

We stress that these theoretical connections do not suggest that increased RM will always be associated with lower RQ. Many respondents might produce low- or high-quality responses regardless of whether they engage in RM or not. Moreover, in some cases secondary activities might contribute to better RQ (e.g., playing some relaxing music in the background).

3.2. Empirical evidence of the link between respondent multitasking and response quality

Lavrakas et al. (2010) and Kennedy (2010) analysed the relationship between RM and RQ in telephone surveys across various indicators. Lavrakas et al. (2010) reported that RM is related to more non-substantive answers (i.e., answering with "Don't know" or refusals). However, they found no such associations for overall item nonresponse, non-differentiation, and refusals to answer specific sensitive questions. In Kennedy's (2010) study, respondents who were eating or drinking during the survey had problems with question comprehension, but there were no associations with other RQ indicators and/or other secondary activities.

Zwarun and Hall (2014) did not directly investigate the relationship between RM and RQ. However, they did ask respondents how distracted they felt while responding to the web questionnaire. The self-reported measures of RM were positively associated with measures of self-reported distraction. The respondents' age was highlighted as an important factor in the strength of the association between the RM and distraction: while the correlation was significant for all age groups, there was a curvilinear trend where the correlation was the strongest for 35- to 44-year-olds.

In addition, Zwarun and Hall (2014) ascertained an important relationship between different RM (sub)types and self-reported distraction. All environmental and non-media forms of RM, except for background music, were significantly positively associated with the level of self-reported distraction. Yet by far the strongest coefficient was associated with the range of media activities measured as a count of how many different forms of media activity respondents engaged in. However, it is interesting to note that the frequencies of specific forms of media

activity were mostly significantly negatively associated with self-reported distractions. The authors provide several explanations for these seemingly contradictory results. In addition to social desirability bias, the authors suggested that respondents may have engaged in multiple media activities in one larger block and thus felt less distracted. Further, media activities might also be "automatic for […] a large portion of the sample" (Zwarun & Hall, 2014, p. 243), which relates to the mitigating influence of the increased knowledge of primary and/or secondary activities that we mentioned at the end of Section 2.1.

In conclusion, we turn our attention to one specific form of media activity described in Zwarun and Hall's (2014, p. 239) study as: "[leaving] the browser screen the survey is on to do another task on my computer, tablet, or phone." This form of activity is interesting for two reasons. First, besides the relatively passive activity of hearing a notification when receiving "an instant message, voice mail, text or email", this is the most common form of media activity in this study. Second, only the frequency of this form of media activity is positively associated with self-reported distraction. These two findings indicate that such tasks could be particularly important aspects of RM in web surveys in terms of the connections with RQ.

4. Measuring respondent multitasking using paradata

4.1. Non-reactive measurements of sequential multitasking on PCs: Focus-out events

In Section 2.2, we proved an overview of the substantive results from studies that used automatically generated logfiles to explore sequential multitasking on PCs. Researchers typically define individual activities based on opened application windows or URL addresses if the application is a web browser, while a switch among activities is defined with regard to changing the focus of the currently active window, tab or URL address to another instance of activity (e.g., Judd, 2013; Mark et al., 2014; Zhang et al., 2015). Some researchers also categorise activities into dichotomous categories such as productive/distractive (Kraushaar & Novak, 2010), work/non-work (Lottridge et al., 2012) or work/entertainment (Yeykelis et al., 2014).

In the web survey research context, some survey tools such as 1KA (Vehovar et al., 2012) provide access to paradata files containing information on when respondents have lost and regained focus on the browser window or tab that contains the web questionnaire. We refer to such an event as a *focus-out event*. In

the context of the RM taxonomy presented in Section 2.4, focus-out events correspond to same-device media activities.

To a certain extent, focus-out events are similar to the switching measures summarised in the first paragraph of this section, particularly those that divide activities into dichotomous categories (e.g., work/non-work). However, these paradata events do not offer any insight into the nature of the secondary activities. In other words, focus-out events only tell us when respondents switched to another activity on their PC – we do not know exactly what they have been doing.

Therefore, we should also be careful not to equate a single focus-out event with a single standalone secondary activity. As an example, one focus-out event could indicate two actual same-device activities (e.g., a respondent switches from the web questionnaire window to write an e-mail and check the weather forecast and then returns back to the questionnaire) or a combination of same-device activities with other types of activities (e.g., a respondent switches from the web questionnaire window to write an e-mail, goes for a walk, and then returns to the questionnaire). On the other hand, multiple focus-out events could be associated with a single activity (e.g., having an IM conversation with a friend throughout the survey process).

Moreover, when we equate focus-out events with same-device media activities, we should note that false positives and false negatives can occur. An example of a false negative is having an audio conversation over Skype while Skype's program window is minimised. An example of a false positive would be automatically switching to screensaver mode while the respondent leaves the PC to do some non-media activities (while this is also RM, it is not a media activity on the same device).

4.2. Non-reactive measurements of long response times in web surveys: Time-out events

Web survey methodologists have used paradata to explore how much time respondents need to complete a question (e.g., Couper & Kreuter, 2013; Heerwegh, 2005), a page (e.g., Gutierrez, Wells, Rao, & Kurzynski, 2011; Kaczmirek, 2008) or a whole questionnaire (e.g., Heerwegh, 2004; Malhotra, 2008) for a variety of reasons. Based on a literature review by Olson and Parkhurst (2013), long response times have been interpreted as a consequence of different respondent-dependent factors (e.g., ambivalent attitudes or greater engagement in the survey) or questionnaire-dependent factors (e.g., poor question wording or usability problems).

We have already suggested that multitasking typically prolongs the time needed to complete tasks (Section 2.1). However, we are not aware of any study in the field of survey methodology that would interpret longer response times as a consequence of RM. We can identify some similarity in studies where authors have tried to identify which respondents have taken longer pauses while responding to a web questionnaire. These studies provide different criteria and different cut-off points for identifying such respondents. In particular, Heerwegh (2005) looked for respondents who spent more than half of their total survey time on a single question (30% of respondents exceeded this cut-off point). Beckers, Siegers, and Kuntz (2011) searched for respondents who spent more than 30 seconds per item, four minutes per page, or more than seven minutes to complete a five-page questionnaire (in total, 23% of respondents exceeded at least one of these cut-off points). Stieger and Reips (2010, p. 1493) defined longer inactivity as when a respondent engages in "no action of any kind for at least 5 minutes" (3.6% of respondents exceeded this cut-off point).

Even though these studies were not focused on RM and do not provide any substantive analysis of the respondents who exceeded the cut-off points, they can provide valuable implications for understanding the importance of cut-off points. Accordingly, we will analyse page response times and define the cut-off point for each page. When a respondent exceeds this cut-off point, we will refer to this event as a *time-out event*. In the context of the RM taxonomy, time-out events correspond to all types of secondary activities so long as they prolong the response time beyond the cut-off point.

In terms of RM measurement, time-out events have similar limitations to focus-out events: we have no insight into the nature of the secondary activities and one time-out event does not necessarily equate to one secondary activity. Moreover, time-out events seem to be more prone to false positives due to non-RM related factors that can also extend the response time past the cut-off point.

Another important difference between focus-out events and time-out events is that the former have no theoretical upper limit (there is no theoretical boundary on how many focus-out events a single respondent can record during a survey session), while the number of possible time-out events per respondent is bounded by the number of questionnaire pages they have visited.

Time-out events also have one pragmatic advantage over focus-out events. While we can observe the former in most web survey tools, the collection of the latter is more technically complicated and so less common.

Figure 2 shows the coverage of focus-out and time-out events. The possibility of false positives and false negatives is intentionally omitted in order to clearly

visualise the relationship between the RM taxonomy, page response times, and both types of events. As shown in the figure, none of the events can capture non-same-device activities that do not exceed the cut-off point. Focus-out events and time-out events overlap when a switch to one or more same-device media activities (which is captured as a single focus-out event) results in a time that is sufficiently prolonged to exceed the cut-off point.

| | ← **Page response time** → | |
Type of secondary activities		**Cut-off point**
Environmental	/	time-out
Non-media activity	/	time-out
Different device media	/	time-out
Same-device media	focus-out	focus-out / time-out

Figure 2. Coverage of focus-out events and time-out events with regard to respondent multitasking taxonomy and page response times.

5. Research questions

Based on the theoretical considerations discussed in Sections 2, 3 and 4, we formulate the following research questions:

RQ1: What is the prevalence of RM in our web survey?

Prevalence in its most basic form can be assessed as a share of respondents with at least one detected RM event. However, the share of such respondents might be too rudimentary. Following the same approach as the Pew Research Center (2006) and Lavrakas et al. (2010), we will try to assess the share of respondents heavily engaged in RM.

RQ2: What is the relationship between the RM and RQ indicators in our web survey?

Our survey allows us to check two RQ indicators: item nonresponse and non-differentiation in matrix questions.

RQ3: What is the relationship between focus-out events and time-out events in our web survey?

As shown in Figure 2, the relationship between focus-out events and time-out events depends on the ratio between such overlapping events and other detected RM events.

It should be noted that RQ3 has in a reciprocal relationship with RQ1 and RQ2. The results related to RQ3 will be important for understanding the findings with regard to RQ1 and RQ2 and *vice versa*. In particular, we would like to assess whether the time-out events have a stronger association with the RQ indicators than the more technically demanding focus-out events.

6. Methods

6.1. The web survey and sample

Empirical research was carried out on a web survey of students from the University of Ljubljana who studied in foreign universities as part of a student exchange programme (i.e., outgoing students) and of foreign students who studied at the University of Ljubljana as part of a student exchange programme (i.e., incoming students) during the 2012/2013 academic year.

The web questionnaire[2] was active for 14 days, from the 17th to 19th week of 2014. The main topics of the questionnaire were students' satisfaction with the student exchange programme and their experiences of living and studying abroad. The questionnaire consisted of 13 pages, containing a total of 28 questions.

It should be acknowledged that the Slovenian students responded to the questionnaire in the Slovenian language – the mother tongue for almost all of them. On the other hand, the foreign students received the questionnaire in the English language, although for almost all of them this was not their first language. Moreover, we should also note that the respondents were randomly assigned to two versions of the same questionnaire. The control group received a questionnaire where the question wordings followed the guideline concerning the use of familiar terms. The question wordings for the experimental group were deliberately not in accordance with this guideline, instead involving the

---

[2] The questionnaire was created using the 1KA survey tool. A preview of the English version of the questionnaire is available at:
https://www.1ka.si/UniLJ&preview=on&disableif=1&disablealert=1&displayifs=1&displayvariables=1

substitution of certain words with more complicated and less well-known synonyms.[3,4]

An invitation was sent to 2,064 potential participants in total, and some 533 of them clicked on the link in the invitation. Of these, 313 respondents reached the final page of the questionnaire and answered at least one question. Among these respondents, the median completion time was ten minutes, not counting the time spent on the first and last pages, which do not contain any questions but rather an introduction and thank-you text, respectively.

In this study, we focus on only those respondents who answered at least half of the questions and who reached the final page. Moreover, to avoid the unwanted influence of outliers, we exclude respondents who needed more than five times the median absolute deviation (MAD)[5] from the median completion time. Finally, we also exclude respondents who did not use a PC to respond to the survey. While the analysis of multitasking by respondents who do not fit these criteria would be an interesting research topic, it is beyond the scope of this study. Thus, in total we included 267 respondents in our analysis.

6.2. Measurement of respondent multitasking and research quality indicators

6.2.1. Focus-out indicators

We parsed the paradata file to identify all the focus-out events. We excluded focus-out events on the introductory and last pages of the questionnaire. Further, we are only interested in focus-out events longer than ten seconds. There are three reasons for this lower limit. First, our tool recorded many one to two second focus-out events that occur when respondents navigate from one questionnaire page to another. Since these short events do not indicate actual RM, we omitted them from further analysis. Second, drawing on the threaded cognition theory, we assume that short switches to other same-device activities do not typically indicate substantially cognitively demanding same-device activities. In the absence of prior research on this topic, we should mention that the exact value of this lower cut-off point was chosen arbitrarily.

---

[3] For more information on this experiment, see Slavec and Vehovar (2014).

[4] While we will briefly address these two manipulations in our results, we are planning to fully investigate the relationship between the questionnaire-dependent factors and RM in our future work (Section 8.3).

[5] The median absolute deviation is a robust measure of variability that is particularly appropriate for describing the variability in univariate non-normal distributions. Compared to standard deviation, it is less sensitive to outliers. MAD is computed as the median of the absolute values of residuals from the data's median.

We derive two basic indicators from the focus-out events: *focus-out count*, which represents the count of all focus-out events per respondent, and *focus-out length*, which is the sum of the lengths of all focus-out events per respondent in seconds.

6.2.2. Time-out indicators

We calculated the median and the MAD of the completion times for each page of the questionnaire (separately for each of the four combinations of the language and question wording groups). We then standardised the values by subtracting the median page completion time from the respondents' page completion time and dividing by the MAD. Similar to the focus-out indicators, the introductory and last pages of the questionnaire were excluded from the development of the time-out indicators.

We defined a time-out event as any instance where a respondent's standardised page completion time was higher than two. Again, as in the case of focus-out indicators, this cut-off point was chosen arbitrarily. Moreover, we also computed the length of such events by calculating the time respondents spent beyond this cut-off point per page.

Thus, we derive two basic indicators from the time-out events: *time-out count*, which refers to the count of all time-out events per respondent, and *time-out length*, which is the sum of the length of all time-out events per respondent in seconds.

6.2.3. Item nonresponse indicator

The item nonresponse variable is computed by counting all question items that were shown to respondents but that remained unanswered by respondents in the questionnaire.

6.2.4. Non-differentiation indicator

We calculated the standardised probability of the differentiation index[6] for three different matrix questions that appeared in the questionnaire. Following the questionnaire preview link provided in Footnote 2, these three questions appear on the 7th (questions Q49 and Q50) and 8th pages (Q52) for the experimental group. For the control group, these questions appear on the 15th (Q71 and Q72) and 16th (Q74) pages of the preview. All of the questions use a five-point scale. The first question asks respondents to evaluate how much information they received on five different aspects of their host country. The second question asks how much information they received from eight different sources. The third

---

[6] For more details on this index, see Berzelak (2014).

question asks respondents to evaluate their satisfaction with six different aspects of their student exchange.

7. Results

7.1. Relationship between focus-out count and time-out count indicators

As shown by the median, variance and skewness values in Table 1, all four RM indicators are strongly skewed to the right. Figure 3 shows the distribution of both count indicators, and that the distribution of both length indicators is similar, but even more skewed.

For almost half (49%) of the respondents, the focus-out count and time-out count share the same values. For almost one-third (32%) of respondents, more time-outs than focus-outs were recorded, while the opposite is true for the remaining 19% of respondents. Moreover, an important difference between these two indicators is that 42% of the respondents who registered at least one time-out event did not register any focus-out events. On the other hand, 25% of respondents who registered at least one focus-out event had no registered time-out events. The results of an exact version of McNemar's Chi-square test (Fay, 2010) show that there are significant differences between these two binary outcomes ($p < 0.01$).

Table 1
Descriptive statistics for respondent multitasking indicators.

	Median	Mean	Variance	Skewness	Kurtosis	Min	Max
Focus-out count[a]	0.00	0.89	2.58	2.80	9.33	0.00	10.00
Time-out count	1.00	1.19	2.88	2.07	5.10	0.00	9.00
Focus-out length[b]	0.00	85.87	78788.72	4.90	25.78	0.00	2064.00
Time-out length	3.17	133.02	126026.53	4.19	19.20	0.00	2385.73

[a] The unit of both count indicators is the frequency of events per respondent.
[b] The unit of both length indicators is the sum of the lengths of events per respondent (in seconds).

Nevertheless, the two count indicators are moderately correlated, as shown by Kendall's tau ($r_\tau = 0.38$, $p < 0.01$). In addition, we should mention that focus-out length and time-out length are moderately correlated as well ($r_\tau = 0.42$, $p < 0.01$).
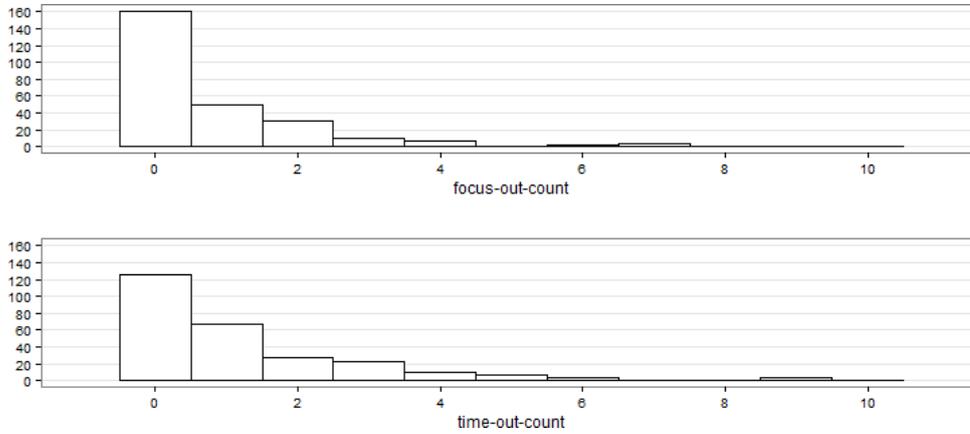

Figure 3. Histogram of respondent multitasking focus-out and time-out count indicators.

7.2. Prevalence of respondent multitasking

Some information about the prevalence of RM can be deduced from Table 1. However, due to the distribution of the RM indicators, it is more appropriate to report prevalence in terms of shares: 40% of respondents had at least one focus-out event, while the same is true for 53% of respondents in the case of the time-out events. If we were to combine these two indicators, 62% of the respondents recorded at least one RM event.

Due to the lack of qualitative information, it is hard to evaluate which respondents were extensively involved in secondary activities. For instance, one criterion for this could be the number of RM events per recorded respondent: 10% of respondents recorded three or more focus-out events and the same is true for 17% of them in the case of time-out events. If we take a look at the length indicators, 19% of the respondents spent more than one minute looking at other content on their PC according to the focus-out length indicator, and if we increase this limit to two minutes this holds true for 11%. For about 30% of the respondents, the duration of the time-out length was over one minute, and for 20% it was over two minutes.

7.3. Relationship between RM and RQ indicators

Similar to the RM indicators, the item nonresponse indicator also has a unimodal distribution that is highly skewed to the right (M=1.30, Me=0, SD=3.44, skewness=3.90, maximum=23).

Kendall's tau (r_τ) does not show a significant correlation of item nonresponse with any of the RM indicators. However, given our elaboration in Section 3.1 and the shape of the distribution of the item nonresponse indicator, we chose to explore this relationship with a negative binomial regression. In this context, we ran four separate procedures[7] with item nonresponse as the dependent variable and one of the four RM indicators as the independent variable. We initially added the questionnaire language and question wording complexity group membership as control variables in all four of the procedures, but since none of them was a significant predictor we dropped them out.

As shown in Table 2, focus-out count is a significant predictor, while time-out count is only marginally significant. Neither of the two length indicators is a significant predictor of item nonresponse. However, we must stress that none of the RM indicators has a tangible value for predicting item nonresponse according to McFadden's pseudo R^2. Instead, these procedures merely serve to show which of the RM indicators is the most strongly associated with item nonresponse.

Table 2
Results of negative binomial regression procedures for the relationship between item nonresponse and respondent multitasking.

Independent variable	β_1	SE(β_1)	p	McFadden's pseudo R2
Focus-out count	0.19	0.09	0.04	< 0.04
Focus-out length	< -0.0004	0.00	0.51	< 0.04
Time-out count	0.15	0.09	0.08	< 0.04
Time-out length	< -0.0004	0.00	0.75	< 0.04

None of the non-differentiation indexes are significantly associated with any of the RM indicators. After using various analytical approaches, we must conclude that there is no relationship between RM and the non-differentiation indexes.

---

[7] All negative binomial generalised linear models were run with R package MASS (Venables & Ripley, 2002) using the function glm.nb(). In addition, the pscl package (Jackman, 2015) was used to calculate McFadden's Pseudo-R^2 using the function p2R().

8. Discussion

8.1. Respondent multitasking: Prevalence and response quality

With reference to RQ1, the results show that 62% of the respondents recorded at least one RM event (i.e., at least one focus- and/or time-out event), whereas in terms of RQ1 and RQ3 it was found that 40% of respondents recorded at least one focus-out event and that 52% recorded at least one time-out event. Although the reported prevalence of RM in our study could be perceived as relatively high, it is comparable with the 18-24 and 25-34 age groups in Zwarun and Hall's (2014) study, as well as with the results in Lavrakas et al.'s (2010) study on telephone surveys, and general findings on the prevalence of MM in everyday life (Section 2.2). In addition, our study seems to confirm previous research by showing that the share of respondents who seemed to be extensively engaged in secondary activities is relatively low.

In relation to RQ2, we report a significant yet weak positive association between the focus-out count and item nonresponse, while focus-out count and time-out count are only marginally significant associated (this difference is also important for RQ3). The second indicator of RQ – non-differentiation – is not associated with any of the RM indicators. Similar to the studies by Lavrakas et al. (2010) and Kennedy (2010) on telephone surveys, we did not find a clear empirical confirmation of the relationship between RM and RQ. In addition, it is noteworthy that none of the RM length indicators are associated with RQ. These findings seem to agree with the suggestion by Zwarun and Hall (2014) that large blocks of secondary activities do not substantially distract respondents.

As the results were obtained as part of a proof of concept study, in what follows we elaborate on the methodological issues and related limitations that could have influenced the substantive results and conceptual conclusions.

8.2. The paradata-based approach

When providing a rationale for RQ3, we have already admitted to being particularly interested in exploring whether time-out events could be sufficient for measuring RM, since these events are easier to measure across a variety of web survey tools and devices used for responding. The results showed that our approach recorded more time-out event than focus-out events and, while the frequencies and durations of the focus-out events and time-out events were moderately correlated, there were also important differences between them. These differences make sense from a conceptual and a substantial point of view, but also give different results regarding RQ2. If we had only been using time-out

events in our analysis, we would not have found a significant relationship between RM and item nonresponse.

There are two possible, although not necessarily mutually exclusive, explanations for such a finding. On one hand, relatively short same-device media activities, which are captured via focus-out events but cannot be measured with time-out events, could be a crucial form of RM that has a large influence on item nonresponse. This explanation is in accordance with Zwarun and Hall's (2014, p. 239) study, which singled out the frequency of "[leaving] the browser screen the survey is on to do another task […]" activity as the only RM form with a significant relationship with self-reported distraction among all the frequency-based predictors of media activities. On the other hand, the results could be different if we develop time-out indicators with different criteria. This second explanation – arbitrary decisions in the design of RM indicators – is part of a more general limitation of our approach. Besides the criteria for time-out events, this refers to the ten-second lower cut-off point for focus-out events. We also see two important reasons to introduce an upper limit for both types of events in the context of analysing the relationship between RM and RQ. First, because of the implication that larger blocks of activities might not typically lead to greater distraction and, second, this would also partially mitigate the problem of outliers and so make our RM indicators more appropriate for more standard analytical procedures.

To control for this issue, we ran additional analyses with slightly changed values for the selected criteria (e.g., a five-second shorter limit for focus-out events, different cut-off points for the time-out events, and the exclusion of RM events lasting more than ten minutes). The results were not significantly different with reference to all three research questions. However, a more thorough and systematic approach to defining these criteria might lead to a more valid and unified methodology for measuring multitasking behaviour in the future.

Moreover, the herein presented paradata-based approach does not provide an insight into secondary activities. Accordingly, we were not able to measure the range of different secondary activities that respondents were involved in, especially if we consider the report by Zwarun and Hall (2014) concerning the range of activities being in a strong relationship with self-reported distraction. Related to this problem is the limited coverage of RM events. As discussed in Section 4, with the paradata-based approach it is impossible to completely avoid false negatives and false positives. This issue is connected with two other limitations: different cut-off points might decrease the possibility for both types of errors, while additional qualitative information could be used to verify the validity of the paradata indicators.

However, it should be remembered that the main motivation for researching RM is not to merely describe the prevalence of this phenomenon, but rather to explore its relationship with RQ. In this context, it is not so crucial to measure all the possible occurrences of RM, but only those types of RM that are most strongly associated with RQ. With this in mind, we believe that the paradata-based approach has the potential to sufficiently measure the most important RM types, at least for web surveys on PCs. Based on both the literature review and our empirical results, focus-out events are especially beneficial: combining different PC activities is one of the most common multitasking scenarios in everyday life; secondary activities on PCs are some of the most common forms of RM in telephone and web surveys; and the combination of two PC activities typically develops in a situation where both activities require the same cognitive resources, which leads to conflict according to the threaded cognition theory. This additional cognitive effort can interact with factors that influence RQ.

Nevertheless, restricting our analysis to only focus-out events could entail a too narrow focus since some other prevalent (sub)types of RM might also be important. Therefore, although based on our results time-out events in the current simplistic form cannot substitute focus-out events in terms of explaining the relationship between RM and RQ, it is also important to take time-out events in future iterations of our procedure into account in the possible future work outlined in Section 8.3.

From the future perspective, the generalisation of findings could also be facilitated by testing our approach on different types of survey questions. Notably, in this study all three matrix questions asked about respondents' past behaviour. Perhaps different results would have been obtained if the survey had contained matrix questions concerning opinions on current matters or a respondent's personality traits, which are commonly asked in such form.

8.3. General limitations and future work

In addition to the identified limitations of the paradata-based approach, this study was also constrained by three general limitations. First, while our sample is probability-based, it is relatively small and based on a specific group of students. The literature shows that adolescents and young adults not only tend to engage in MM more frequently (Section 2.2), but that they are also generally more capable of multitasking than older populations (Section 2.1). Further, familiarity with tasks also influences the quality of results in multitasking situations. Therefore, it is safe to assume that students are commonly engaged in information retrieval tasks that are akin to responding to a survey (e.g., doing homework assignments, taking tests, etc.). For example, Zwarun and Hall (2014) show that younger age

groups self-reported that they engage in more RM, although the correlation between RM and distraction was stronger for older respondents. However, although comparing our results with those of other studies is useful to show that the paradata-based approach produces meaningful results, our sample is too specific to directly match our findings with other studies.

The specific sample in our study is also related to the second general limitation – a lack of control variables. There are several respondent-dependent factors that could influence RM and/or RQ such as age, socioeconomic status, personality traits, individual capability to cope with specific RM situations and so on, which were not available for consideration in this empirical study. Moreover, it would also be insightful to check for different questionnaire-dependent factors. As suggested in Section **Error! Reference source not found.**, we will investigate the relationship between RM and the different levels of cognitive demand determined by differences in question wording. In addition, we are also planning to explore other related factors, including questionnaire design, length and topic.

Finally, in this study a relatively limited set of RQ indicators was tested, namely item nonresponse and non-differentiation. In future work, it would be highly valuable to include more RQ indicators such as those studied by Kennedy (2010) and by Lavrakas et al. (2010).

Future work could also address issues directly related to the paradata-based approach. Namely, as part of its further development, researchers could explore what approaches are the most valid for computing and analysing focus-out and especially time-out events. Second, scholars could look into the possibility of combining focus-out and time-out indicators within a single RM indicator. Third, research could also focus on temporal trends, for example to determine whether RM is more prevalent towards the end of the web survey or for certain pages or questions, and whether the exclusion of (highly) distracted respondents (an approach used in this study) from the sample affects the substantial findings of surveys. In an attempt to address these issues, we have already looked into both of these questions in our study, although we did not find any significant differences. Lastly, future research designs could compare self-reports on RM with the herein introduced paradata-based RM indicators. Although our questionnaire contained questions about RM, the triangulation of both data types was beyond the scope of this paper. Nevertheless, we strongly believe that an integrative approach to combining paradata and self-reports, if carefully implemented and analysed, might offer a comprehensive insight into RM.

9. Conclusion

In this paper, we presented a proof of concept study that investigated RM in web surveys using paradata. On the theoretical level, we provided an elaboration of the conceptual relationships between factors of RM and RQ, as well as a refinement of the RM taxonomy in web surveys suggested by Zwarun and Hall (2014). We then introduced a new methodological paradata-based approach for measuring RM in web surveys and reported the results.

The literature review showed not only that MM is a common everyday behaviour, but also that the human cognitive system has limited resources available when dealing with multiple activities. In addition, it demonstrated that RM, RQ, and their underlying factors form an intertwined network of relationships. While measuring multifaceted behaviour such as RM is a challenging task, measuring complex cognitive states such as motivation and ability is even more demanding. As such, RM has the potential to be a tangible indicator of parts of these complex relationships.

Based on these theoretical inputs, different types of secondary activities and available paradata events, we developed a new methodological approach for measuring RM in web surveys. We tested this new approach in a study and, in addition, provided new substantive empirical insights into this under-researched field.

Despite the discussed limitations such as less qualitative insights into secondary activities and the lack of capacity to detect non-same-device secondary activities that do not prolong page response times past the cut-off points, which all warrant future research, the current results show that a paradata-based approach is a feasible means of investigating respondent multitasking that is worthy of further consideration and refinement. Specifically, we showed that it is possible to measure several important aspects of RM via paradata. Moreover, when compared to self-reporting data collection strategies, this non-reactive approach has some additional advantages, including more detailed quantitative information about RM, no respondent bias, and no additional respondent burden.

While the empirical findings from our and others' studies do not establish RM as a significant threat to RQ, it is important to continue to study this topic for several reasons. The first stems from theoretical and empirical elaborations of the influence of multitasking on performance in goal-oriented activities. Secondly, there is a general lack of research on RM in different survey contexts. Finally, the overall rising trends of media multitasking in everyday life – especially among younger generations – have to be considered. When the so-called millennial

generation starts to represent an increasingly larger proportion of the general population and when media devices become even more ubiquitous and capable of performing different demanding processes at once, we can anticipate that the importance of RM will also grow for other age groups. Thus, it is important that a survey methodology is ready to meet this challenge and we believe that the presented paradata-based approach could be helpful in this regard. Last but not least, we hope that the herein presented methodological findings could also be applied to the broader context of research on the prevalence of media multitasking.

References

Adler, R. F., & Benbunan-Fich, R. (2012). Juggling on a high wire: Multitasking effects on performance. *International Journal of Human-Computer Studies*, *70*(2), 156–168. http://doi.org/10.1016/j.ijhcs.2011.10.003

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.

Beckers, T., Siegers, P., & Kuntz, A. (2011). *Speeders in online value research: Cross-checking results of fast and slow respondents in two separate samples answering the 40 item "Portrait value questionnaire"*. Presented at the General Online Research Conference, Düsseldorf, Germany. Retrieved from http://conftool.gor.de/conftool11/index.php?page=downloadPaper&file name=Beckers-Speeders_in_Online_Value_Research-172.pdf&form_id=172&form_index=&form_version=final

Berzelak, J. (2014). *Mode effects in web survey* (Unpublished doctoral dissertation). University of Ljubljana. Retrieved from http://dk.fdv.uni-lj.si/doktorska_dela/pdfs/dr_berzelak-jernej.pdf

Berzelak, N., Lozar Manfreda, K., Slavec, A., & Vehovar, V. (2012). *Advanced paradata in web surveys: What can they tell about the response process?* Presented at the annual conference of the American Association for Public Opinion Research, Orlando, FL.

Bosnjak, M., & Tuten, T. L. (2001). Classifying response behaviors in web-based surveys. *Journal of Computer Mediated Communication*, *6*(3). http://doi.org/10.1111/j.1083-6101.2001.tb00124.x

Callegaro, M., Lozar Manfreda, K., & Vehovar, V. (2015). *Web survey methodology*. London, United Kingdom: SAGE Publications Ltd.

Carrier, L. M., Cheever, N. A., Rosen, L. D., Benitez, S., & Chang, J. (2009). Multitasking across generations: Multitasking choices and difficulty ratings in three generations of Americans. *Computers in Human Behavior*, *25*(2), 483–489. http://doi.org/10.1016/j.chb.2008.10.012

Couper, M. P. (2000a). Usability evaluation of computer-assisted survey instruments. *Social Science Computer Review*, *18*(4), 384–396. http://doi.org/10.1177/089443930001800402

Couper, M. P. (2000b). Web surveys: A review of issues and approaches. *The Public Opinion Quarterly*, *64*(4), 464–494. http://doi.org/10.1086/318641

Couper, M. P., & Kreuter, F. (2013). Using paradata to explore item level response times in surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(1), 271–286.

Courage, M. L., Bakhtiar, A., Fitzpatrick, C., Kenny, S., & Brandeau, K. (2015). Growing up multitasking: The costs and benefits for cognitive

development. *Developmental Review*, *35*, 5–41.
http://doi.org/10.1016/j.dr.2014.12.002

Fay, M. P. (2010). Two-sided exact tests and matching confidence intervals for discrete data. *R Journal*, *2*(1), 53–58. Retrieved from http://journal.r-project.org/archive/2010-1/RJournal_2010-1_Fay.pdf

Foehr, U. G. (2006). *Media multitasking among American youth: Prevalence, predictors and pairings.* Henry J. Kaiser Family Foundation. Retrieved from http://kaiserfamilyfoundation.files.wordpress.com/2013/01/7592.pdf

Fried, C. B. (2008). In-class laptop use and its effects on student learning. *Computers & Education*, *50*(3), 906–914. http://doi.org/10.1016/j.compedu.2006.09.006

Ganassali, S. (2008). The influence of the design of web survey questionnaires on the quality of responses. *Survey Research Methods*, *2*(1), 21–32.

Google. (2012). *The new multi-screen world: Understanding cross-platform consumer behavior.* Retrieved from https://ssl.gstatic.com/think/docs/the-new-multi-screen-world-study_research-studies.pdf

Greenberg, B. S., Eastin, M. S., Skalski, P., Cooper, L., Levy, M., & Lachlan, K. (2005). Comparing survey and diary measures of internet and traditional media use. *Communication Reports*, *18*(1-2), 1–8. http://doi.org/10.1080/08934210500084164

Gutierrez, C., Wells, T., Rao, K., & Kurzynski, D. (2011). *Catch them when you can: Speeders and their role in online data quality.* Presented at the annual

meeting of the Midwest Association for Public Opinion Research, Chicago, IL. Retrieved from http://www.mapor.org/confdocs/absandpaps/2011/2011_slides/6d2Gutierrez.pdf

Heerwegh, D. (2004). *Using progress indicators in web surveys*. Presented at the annual conference of the American Association for Public Opinion Research, Phoenix, AZ. Retrieved from https://perswww.kuleuven.be/~u0034437/public/Files/Heerwegh%20Using%20Progress%20Indicators.pdf

Heerwegh, D. (2005). *Web surveys: Explaining and reducing unit nonresponse, item nonresponse and partial nonresponse* (Unpublished doctoral dissertation). University of Leuven. Retrieved from http://perswww.kuleuven.be/~u0034437/public/Files/Heerwegh_PhD_Diss_distribute.pdf

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, *67*(1), 79–125. http://doi.org/10.1086/346010

Ie, A., Haller, C. S., Langer, E. J., & Courvoisier, D. S. (2012). Mindful multitasking: The relationship between mindful flexibility and media multitasking. *Computers in Human Behavior*, *28*(4), 1526–1532. http://doi.org/10.1016/j.chb.2012.03.022

Iqbal, S. T., & Horvitz, E. (2007). Disruption and recovery of computing tasks: Field study, analysis, and directions. In M. B. Rosson & D. Gilmore (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 677–686). http://doi.org/10.1145/1240624.1240730

Jackman, S. (2015). *pscl: Classes and methods for R developed in the Political Science Computational Laboratory, Stanford University* (Unpublished doctoral dissertation). Stanford, California: Department of Political Science, Stanford University. Retrieved from http://pscl.stanford.edu/

Jeong, S. H., Zhang, W., Davis, E., Jordan, A., Fishbein, M., Hennessy, M., & Martin, S. (2005). *Multiple media use and multitasking with media among high school and college students*. Presented at the General Online Research Conference, Zurich, Switzerland. Retrieved from http://www.academia.edu/227042/Multiple_media_use_and_multitasking_with_media_among_high_school_and_college_students_A_Diary_method

Judd, T. (2013). Making sense of multitasking: Key behaviours. *Computers & Education*, *63*, 358–367. http://doi.org/10.1016/j.compedu.2012.12.017

Junco, R., & Cotten, S. R. (2012). No A 4 U: The relationship between multitasking and academic performance. *Computers & Education*, *59*(2), 505–514. http://doi.org/10.1016/j.compedu.2011.12.023

Kaczmirek, L. (2008). *Human-survey interaction: Usability and nonresponse in online surveys* (Unpublished doctoral dissertation). University of Mannheim. Retrieved from https://ub-madoc.bib.uni-mannheim.de/2150/

Kennedy, C. K. (2010). *Nonresponse and measurement error in mobile phone surveys* (Unpublished doctoral dissertation). University of Michigan. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/75977

Kirschner, P. A., & Karpinski, A. C. (2010). Facebook® and academic performance. *Computers in Human Behavior*, *26*(6), 1237–1245. http://doi.org/10.1016/j.chb.2010.03.024

Kraushaar, J. M., & Novak, D. C. (2010). Examining the effects of student multitasking with laptops during the lecture. *Journal of Information Systems Education*, *21*(2), 241–251.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236. http://doi.org/10.1002/acp.2350050305

Lang, A., Shin, M., Bradley, S. D., Wang, Z., Lee, S., & Potter, D. (2005). Wait! Don't turn that dial! More excitement to come! The effects of story length and production pacing in local television news on channel changing behavior and information processing in a free choice environment. *Journal of Broadcasting & Electronic Media*, *49*(1), 3–22.

Lavrakas, P. J., Shuttles, C. D., Steeh, C., & Fienberg, H. (2007). The state of surveying cell phone numbers in the United States 2007 and beyond. *Public Opinion Quarterly*, *71*(5), 840–854. http://doi.org/10.1093/poq/nfm054

Lavrakas, P. J., Tompson, T. N., Benford, R., & Fleury, C. (2010). *Investigating data quality in cell phone surveying*. Presented at the annual conference of the American Association for Public Opinion Research, Chicago, IL.

Lottridge, D., Marschner, E., Wang, E., Romanovsky, M., & Nass, C. (2012). Browser design impacts multitasking. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 1957–1961). http://doi.org/10.1177/1071181312561289

Lynn, P., & Kaminska, O. (2011). *The impact of mobile phones on survey measurement error* (ISER Working Paper Series No. 2011-07). Retrieved from http://hdl.handle.net/10419/65992

Lynn, P., & Kaminska, O. (2012). Factors affecting measurement error in mobile phone interviews. In S. Häder, M. Häder, & M. Kühne (Eds.), *Telephone Surveys in Europe* (pp. 211–228). http://doi.org/10.1007/978-3-642-25411-6_14

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly*, *72*(5), 914–934. http://doi.org/10.1093/poq/nfn050

Mark, G., Voida, S., & Cardello, A. (2012). A pace not dictated by electrons: An empirical study of work without email. In M. B. Rosson & D. Gilmore (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 555–564). http://doi.org/10.1145/2207676.2207754

Mark, G., Wang, Y., & Niiya, M. (2014). Stress and multitasking in everyday college life: An empirical study of online activity. In M. B. Rosson & D. Gilmore (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 41–50). http://doi.org/10.1145/2556288.2557361

Meng, J., & McDonald, D. (2009). *Predictors and impacts of TV multitasking and simultaneous multiple media use*. Presented at the annual meeting of the International Communication Association, Chicago, IL. Retrieved from http://citation.allacademic.com//meta/p_mla_apa_research_citation/2/9/8/1/0/pages298106/p298106-1.php

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part I. Basic mechanisms. *Psychological Review*, *104*(1), 3–65. http://doi.org/10.1037/0033-295x.104.1.3

Möller, A., Kranz, M., Schmid, B., Roalter, L., & Diewald, S. (2013). Investigating self-reporting behavior in long-term studies. In M. B. Rosson & D. Gilmore (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2931–2940). http://doi.org/10.1145/2470654.2481406

Olson, K., & Parkhurst, B. (2013). Collecting paradata for measurement error evaluations. In F. Kreuter (Ed.), *Improving Surveys with Paradata: Analytic Uses of Process Information* (pp. 43–72). London, UK: John Wiley. http://doi.org/10.1002/9781118596869.ch3

Petrič, G., Petrovčič, A., & Vehovar, V. (2011). Social uses of interpersonal communication technologies in a complex media environment. *European Journal of Communication*, *26*(2), 116–132.

Pew Research Center. (2006). *The cell phone challenge to survey research*. Retrieved from http://www.people-press.org/files/legacy-pdf/276.pdf

Rideout, V. J., Foehr, U. G., & Roberts, D. F. (2010). *Generation M2: Media in the lives of 8-to 18-year-olds*. Henry J. Kaiser Family Foundation. Retrieved from https://kaiserfamilyfoundation.files.wordpress.com/2013/04/8010.pdf

Salvucci, D. D., & Taatgen, N. A. (2011). *The multitasking mind*. New York, NY: Oxford University Press.

Slavec, A., & Vehovar, V. (2014, July). *Reducing response burden by improving question wording: A split-ballot experiment*. Presented at the European Congress of Methodology, Utrecht, Netherlands.

Spink, A., Cole, C., & Waller, M. (2008). Multitasking behavior. *Annual Review of Information Science and Technology*, *42*(1), 93–118. http://doi.org/10.1002/aris.2008.1440420110

Stieger, S., & Reips, U.-D. (2010). What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study. *Computers in Human Behavior*, *26*(6), 1488–1495. http://doi.org/10.1016/j.chb.2010.05.013

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press. http://doi.org/10.1017/CBO9780511819322

Tran, P., Carrillo, R., & Subrahmanyam, K. (2013). Effects of online multitasking on reading comprehension of expository text. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *7*(3). http://doi.org/10.5817/CP2013-3-2

Van Cauwenberge, A., Schaap, G., & van Roy, R. (2014). "TV no longer commands our full attention": Effects of second-screen viewing and task relevance on cognitive load and learning from news. *Computers in Human Behavior*, *38*, 100–109. http://doi.org/10.1016/j.chb.2014.05.021

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York, NY: Springer. http://doi.org/10.1037/11019-008

Voorveld, H. A. M., & van der Goot, M. (2013). Age differences in media multitasking: A diary study. *Journal of Broadcasting & Electronic Media*, *57*(3), 392–408. http://doi.org/10.1080/08838151.2013.816709

Wallis, C. (2010). *The impacts of media multitasking on children's learning and development: Report from a research seminar.* The Joan Ganz Cooney Center and Stanford University. Retrieved from http://multitasking.stanford.edu/MM_FinalReport_030510.pdf

Wang, Z., David, P., Srivastava, J., Powers, S., Brady, C., D'Angelo, J., & Moreland, J. (2012). Behavioral performance and visual attention in communication multitasking: A comparison between instant messaging and online voice chat. *Computers in Human Behavior*, *28*(3), 968–975. http://doi.org/10.1016/j.chb.2011.12.018

Wang, Z., & Tchernev, J. M. (2012). The "Myth" of Media Multitasking: Reciprocal Dynamics of Media Multitasking, Personal Needs, and Gratifications. *Journal of Communication*, *62*(3), 493–513. http://doi.org/10.1111/j.1460-2466.2012.01641.x

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. http://doi.org/10.1518/001872008X288394

Wood, E., Zivcakova, L., Gentile, P., Archer, K., De Pasquale, D., & Nosko, A. (2012). Examining the impact of off-task multi-tasking with technology on real-time classroom learning. *Computers & Education*, *58*(1), 365–374. http://doi.org/10.1016/j.compedu.2011.08.029

Yeykelis, L., Cummings, J. J., & Reeves, B. (2014). Multitasking on a single device: Arousal and the frequency, anticipation, and prediction of switching between media content on a computer. *Journal of Communication*, *64*(1), 167–192. http://doi.org/10.1111/jcom.12070

Zhang, T., Sun, X., Chai, Y., & Aghajan, H. (2015). A look at task-switching and multi-tasking behaviors: From the perspective of the computer usage among a large number of people. *Computers in Human Behavior*, *49*, 237–244. http://doi.org/10.1016/j.chb.2015.03.012

Zwarun, L., & Hall, A. (2014). What's going on? Age, distraction, and multitasking during online survey taking. *Computers in Human Behavior*, *41*, 236–244. http://doi.org/10.1016/j.chb.2014.09.041