



Automatic Versus Manual Forwarding in Web Surveys - A Cognitive Load Perspective on Satisficing Responding

Arto Selkälä¹(✉), Mario Callegaro² , and Mick P. Couper³

¹ University of Lapland, Rovaniemi, Finland
arto.selkala@ulapland.fi

² Google, London, UK
callegaro@google.com

³ University of Michigan, Ann Arbor, MI, USA
mcouper@umich.edu

Abstract. We examine the satisficing respondent behavior and cognitive load of the participants in particular web survey interfaces applying automatic forwarding (AF) or manual forwarding (MF) in order to forward respondents to the next item. We create a theoretical framework based on the Cognitive Load theory (CLT), Cognitive Theory of Multimedia Learning (CTML) and Survey Satisficing Theory taken also into account the latest findings of cognitive neuroscience. We develop a new method in order to measure satisficing responding in web surveys. We argue that the cognitive response process in web surveys should be interpreted starting at the level of sensory memory instead of at the level of working memory. This approach allows researchers to analyze an accumulation of cognitive load across the questionnaire based on observed or hypothesized eye-movements taken into account the interface design of the web survey. We find MF reducing both average item level response times as well as the standard deviation of item-level response times. This suggests support for our hypothesis that the MF interface as a more complex design including previous and next buttons increases satisficing responding generating also the higher total cognitive load of respondents. The findings reinforce the view in HCI that reducing the complexity of interfaces and the presence of extraneous elements reduces cognitive load and facilitates the concentration of cognitive resources on the task at hand. It should be noted that the evidence is based on a relatively short survey among university students. Replication in other settings is recommended.

Keywords: Web surveys · Interactivity · Human-computer interaction · Interface design · Cognitive load theory · Cognitive Theory of Multimedia Learning · Survey satisficing theory · Sensory memory · Working memory · Long-term memory · Satisficing responding · Automatic forwarding · Manual forwarding · Item-level response times · Standard deviation · Cognitive burden · Nondifferentiation · Straightlining · Intrinsic cognitive load · Extraneous cognitive load · Eye-tracking · Eye-movements · Proactive attention · Contiguity principle · Spatial orienting · Attentional capture · Group-mean centering · Log-linear model

1 Introduction

The e-environment consists of various technologies, applications and functionalities like web pages, web browsers, multimedia presentations and web surveys, all of them involving human-computer interaction (HCI). In this article, we apply *Cognitive Load theory* (CLT), *Cognitive Theory of Multimedia Learning* (CTML) and *Survey Satisficing Theory*. CLT and CTML are widely applied in human computer interaction research when Survey Satisficing theory is applied in web survey methodology, the field which has much in common with human-computer interaction research. We create a theoretical framework based on these theories taken also into account the latest findings of cognitive neuroscience in order to understand human-computer interaction in a more coherent way in order to apply this understanding in a particular web survey interface employing automatic page forwarding. In addition, we develop a new method to measuring satisficing responding in web surveys which we introduce in this article applying it in the web survey data.

We investigate a particular functionality of web surveys, the page forwarding procedure and its implications for satisficing respondent behavior in an interaction of other task completion elements. Web surveys are a widely-used data gathering method in social science and market research representing an important advancement in the evolution of self-administered questionnaires making large samples affordable to a wide range of researchers (Tourangeau et al. 2013). However, the prevailing lack of national e-mail registries appears as a major challenge in applying web surveys as a scientific data gathering method making it difficult to constitute a representative sample at the general population level yet overcome by online panels and weighting/post-stratification methods (Callegaro et al. 2015). In addition, web surveys in particular have been found to be subject to declining participation rates (Callegaro et al. 2015). This challenge has been overcome by advanced invitation methods (Bandilla et al. 2012; Callegaro et al. 2015; Dillman 2019; Kaplowitz et al. 2012; Selkälä et al. 2019). It should also be noted that a poor participation rate of any given survey data does not necessarily imply poor statistical representativeness given that a nonresponse rate correlates only weakly with a nonresponse bias (Davern 2013; Groves and Peytcheva 2008).

Computer and web-based survey technology has made possible question formats, layouts and functionalities that would be impossible or difficult to implement using traditional paper questionnaires. Interactivity is then a key element of web surveys, unlike paper and pencil surveys (Couper 2008). One element of interactivity is automatic forwarding, which advances respondents automatically to the next item without the need to click on a “next” button (as in manual forwarding). Another well-known example of interactivity in web surveys is a drop-down question; a format which is impossible to conduct as such on a paper and pencil survey.

In the perspective of a human-computer interaction, a major challenge becomes how to manage a cognitive burden of the participants of any given task. This becomes essential given that in most cases a poor task performance is found to occur due to an excessive cognitive burden of the participants. One of the most influential theories to understand the formation of a cognitive burden in different contexts is the cognitive load theory (CLT) largely adopted in the field of human-computer interaction (HCI)

(Hollender et al. 2010), in usability research (<https://www.nngroup.com/articles/minimize-cognitive-load/>), in internet psychology (Sundar 2007), in the field of instruction science (Clark and Mayer 2016), in the research of multimedia learning (Mayer and Fiorella 2014) and in the web survey methodology.

Given that our aim in this study was not only to analyze how the individual questions affect satisficing responding in different interfaces, automatically (AF) and manually (MF) forwarded web surveys but also to analyze how these interface features interact with the individual questions generating satisficing responding we argue that the conventional approach to understand a cognitive survey response process (Tourangeau 1984; 2018) is in this respect insufficient. This is due to the fact that the different versions of this description start at the level of interpreting the meaning of each question or at the level of comprehension. In both of these cases, the response process starts at the level of working memory turned to be insufficient in terms of analyzing the interface features. In order to understand the cognitive response process as a whole in web surveys the process should be interpreted starting at the level of *sensory memory* (Mayer 2014). As a cognitive response process starts at the level of sensory memory it occurs by selecting appropriate elements into the working memory based on spatial orienting and attentional capture in the preparation of eye movements further generating intrinsic or extraneous load of respondents (Theeuwes 2014). In addition, we expect the total cognitive load of respondents accumulating as responding proceeds across the questionnaire based on the interaction of intrinsic and extraneous loads and the proactive nature of recall (Heideman et al. 2018; Nobre and Stokes 2019).

2 Theoretical Background

2.1 Automatic Versus Manual Forwarding

Auto forwarding or automatic advance in web surveys is a functionality that can be used when a question type has mutually exclusive answers (such as radio buttons) where clicking on a radio button response, for example, takes the respondent directly to the next page (question) without the need to click on a “next” button. Question types like check-all-that-apply and open-ended questions cannot use automatic forwarding. The main arguments in support of auto forwarding are that it (1) reduces respondent burden (number of clicks) and (2) serves as a “forcing function,” requiring the selection of a response to proceed (Selkälä and Couper 2018). With the recent rise in the proportion of respondents completing web surveys on mobile devices (specifically smartphones; see Couper et al. 2017) and the corresponding finding that surveys completed on smartphones take longer to complete than those completed on personal computers (PCs) (Couper and Peterson 2016), researchers are trying to find ways to make such surveys more efficient, especially for the sets of questions with similar response options (see de Bruijne et al. 2015; de Leeuw et al. 2012; Klausch et al. 2012).

As Selkälä and Couper (2018) have noted, while there have been many arguments for and against auto forwarding, empirical research on the topic is scarce. In the first known study to examining auto forwarding, Rivers (2006) reported significantly ($p < .001$) fewer break offs in the AF (40.6%) than in the MF (49.4%) version. He also reported significantly ($p < .01$) shorter completion times for the AF (median = 19.3 min) than

the MF (median = 23.1 min) version and higher levels of user satisfaction with the AF version ($p < .001$). Both of these results were found also in the Selkälä and Couper study (2018). Hays et al. (2010) also found that the survey took about 50% longer ($p < .025$) in the MF version: mean completion times were 9.1 min for AF and 13.5 min for MF. Missing data, reliability, and mean scale scores were similar across the groups. Somewhat similar findings were reached by Giroux et al. (2019) given that they did not find significant differences in survey duration time, straight-lining, breakoff rates, or item nonresponse (for mobile users) between the two experimental groups, but desktop users without the automatic advancement feature had higher item nonresponse.

The research findings regarding item nondifferentiation¹ also called straightlining between auto and manual forwarding are not entirely consistent. Auto forwarding has been shown to increase non-differentiation and primacy effect (Hammen 2010) but also decreasing it in the case of particular horizontal scrolling matrix” (HSM) version (de Leeuw et al. 2012) leading the authors considering findings as an evidence of deeper processing in the AF version. On the other hand, automatic forwarding has been almost consistently shown to decrease web survey completion times in comparison with manual forwarding in the Hays et al. (2010), Rivers (2006) and Selkälä and Couper (2018) studies but not in the Giroux et al. (2019) study. When it comes to item level response times Selkälä and Couper (2018) found AF respondents taking on average 0.4 s ($p < .001$) longer to provide an initial answer to each item. They (2018, p. 11) argue this suggests support for the hypothesis that by simplifying the response process, auto forwarding allows the respondent to focus more fully on the item under consideration.

From the more specific perspective focusing on an opportunity to change an already given answer auto forwarding and manual forwarding represent different solutions. Respondents in the MF condition could change answers before proceeding, whereas AF respondents would need to return to the item to make a change. Respondents can also return to review previous items without making changes. Selkälä and Couper (2018) found that MF respondents change responses significantly more on experimentally manipulated items conveying a low information accessibility or a consistency requirement in comparison to corresponding neutral items in the control groups: 15.5% of the respondents exposed to the low information accessibility version changed answers to this item, compared with 3.3% for the control groups; similarly, 14.9% of those exposed to the consistency requirement changed answers, compared to 2.9% for the control group. They did not find such differences in the AF group; overall changes were very low (0.6% to zero). Taking into account the response time findings that experimentally manipulated items took longer to complete, Selkälä and Couper (2018) conclude that the questions conveying low accessible information or consistency requirement increase cognitive burden of respondents. To some extent, this leads respondents to revisit those items and change their responses. However, no evidence was found to support their hypothesis that respondents in the AF groups return more to experimentally manipulated items in order to change their responses. Instead, they found higher rates of returns for both MF and AF groups to experimentally manipulated items and higher rates of changed answers to the MF groups but not the AF groups. Giroux et al. (2019) found similar results given that

¹ Item nondifferentiation is a response style where the answers to a battery of questions with the same response options (e.g. a table or grid) are the same or very similar among each other.

in their study respondents receiving the automatic advancement treatment on average changed about 50% fewer answers across the survey instrument than those who did not receive the automatic advancement design.

2.2 Cognitive Load Theory, Cognitive Response Process, and an Expected Accumulation of Cognitive Load in the AF and MF Interfaces

Cognitive load theory (CLT) provides a theoretical framework addressing individual information processing and learning (Paas and Sweller 2012). CLT is concerned with the learning of complex cognitive tasks, in which learners are often overwhelmed by the number of interactive information elements that need to be processed simultaneously (Paas et al. 2010). CLT is based on the definition of different types of cognitive loads: intrinsic, extraneous, and germane (Paas et al. 2003). Intrinsic load is the load caused by the complexity of the materials to be learned and therefore the complexity of the schemas that must be acquired (Paas et al. 2010). Extraneous load is caused by inadequately designed instructional procedures that interfere with schema acquisition. Germane load is generated as a result of beneficial instructional design factors that support schema creation, learning, instructional task performance, and transfer (Ayres and van Gog 2009; Hollender et al. 2010; Leppink et al. 2013; van Merriënboer et al. 2006; Paas et al. 2010).

CLT is based on understanding how these different types of loads interact with each other in any learning process or a task completion process. An essential clarification in this respect is offered by Paas et al. (2010). They argue that intrinsic load is dependent upon element interactivity, the number of elements that need to be processed simultaneously by the learner. If element interactivity is high, learning becomes difficult and WM-resource intensive [WM: working memory], whereas for low element interactivity material, learning is easier, requiring fewer WM resources. They also argue (2010) that when instructional material is poorly constructed, extraneous load is generated because the learner is diverted away from schema acquisition and uses up precious WM resources by trying to deal with a suboptimal learning environment. Because intrinsic and extraneous cognitive load are additive, an increase in extraneous cognitive load reduces the WM resources available to deal with intrinsic cognitive load and hence reduces germane cognitive load. On the other hand, when intrinsic cognitive load is high, it becomes important to decrease extraneous cognitive load; otherwise, the combination of both might exceed the maximum cognitive capacity and thus prevent effective, or germane, processing activities to occur.

In most cases when CLT is applied an analysis of a cognitive process is based on intrinsic and extraneous types of loads. This is probably due to practical reasons given that these concepts offer a necessary but also sufficient theoretical basis to understand most of the cognitive processes. As discussed above, intrinsic elements refer to the task completion elements that are essential to the task and cannot be separated of it without jeopardizing the accomplishment of the task. In other words, they are necessary to learn in terms of the task completion. On the other hand, the elements capable to generate extraneous load are in most cases irrelevant in terms of the task completion. They consist of technical procedures or information which is redundant or overlapping with the intrinsic information of the task. An excessive total cognitive load also understood

as a working memory overload can occur either as a result of an interaction between simultaneously occurring intrinsic elements or an interaction between intrinsic and extraneous elements. In a detail the relationship of intrinsic and extraneous loads should be understood as follows. Because intrinsic and extraneous cognitive load are additive, an increase in extraneous cognitive load reduces the working memory resources available to deal with intrinsic cognitive load (Paas et al. 2010). Therefore, if intrinsic load is high, extraneous cognitive load must be lowered. Inversely, if intrinsic load is low, a high extraneous cognitive load may not be harmful because the total cognitive load occurs within working memory limits (van Merriënboer and Sweller 2005).

From the perspective of different memory types CLT focuses on an interaction between working memory (WM) and long-term memory (LTM) which is the key to understanding how learning takes place and how complex problems get solved (Ayres 2018). Like *Cognitive Theory of Multimedia Learning*, CLT argues that new information needs to be first processed and integrated with prior knowledge in WM before it is encoded in LTM as new knowledge (Ayres 2018). These theories also share an understanding that the role of participants in any type of task of Web should be understood as the active participants rather than passive recipients of communication (Sundar 2007). In web surveys this feature of the task completion environment is introduced as an interactive principle (Couper 2008).

However, what differs between CLT and CTML from the perspective of memory processes is that CLT focuses on interaction between working memory and long-term memory when CTML additionally recognizes the importance of *sensory memory* (Mayer 2014). The sensory memory operates at the level of spatial orienting and attentional capture that participate in the preparation of eye movements, in turn responsible for cognitive load formation (Mayer 2014; Theeuwes 2014). This process as a whole is concentrated as a *selection* of relevant information transferred to *working memory* further organizing it in order to create mental representations that are integrated with a prior knowledge of long-term memory (Mayer and Moreno 2003). From this perspective learning or a task completion based on any visual stimuli starts at the level of *sensory memory* proceeding towards a *working memory* and *long-term memory* interaction enabling a deeper information processing. Regarding the common descriptions of a cognitive survey response process the organizing and integrating information processing levels are well represented, unlike the sensory memory level. We therefore suggest that these descriptions should be completed by the selection process occurring at the level of sensory memory.

The widely applied description of a cognitive survey response process is as follows (Tourangeau 1984, 2018):

1. Comprehension (interpreting the meaning of each question)
2. Retrieval (searching and retrieving information stored in memory)
3. Judgment and estimation (integrating the information into an opinion or judgment)
4. Reporting an answer (expressing this opinion appropriately).

A more recent description of the cognitive survey response process with the addition of the level of sensory memory is as follows:

1. Selection (transfer information from sensory memory to working memory based on eye movements)
2. Comprehension (interpret the intended meaning of question)
3. Retrieval (retrieve relevant information from memory)
4. Judgment and estimation
5. Reporting an answer.

As a consequence of the addition introduced above, it becomes possible to analyze the web survey response process from the perspective in what extent the interface features likely generate a cognitive load of the respondents as information is transferred from sensory memory (SM) to working memory (WM). This framework turns out to be beneficial given that in most cases when web survey responding is evaluated, the interface features are excluded from the analysis, focusing instead on the cognitive response process in terms of the substantive nature of individual questions. What is needed instead when trying to understand the web survey response process from the perspective of cognitive burden or cognitive load is to take into account the interface features together with the individual questions.

With regard to the interaction between participants and an interface it has been found that greater interactivity of users with a website engenders greater navigational - and hence cognitive load on users (Sundar 2007). This relationship can also be expected to occur in web surveys in a way that greater effort navigating through the web survey interface increases the total cognitive load of the participants. In addition, an excessive total cognitive load should be expected to occur either as a result of an interaction of intrinsic elements or an interaction of intrinsic and extraneous elements.

A more elaborate view of how the navigation occurs on the interface can be reached by using eye tracking method measuring the eye movements of the participants (Kim et al. 2016; Krejtz et al. 2018; Zagermann et al. 2016). Eye-tracking can be used to detect intrinsic as well as extraneous cognitive load (Makransky et al. 2019). An extraneous load is generated for instance, when two elements, both necessary for learning, are located visually separated on the interface, making it difficult for the participants to reach a coherent understanding about their interrelationship. In order to improve their understanding, the participants are therefore forced to scan back and forth wasting precious cognitive processing capacity (Mayer 2017). This kind of interface design violates *a contiguity principle* generating an extraneous *sensory memory* load further leading to working memory overload (Clark and Mayer 2016; Makransky et al. 2019). The harmful consequences of this design can be explained by the law of proximity referring to a phenomenon fostering learning when related representations are spatially integrated or close to each other (Beege et al. 2019; Clark and Mayer 2016). With regard to surveys it is shown that when items are presented in proximity to each other, the likelihood for an assimilation effect increases (Couper et al. 2001; Tourangeau et al. 2013). The reason for this lies in the law of proximity, which causes items to be perceived as a group (Toepoel et al. 2009).

In a typical situation when CLT is applied to an evaluation of usability of any given interface the interface elements affecting usability remain as permanent across the task completion. This is the case for instance regarding navigation through an ordinary web page not including any interactive elements. However, when CLT is applied to the web

survey response process its role should be understood differently. What makes the web survey task completion process different is the repetition. In the web survey response process the same type of intrinsic-, and extraneous elements repeatedly follow each other instead of occurring once. Taken into account the particular nature of a web survey responding process, we expect the total cognitive load to be generated across the response process based on the permanent elements of the process like interface features and non-permanent elements like individual questions and their interaction. We also expect the total cognitive load of respondents accumulating across the responding occurring at the highest level at the end of the questionnaire assuming that the capacity of individual web survey questions to generate a cognitive load of respondents remains somewhat equal across the survey. The cognitive load accumulation is recognized in CLT and measured by subjective measures like rating scales as well as objective measures like the amount of time the learner spends on completing the task or navigation behavior (Antonenko and Keil 2018).

Given that we were not able to use an eye tracking method we nevertheless offer a hypothesis with regard to the eye movements on the studied interfaces AF and MF. The hypothetical framework regarding eye movements makes it possible to understand how the studied interface features contribute and interact in the formation of different types of cognitive loads during the response process. It should be noted that the respondents in both studies were encouraged to use PCs or tablets but were not prevented from using a smartphone. Despite this opportunity, only 16% of the respondents completed the survey using a smartphone (Selkälä and Couper 2018). Thus, although the interfaces for PC and smartphone respondents differ considerably (Appendix 2) we find the proportion of smartphone respondents so low that we do not expect it to affect the results.

The expected eye movements on a manually forwarded interface are illustrated in Fig. 1. Given that the interface in both versions (AF/MF) is divided in two parts, on the left side of the screen appears the list of the items when on the right side of the screen appears the particular question and the response options of it. As the response has been given to the particular question the next question on the list is activated with a shaded background. At the same time, it appears visible on the right side of the screen. It should be noted as well that the individual items are shared into groups. The headline of the question group appears visible on the left side of the screen when the items within it appear visible below of the headline. We expect the eye movements occurring in the response process as follows. Firstly (1) respondents are likely to focus on the relationship of the headline of a question group and the first item within it. Secondly (2) we expect them to focus on the relationship of the headline of a question group and the headline of the first item on the right side of the screen. Thirdly (2.b) we expect the respondents focusing on the relationship of the headline of the first item on the right side of the screen and other items on the item list. We expect these particular eye movements to generate lower cognitive load than other expected eye movements given that their intention is more to become aware of the task completion elements as a whole than focus on the particular question under the consideration. We expect the third (3) particular eye movements occurring between the headline of the first item on the right side of the screen and the response options below of it. In the automatically forwarded interface (AF) we do not expect any other notable eye movements to occur but in the manually forwarded version (MF) we

expect the fourth (4) major eye movement to occur between the response options of the responded question and previous and next buttons below them. We expect the first, second and the third eye movements generating intrinsic load in both of the versions. In addition we expect the fourth (4) eye movements in manually forwarded (MF) version generating extraneous load.

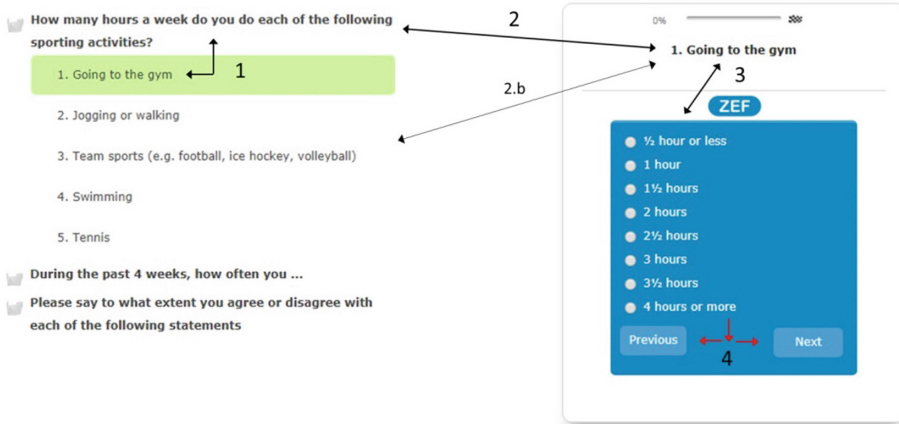


Fig. 1. The expected eye movements on manually forwarded interface.

In addition to permanent interface features discussed above the repetition of individual questions should take into account in order to understand how a cognitive load accumulates in detail in the AF and MF conditions. As illustrated in the popular descriptions of the cognitive response process (Tourangeau 1984, 2018) the retrieval is an essential part of responding to individual questions. However, this part of the response process should be understood differently when trying to achieve a coherent understanding about an accumulation of cognitive load in terms of the individual questions and the interface as a whole. The recent findings from the field of neuroscience show that recall cannot be considered just as a passive operation to retrieve something from the long-term memory. It should instead be considered as an active process especially in terms of anticipating upcoming events.

An increasing variety of experimental approaches is being used to explore how long-term memory (LTM) content is used proactively to guide adaptive behavior. The approaches share the notion that the brain uses LTM information constantly, proactively, and predictively (Nobre and Stokes 2019). Information in working memory has been considered the major source of top-down proactive attention. Even before the target stimuli appear, these memory traces influence the pattern of brain activity in a proactive fashion to facilitate the processing of signals associated with likely relevant items (Chelazzi et al. 1993; Kastner et al. 1999; Stokes et al. 2009). In addition, a recent neurophysiological study of sequential learning in a serial response task has also revealed a proactive anticipation of upcoming stimuli and associated responses based on learned spatiotemporal expectations (Heideman et al. 2018; Nobre and Stokes 2019).

These findings receive support from the survey satisficing studies showing that satisficing respondent behavior and straightlining occur more likely towards the end of the questionnaire than toward the beginning (Knowles 1988; Krosnick and Alwin 1988). When the retrieval process is understood in a proactive fashion as discussed above, anticipating the upcoming stimuli, it becomes easier to understand why satisficing respondent behavior becomes more likely towards the end of the questionnaire. When the respondent retrieves an activated material from long term memory integrating it with the mental representation based on the present question the invested cognitive effort depends also on other sources to increase total cognitive load. These sources generate intrinsic or extraneous load, or the load based on their interaction. The anticipatory nature of retrieval can explain why respondents tend to relieve excessive load through satisficing in repetitive tasks like web surveys. It can be expected to occur when the repetitive task includes certain permanent interface elements increasing the extraneous load on the respondents. This is the case in the manually forwarded interface. Under these circumstances, the respondents can easily anticipate the burdening nature of the upcoming task and adjust the invested cognitive effort across the individual items (see Heideman et al. 2018; Nobre and Stokes 2019). A crucial part of this process in the MF interface generating extraneous load is the selection between the previous and next buttons illustrated in Fig. 2.

As a consequence of an interaction between the individual questions and the MF mechanism we expect a total cognitive load of respondents accumulating and turning excessive as individual items follow each other. The excessive load is then relieved through satisficing. Inversely arguing, if we expect the accumulation of total cognitive load not occurring, we shouldn't be able to observe an increased satisficing towards the end of the questionnaire. In this case, each load in relation with the individual items should completely be relieved after giving a response. What follows is that in this case a respondent should be able to start a response process on each question without an accumulated load originated from previous questions. Given that the empirical findings and theoretical reasoning discussed above suggest otherwise we accept the accumulation hypothesis.

2.3 Survey Satisficing and Response Time Based Measurements

Krosnick's survey satisficing theory (Krosnick 1991, 1999) is probably the most influential theory regarding satisficing respondent behavior in surveys. It is based on Tourangeau's (1984) description of the cognitive process taken place in survey responding borrowing also from Simon's (1957) more general theory of decision-making. Survey satisficing occurs when instead of investing a sufficient effort to respond thoughtfully or optimally respondents take shortcuts in order to minimize cognitive effort (Kim et al. 2019; de Rada and Dominguez-Alvarez 2014; Zhang and Conrad 2018). It is this resort to a satisfactory rather than an optimal decision strategy that gives satisficing theory its name (Hamby and Taylor 2016).

Krosnick introduced the idea of "weak" and "strong" forms of satisficing. Weak satisficing occurs when the four cognitive stages of survey responding—comprehension, recall, retrieval, and judgment (Cannell et al. 1981; Tourangeau et al. 2000; Tourangeau 2018) are undertaken but less thoroughly than they might be. Strong satisficing occurs when one or more of these stages are skipped entirely. Examples of weak satisficing are

The selection of back or forward buttons in the MF interface

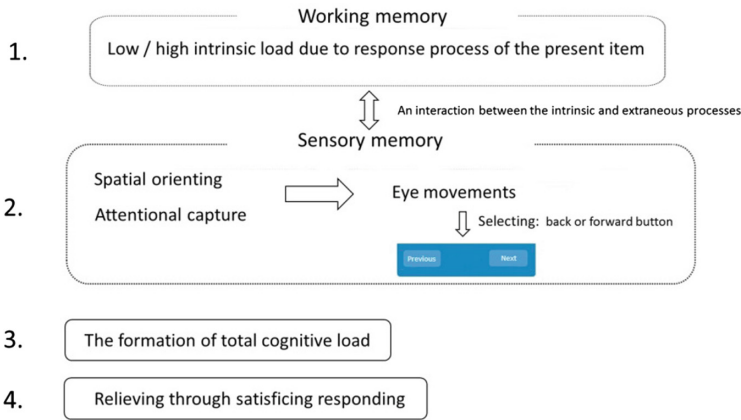


Fig. 2. The hypothesized process selecting between previous and next buttons in the MF interface.

acquiescence, where respondents show a tendency to agree with statements in attitude questions, selecting the midpoint on opinion questions with odd numbers of response options, and selecting the first reasonable option from a list rather than considering all options and selecting the most appropriate (Krosnick and Alwin 1987). The strong form of satisficing occurs for instance when respondents select “*Don’t Know*” when they could provide a substantive answer, when they select a substantive response option randomly or when they do not differentiate their responses on a battery of scale items (Hamby and Taylor 2016; Lipps 2007; Kaminska et al. 2011; Vannette and Krosnick 2014).

Non-differentiation—in other words, straightlining—a tendency to give the same answers across several items is a widely applied satisficing measure in survey methodology. It can be detected by Cronbach alpha, simple nondifferentiation, mean root of pairs, maximum identical rating, standard deviation of battery and scale point variation (Kim et al. 2019). Non-differentiation is more common among respondents with less education and low verbal ability and it is more common toward the end of a questionnaire than toward the beginning (Krosnick and Alwin 1988; Vannette and Krosnick 2014).

Survey satisficing is often suggested as occurring as a consequence of an excessive cognitive burden on respondents. Following this hypothesis Vannette and Krosnick (2014) suggest that in order to minimize the likelihood of satisficing, questionnaire designers should take steps to maximize respondent motivation and minimize task difficulty. This can be reached by making it easy for respondents to interpret questions, to retrieve information from memory, to integrate the information into a judgment, and to report the judgment (Vannette and Krosnick 2014). Regarding to measure the cognitive burden of respondents, response times are widely applied (Zhang and Conrad 2014). Their popularity in this respect is based on the assumption that if a survey question is in some way difficult or complex, it takes more time to answer due to greater thought and attention to determine a response (Turner et al. 2015).

In terms of satisficing, shorter item-level response times have been suggested to indicate sub-optimal responding or satisficing (Di et al. 2016; Zhang and Conrad 2014; Conrad et al. 2017). Zhang and Conrad (2014) showed that straightlining is significantly associated with speeding, a tendency to answer faster than is necessary in order to offer a response processed at least at the minimum level of cognitive effort. The findings of Callegaro et al. (2009) suggest support for these findings given that in their study the participants (job applicants) who inherently were expected holding stronger motivation towards the task, spent more time to accomplish the task than their counterparts, the less motivated participants. Thus, faster task accomplishment indicates satisficing task performance in their study as well. However, regarding the response time based satisficing measurements, Turner, Sturgis, and Martin (Turner et al. 2015) found the opposite results given that a higher proportion of “*Don’t Know*” answers and a tendency for rounding—typical satisficing measures—were associated with *longer* response latencies.

Using response latencies as a satisficing measure is not unproblematic given that the item level response times do not vary just because of the motivation and ability of respondents (Krosnick 1991; Vannette and Krosnick 2014) but other individual level factors as well. One of these factors is related with the information accessibility of respondents. It is well known that respondents with strong attitudes tend to offer their initial answer faster in attitude questions compared with their counterparts with less strongly attitudes (Fazio 1990). Regarding these respondents, shorter item level response times do not indicate satisficing but a tendency to answer faster due to more easily accessible information in their working memory. On the other hand when a question is complex and relevant information more difficult to recall, answering takes longer due to a greater attention and a cognitive process to formulate an answer (Turner et al. 2015; Yan and Tourangeau 2008). It is also well known that satisficing respondent behavior becomes more prevalent as responding proceeds towards the end of the questionnaire (Vannette and Krosnick 2014). What follows is that the item level response times should be decreased correspondingly towards the end of the questionnaire if shorter item-level response times are interpreted to indicate satisficing responding. On the other hand, as the responding proceeds towards the end of the questionnaire respondents become more fatigued due to accumulated cognitive burden. This should be detected as longer item level response times in case an appropriate amount of cognitive effort is invested in each item. However, when the respondents take shortcuts (satisficing) the item level response times should to decrease instead.

As discussed above, applying item level response times as a measure of satisficing is not as straightforward as suggested in the previous literature. This is mainly due to various factors affecting item level response times including individual level factors like motivation and ability or an information accessibility but also questionnaire level factors like a length of the questionnaire, an interface design in web surveys as well as question level factors like a substantial complexity of survey questions. It seems obvious that the interrelationship of these different sources affecting item level response times requires a clarification.

As the traditional survey satisficing theory is extended by the cognitive load theory, the interrelationship of its core elements; respondent ability, respondent motivation and task difficulty can be understood in a more advanced fashion. The starting point to

understand the relationship of these factors with an understanding how the cognitive load is generated is to realize that in case when the task difficulty is increased to a very high level generating correspondingly high cognitive load, the respondents inherently susceptible to satisficing respondent behavior tend to relieve an excessive load through satisficing. According to survey satisficing theory, these respondents are more likely less educated (individuals with lower abilities) and less motivated in comparison with their counterparts. At a more detailed level, the satisficing can be expected to occur more frequently in relation with more complex and cognitively demanding questions given that despite whether satisficing occurs at the conscious or unconscious level of responding it can be explained by an increased pressure to relieve an excessive cognitive load through it. This becomes most effectively executed regarding the most demanding questions generating the highest cognitive load because by taking shortcuts (skipping the entire stages of a response process) with regard to these questions the highest amount of excessive load can be relieved with minimum effort.

Regarding the less demanding questions in the survey, the satisficers can either be expected to follow the similar satisficing pattern than with the more complex questions or invest their major cognitive capacity in order to respond to these questions in particular. The latter option becomes more likely when the satisficers are truly less motivated individuals with lower abilities as the survey satisficing theory claims given that these types of individuals probably find the easiest questions the most convenient to answer. It should also be noted that from the perspective of an accumulation of cognitive load across the survey the less motivated respondents with lower abilities should become even more susceptible to satisficing respondent behavior towards the end of the questionnaire. Consequently, this tendency becomes even stronger when the individual questions are substantially complex or the web survey interface elements increase an extraneous load of the participants.

Table 1. The examples of satisficing responding patterns on item level response times.

Items	load	A	B	C
1	-	↓	↑	↗
2	+	↘	↘	↘
3	-	↓	↑	↗
4	+	↘	↘	↘
5	-	↓	↑	↗
6	+	↘	↘	↘

Note. The direction of arrow represents an expected change in direction of item item level response time. The size of the arrow represents an expected change in the magnitude of the item level response time.

As a result of an extension of survey satisficing theory discussed above at least three satisficing responding patterns can be expected to occur under cognitively burdening

circumstances (Table 1). As expected in the traditional survey literature, shorter item level response times could occur evenly spread across the survey items as a consequence of increased satisficing responding. This satisficing pattern results shorter average item level response times given that all or most of the individual response latencies occur shorter in comparison with less burdening situation generating weaker satisficing respondent behavior. In Table 1 column A represents this satisficing responding pattern. However, as discussed above it is actually more likely that satisficers do not invest less response time consistently in each item across the survey but vary their responding pattern item by item. This becomes likely in particular when there are different types of questions on the survey in terms of their tendency to increase a cognitive load of respondents. The columns B and C represent these satisficing patterns. In both of these cases satisficers invest their major cognitive capacity to focus more carefully on cognitively less demanding questions. As a result, the item level response times do not decrease evenly across the items but decrease in terms of cognitively demanding questions and increase in terms of cognitively less demanding questions. As a result, the difference in average item level response times between satisficers and non-satisficers diminishes and may even disappear with regard to certain question combinations. However, what remains is converging item level response times in each of the illustrated satisficing responding patterns A, B and C. Thus, even though the difference in average item level response times cannot be treated as a reliable satisficing measure, the variation of item level response times appears to be reliable given that in all of the illustrated cases it would decrease. This makes the standard deviation (SD) of item level response times a more prominent measure of satisficing when an average item level response times should be treated as a secondary, a complementary measure of satisficing.

3 Subjects

In the first study (the University of Lapland, Finland) 3,023 undergraduate students were randomly assigned to six independent experimental conditions. This survey was fielded from October 7 to October 28, 2015. In the second study (the Lapland University of Applied Sciences) 5,004 undergraduate students were randomly assigned to six independent conditions following the same procedures as Study 1. This survey was fielded from April 18 to May 8, 2016. Respondents in both studies were encouraged to use PCs or tablets, but were not prevented from using a smartphone. The breakoffs and response rates of the aggregated data of two samples are shown in Appendix 3. The final data was applied in the present study in combining the automatically forwarded (AF) and manually forwarded (MF) groups resulting in two groups; AF ($n = 863$) and MF ($n = 900$).

4 Method

4.1 Measuring Response Times

Both client-side and server-side paradata were captured, and response time was measured at both the respondent (survey) level and the item level (Yan and Olson 2013). The total response time (TRT) for a survey for a particular respondent was calculated by taking

the difference between the first and last time stamps in the survey. **Item-level response times were also calculated as the difference between mouse clicks on two radio buttons or between the mouse clicks of the forward/backward button and a radio button. This is a measure of the time to select an initial response for an item after a page has loaded. It does not include the time following this selection** (i.e., the time taken to change an answer or to click the next button in the MF version).

In order to avoid the drawbacks in measuring satisficing and to develop a methodological solution taking into account the individual-level influence we use a standard deviation (SD) of item-level response times (calculated within the individuals) as a measure of satisficing. We interpret a decreased standard deviation of item-level response times indicating increased satisficing. This becomes intelligible when realizing that substantially different items require different amount of cognitive effort to become comprehended, the necessary information retrieved, an appropriate judgment completed, and a response given. As the standard deviation of item-level response times decreases, it reveals respondents investing response time in a more similar fashion in different types of items. This suggests an increased total cognitive load occurring across the items further relieved through satisficing responding. Empirically it can be recognized as a decreased standard deviation. Given that as we expect the MF to be associated with a more cognitively burdening procedure due to the previous and next buttons, we should observe lower SD in the MF group in comparison with the AF group.

However, because the standard deviation of item level response times is affected by an individual tendency to respond slower or faster (respondent baseline speed), we took an average item level response times of individuals account within the examined question batteries as a nuisance variable. In other words, the individual tendency to respond slower or faster was removed from the estimate of standard deviation by modelling it as an independent variable. Given that the original relationship of an average response speed of individuals and the SD appeared to be curvilinear in order to take this into account we were able to achieve more accurate estimates as well as compare the three question batteries including different types of items.

We applied the standard deviation (SD) of item-level response times as a measure of satisficing in several log-linear regression models. The approach was conducted within three question batteries separately including the items varying in their expected tendency to increase intrinsic cognitive load. In order to control the confounded influence of average respondent-level response speed on SD we added it in the models as an explanatory variable with the dummy variable “AF/MF”. To make the explanatory variables uncorrelated we group-mean centered the values of the average respondent-level response speed within the AF/MF groups (see Bell et al. 2018; Dalal and Zickar 2012; Enders and Tofghi 2007; Paccagnella 2006). We excluded individuals from the analysis whose average item-level response times exceeded an upper outer fence in the boxplot². The model is:

$$\log Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

² When answering web surveys some respondents leave a question “open” for a long time because they interrupt answering the survey due to a variety of reasons. On that specific screen or question, the time latency is very high and therefore such cases need to be excluded from the analysis.

Y = The log of standard deviation of item-level response times

X_1 = An average respondent-level response speed based on item-level response times

X_2 = AF/MF.

5 Results

The examination of the first question battery (4 items) revealed (Table 2) that one unit change in the “AF/MF” - variable (X_2) is associated with a 16% decrease ($p < 0.001$) in the expected geometric mean of the SD indicating that we expect to see about a 16% decrease in the geometric mean of the standard deviation of item-level response times for the MF group ($n = 863$) compared with the AF group ($n = 846$). The corresponding decrease regarding the second question battery (9 items; AF, $n = 855$; MF, $n = 881$) was 1% (non-significant) when for the third battery (4 items; AF, $n = 842$; MF, $n = 873$), it was 6% (non-significant; $p = 0.108$). A one-unit change in an average respondent-level response speed (X_1) was associated with 0.02% to 0.03% increase in SD in all three question batteries.

Table 2. An average respondent-level response speed and AF/MF associated with log of standard deviation of item-level response times.

	β	Exp. β	95% CI	p
<i>Items 2–5</i>				
RT	0.00026	1.00	1.00–1.00	<0.001
AF/MF	– 0.17439	0.84	0.78–0.91	<0.001
<i>Items 6–14</i>				
RT	0.00022	1.00	1.00–1.00	<0.001
AF/MF	– 0.01189	0.99	0.94–1.04	n.s
<i>Items 15–18</i>				
RT	0.00015	1.00	1.00–1.00	<0.001
AF/MF	– 0.06626	0.94	0.86–1.02	n.s

Note. RT = An average respondent-level response time. Exp. β - exponentiated coefficient represents the ratio of expected geometric mean difference in the standard deviation of item-level response times for a 1-unit change in an explanatory factor. Maximum likelihood estimation.

The results are consistent with the theoretical expectations given that the largest difference (16%) in the expected geometric mean of SD between the AF and MF groups was found in the first question battery (items 2–5) including the most demanding questions. The second largest difference (6%) was found in the third question battery (items 15–18), including cognitively demanding attitude questions. The smallest (non-significant) difference (1%) was found in the second question battery (items 6–14), including the easiest items to answer; the mood items. The direction of the observed effects occurs

as expected, suggesting support for the hypothesis that as the total cognitive load of respondents increases, the standard deviation of item-level response times decreases. Additionally, the results suggest indirect support for the hypothesized eye-movements in the examined interfaces given that in the MF group a number of hypothesized eye-movements to navigate through the interface is bigger (4) than in the AF group (3) due to the previous and next buttons (Fig. 1). As a result, the total cognitive load of manually forwarded respondents tends to exceed their working memory capacity especially when the content of the questions (the first and third question batteries) with the required eye movements to navigate through the interface generates a high intrinsic load. In an interaction with a generated extraneous load due to the previous and next buttons the total cognitive load of respondents in the MF condition increases to a very high level exceeding their working memory capacity. As a consequence, the respondents relieve the excessive load through satisficing responding which was detected by a decreased standard deviation of item level response times.

In the second question battery, the total cognitive load of respondents is smaller due to the content of items. What follows is that the working memory capacity is not exceeded even in the MF group despite the extraneous load generated by the previous and next buttons. Empirically this was observed as a very small difference in standard deviation between the AF and MF groups. One should note however that this finding indicates nothing with regard to the overall level of satisficing across the mood items. It reveals only that the generated cognitive load of the respondents between the AF and MF groups does not differ from enough in order to lead the second of the groups relieving it through satisficing respondent behavior.

Despite the fact that satisficing respondent behavior is in other studies found to occur more frequently towards the end of the questionnaire we did not find support for these findings regarding the standard deviation. Despite the burdening content of the questions in the third response battery, the decrease in standard deviation was smaller in the MF group (6%) than corresponding difference in the first question battery (16%). One possible explanation for this is that the questionnaire was quite short, and the respondents were university students with high cognitive skills.

On the other hand, the differences between the average item-level response times in the AF and MF groups (Table 3) support the previous findings of the satisficing occurrence towards the end of the questionnaire given that the differences in average item-level response times between the AF and MF groups occurred the smallest in the first question battery, the second largest in the second question battery and the largest in the third question battery. The corresponding decrease in percentages in MF group compared with AF group were in the first question battery 4.2%, in the second question battery 5.9% and in the third question battery 7.3%. This suggests the likelihood of satisficing responding increasing in MF group towards the end of the questionnaire if the difference in average item-level response times in comparison with AF group is considered as a measure of satisficing.

6 Conclusion

In the perspective of a human-computer interaction, a major challenge becomes how to manage a cognitive burden of the participants which is also interpreted as increasing

Table 3. The differences in average item-level response times between the AF and MF groups in milliseconds.

	AF	MF	AF – MF	<i>t</i>	<i>p</i>
Items 2–5	3885	3720	165	1.932	<0.05
Items 6–14	4509	4241	268	3.801	<0.001
Items 15–18	6827	6328	499	3.855	<0.001

Note. The mean values are the average item-level response times calculated separately within the three question batteries based on the combined 3 groups allocated to AF and the three groups allocated to MF. Extreme outliers excluded based on Tukey's method. One-tailed t-tests, equal variances not assumed.

satisficing in the survey literature. In order to estimate satisficing between the different types of interfaces (AF/MF) we take into account that satisficers do not necessarily invest response time consistently in each item across the survey. Consequently, we introduce a more elaborated method, the standard deviation of item level response times, to measure satisficing and estimate the amount of total cognitive load of respondents.

We find support for our hypothesis that the MF version increases satisficing responding given that MF reduces both average item level response times as well as the standard deviation of item-level response times. This suggests support also for the hypothesis that the MF generates higher total cognitive load of respondents due to a more complex interface design. On the other hand, AF has shown to reduce completion times across the whole questionnaire (Selkälä and Couper 2018). This is consistent with our findings given that a shorter completion time allows the respondent to focus on individual items more carefully which is observed as a longer average item level response times. The findings reinforce the view in HCI that reducing the complexity of interfaces and the presence of extraneous elements reduces cognitive load and facilitates the concentration of cognitive resources on the task at hand.

To test these ideas further, we need longer questionnaires to detect fatigue effects towards the end of the questionnaire. We also need questions with varying levels of cognitive effort. We also need eye-tracking studies to determine the effect of the back and the next button on cognitive load. In addition, should be noted that the evidence is based on a relatively short survey among university students. Replication in other settings is recommended. We caution that this applies primarily to highly repetitive tasks, such as answering a series of questions using the same response format (as is often the case in the batteries of psychological measures). The effect may be different in surveys that vary the format and content of items. In other words, AF may be suitable for some types of survey questionnaires, but maybe not others. Further research is needed to figure out under what survey conditions AF is optimal.

Appendix 1. The Questionnaires of the Experiment

Conditions: 1(MF) & 2(AFM) (control groups)	Conditions: 3(MF) & 4(AFM) (low accessible information)	Conditions: 5(MF) & 6(AFM) (consistency requirement)
<i>The first file (items 1-5)</i>	<i>The first file</i>	<i>The first file</i>
How many hours a week do you do each of the following sporting activities? (The response options: ½ hour or less, 1 hour, 1½ hours, 2 hours, 2½ hours, 3 hours, 3½ hours, 4 hours or more)		
1. Going to the gym	1. Going to the gym	1. Going to the gym
2. Horse sport	2. Jogging or walking	2. Horse sport
3. Team sports (e.g. football, ice hockey, volleyball)	3. Team sports (e.g. football, ice hockey, volleyball)	3. Team sports (e.g. football, ice hockey, volleyball)
4. Swimming	4. Swimming	4. Swimming
5. Tennis	5. Tennis	5. All sports put together
<i>The second file (items 6-14)</i>	<i>The second file</i>	<i>The second file</i>
During the past 4 weeks, how often you... (The response options: All of the time; Most of the time; Some of the time; A little bit of the time; None of the time)		
6. Have been happy?	6. Have been happy?	6. Have been happy?
7. Have felt that you can reach all the goals you have set for yourselves?	7. Have felt that you can reach all the goals you have set for yourselves?	7. Have felt that you can reach all the goals you have set for yourselves?
8. Have felt yourself as energetic	8. Have felt yourself as energetic	8. Have felt yourself as energetic
9. Have felt downhearted	9. Have felt downhearted	9. Have felt downhearted
10. Have felt calm and peaceful	10. Have felt calm and peaceful	10. Have felt calm and peaceful
11. Have felt full of pep	11. Have felt full of pep	11. Have felt full of pep
12. Have been nervous	12. Have been nervous	12. Have been nervous
13. Have felt so down in the dumps that nothing could cheer you up	13. Have felt so down in the dumps that nothing could cheer you up	13. Have felt so down in the dumps that nothing could cheer you up
14. Have felt tired	14. Have felt tired	14. Have felt tired
<i>The third file (items 15-18)</i>	<i>The third file</i>	<i>The third file</i>
Please say to what extent you agree or disagree with each of the following statements. (The response options: Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly disagree)		
15. Exercising has health benefits	15. Exercising increases body's ability to buffer lactic acid	15. Exercising has health benefits
16. Exercising becomes more difficult for students with low income than for students with high income	16. Exercising becomes more difficult for students with low income than for students with high income	16. Exercising becomes more difficult for students with low income than for students with high income
17. I can easily afford chargeable sporting activities	17. I can easily afford chargeable sporting activities	17. I can easily afford chargeable sporting activities
18. Students would benefit in many ways by increasing student grants	18. Students would benefit in many ways by increasing student grants	18. Increasing student grants would increase the time I will spend in chargeable sporting activities

Appendix 2. Screen Shots of ZEF Interface

a. Manually forwarded ZEF survey on PC.

Arvioi kuinka monta tuntia viikossa käytät seuraavien liikuntamuotojen harrastamiseen?

- Kuntosali
- Hevosurheilu
- Joukkuelajit (esim. jalkapallo, jääkiekko, lentopallo)
- Uinti
- Tennis

Kuinka usein viimeisen 4 viikon kuluessa olet...

Mitä mieltä olet seuraavista väittämistä?

0% 38%

1. Kuntosali

ZEF

- ½ tuntia tai vähemmän
- 1 tunti
- 1½ tuntia
- 2 tuntia
- 2½ tuntia
- 3 tuntia
- 3½ tuntia
- 4 tuntia tai enemmän

Edellinen Seuraava

b. Automatically forwarded ZEF survey on PC.

Arvioi kuinka monta tuntia viikossa käytät seuraavien liikuntamuotojen harrastamiseen?

- Kuntosali
- Hevosurheilu
- Joukkuelajit (esim. jalkapallo, jääkiekko, lentopallo)
- Uinti
- Tennis

Kuinka usein viimeisen 4 viikon kuluessa olet...

Mitä mieltä olet seuraavista väittämistä?

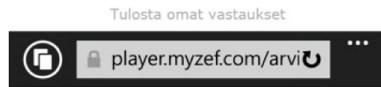
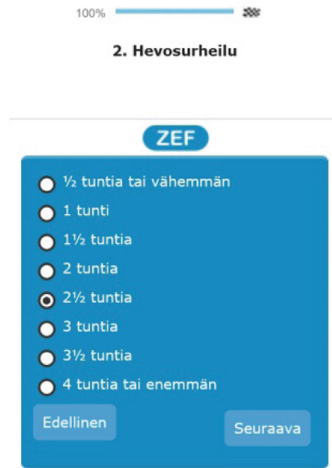
0% 38%

1. Kuntosali

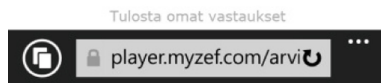
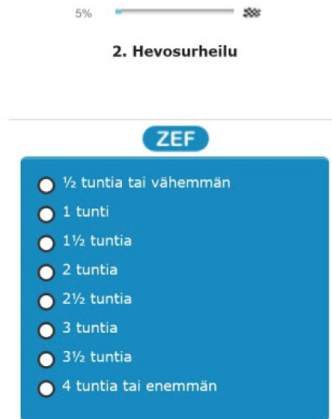
ZEF

- ½ tuntia tai vähemmän
- 1 tunti
- 1½ tuntia
- 2 tuntia
- 2½ tuntia
- 3 tuntia
- 3½ tuntia
- 4 tuntia tai enemmän

c. Manually forwarded ZEF survey on smartphone.

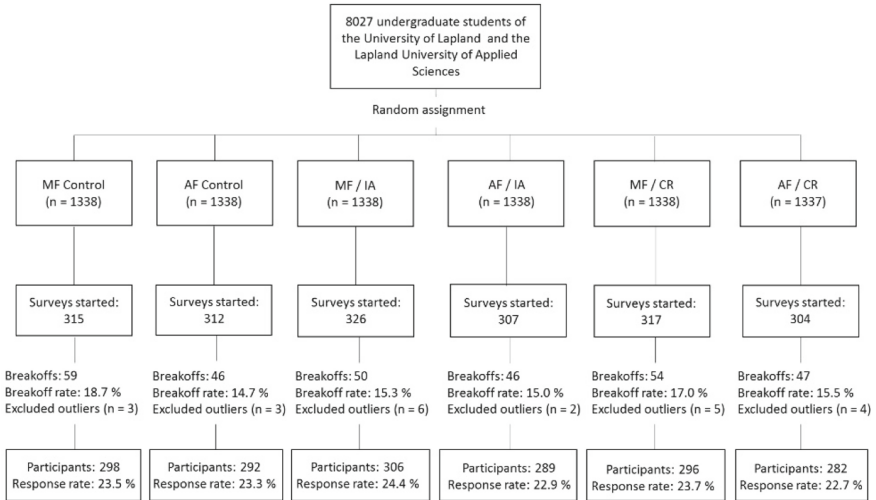


d. Automatically forwarded ZEF survey on smartphone.



Appendix 3. Experimental Design of the Original Study

(Selkälä and Couper 2018)



Note. MF = Manual forwarding; AF = automatic forwarding; IA = information accessibility; CR = consistency requirement. The response rates are calculated by following the RR2 standard definition (AAPOR 2008, pp. 34, 48; Bethlehem and Biffignandi 2011, 439). The breakoff rates are calculated by the QBR definition (Callegaro et al. 2015) representing the combined proportion of breakoffs and partially completed questionnaires of all respondents starting the questionnaire.

References

- Antonenko, P.D., Keil, A.: Assessing working memory dynamics with electroencephalography. Implications for research on cognitive load. In: Zheng, R.Z. (ed.) *Cognitive Load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice*, 1st edn. Routledge, New York (2018)
- Ayres, P.: Subjective measures of cognitive load. What can they reliably measure? In: Zheng, R.Z. (ed) *Cognitive Load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice*, 1st edn. Routledge, New York (2018)
- Ayres, P., van Gog, T.: State of the art research into cognitive load theory. *Comput. Hum. Behav.* **25**, 253–257 (2009). <https://doi.org/10.1016/j.chb.2008.12.007>
- Bandilla, W., Couper, M.P., Kaczmirek, L.: The mode of invitation for web surveys. *Surv. Pract.* **5**, 1–5 (2012). <https://doi.org/10.29115/sp-2012-0014>
- Beege, M., Wirzberger, M., Nebel, S., Schneider, S., Schmidt, N., Rey, G.D.: Spatial continuity effect vs. spatial contiguity failure. Revising the effects of spatial proximity between related and unrelated representations. *Front. Educ.* **4**, 86 (2019). <https://doi.org/10.3389/educ.2019.00086>
- Bell, A., Jones, K., Fairbrother, M.: Understanding and misunderstanding group mean centering: a commentary on Kelley et al.'s dangerous practice. *Qual. Quant.* **52**, 2031–2036 (2018)

- Callegaro, M., Yang, Y., Bhola, D.S., Dillman, D.A., Chin, T.-Y.: Response latency as an indicator of optimizing in online questionnaires. *Bull. Méthodologie Sociologique* **103**, 5–25 (2009)
- Callegaro, M., Lozar Manfreda, K., Vehovar, V.: *Web Survey Methodology*. Sage Publications, London (2015)
- Cannell, C.F., Miller, P.V., Oksenberg, L.: Research on interviewing techniques. *Sociol. Methodol.* **12**, 389–437 (1981)
- Chelazzi, L., Miller, E.K., Duncan, J., Desimone, R.: A neural basis for visual search in inferior temporal cortex. *Nature* **363**, 345–347 (1993)
- Clark, R.C., Mayer, R.E.: *e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*, 4th edn. Wiley, Hoboken (2016)
- Conrad, F.G., Tourangeau, R., Couper, M.P., Zhang, C.: Reducing speeding in web surveys by providing immediate feedback. *Surv. Res. Methods* **11**, 45–61 (2017)
- Couper, M.P.: *Designing Effective Web Surveys*. Cambridge University Press, New York (2008)
- Couper, M.P., Peterson, G.: Why do web surveys take longer on smartphones? *Soc. Sci. Comput. Rev.* **35**, 357–377 (2016)
- Couper, M.P., Traugott, M.W., Lamias, M.J.: Web survey design and administration. *Public Opin. Q.* **65**, 230–253 (2001)
- Dalal, D.K., Zickar, M.J.: Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organ. Res. Methods* **15**, 339–362 (2012)
- Davern, M.: Nonresponse rates are a problematic indicator of nonresponse bias in survey research. *Health Serv. Res.* **48**, 905–912 (2013)
- de Bruijne, M., Das, M., van Soest, A., Wijnant, A.: Adapting grid questions on mobile devices. Paper presented at the European Survey Research Association Conference, Reykjavik, July 2015
- de Leeuw, E.D., Hox, J.J., Klausch, T., Roberts, A., de Jongh, A.: Design of web questionnaires: matrix questions or single question formats. Paper presented at the Annual Conference of the American Association for Public Opinion Research, Orlando, FL, May 2012
- de Rada, V.D., Dominguez-Alvarez, J.A.: Response quality of self-administered questionnaires: a comparison between paper and web questionnaires. *Soc. Sci. Comput. Rev.* **32**, 256–269 (2014). <https://doi.org/10.1177/0894439313508516>
- Di, J.: Determining survey satisficing of online longitudinal survey data in the multicenter AIDS cohort study: a group-based trajectory analysis. *J. Med. Internet Res. Public Health Surveill.* **2**, 1–10 (2016)
- Dillman, D.: Web-push surveys; origins, uses and unsolved challenges. 2019 JPSM distinguished lecture, University of Maryland, 12 April 2019
- Enders, C.K., Tofighi, D.: Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychol. Methods* **12**, 121–138 (2007)
- Fazio, R.H.: Multiple processes by which attitudes guide behavior: the mode model as an integrative framework. *Adv. Exp. Soc. Psychol.* **23**, 75–109 (1990)
- Giroux, S., Tharp, K., Wietelman, D.: Impacts of implementing an automatic advancement feature in mobile and web surveys. *Surv. Pract.* **12**, 1–12 (2019). <https://doi.org/10.29115/sp-2018-0034>
- Groves, R.M., Peytcheva, E.: The impact of nonresponse rates on nonresponse bias: a meta-analysis. *Public Opin. Q.* **72**, 167–189 (2008)
- Hamby, T., Taylor, W.: Survey satisficing inflates reliability and validity measures: an experimental comparison of college and amazon mechanical turk samples. *Educ. Psychol. Measur.* **76**, 912–932 (2016). <https://doi.org/10.1177/0013164415627349>
- Hammen, K.: The impact of visual and functional design elements in online survey research. Paper presented at the General Online Research Conference, Mannheim, Germany, 26–28 May 2010 (2010)

- Hays, R.D., Bode, R., Rothrock, N., Riley, W., Cella, D., Gershon, R.: The impact of next and back buttons on time to complete and measurement reliability in computer-based surveys. *Qual. Life Res.* **19**, 1181–1184 (2010)
- Heideman, S.G., Rohenkohl, G., Chauvin, J.J., Palmer, C.E., van Ede, F., Nobre, A.C.: Anticipatory neural dynamics of spatial-temporal orienting of attention in younger and older adults. *Neuroimage* **178**, 46–56 (2018). <https://doi.org/10.1016/j.neuroimage.2018.05.002>
- Hollender, N., Hofmann, C., Deneke, M., Schmitz, B.: Integrating cognitive load theory and concepts of human-computer interaction. *Comput. Hum. Behav.* **26**, 1278–1288 (2010)
- Kaminska, O., McCutcheon, A.L., Billiet, J.: Satisficing among reluctant respondents in a cross-national context. *Public Opin. Q.* **74**, 956–984 (2011)
- Kaplowitz, M.D., Lupi, F., Couper, M.P., Thorp, L.: The effects of invitation design on web survey response rates. *Soc. Sci. Comput. Rev.* **30**, 339–349 (2012)
- Kastner, S., Pinsk, M.A., De Weerd, P., Desimone, R., Ungerleider, L.G.: Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron* **22**, 751–761 (1999)
- Kim, Y., Dykema, J., Stevenson, J., Black, P., Moberg, D.P.: Straightlining: overview of measurement, comparison of indicators, and effects in mail–web mixed-mode surveys. *Soc. Sci. Comput. Rev.* **37**, 214–233 (2019). <https://doi.org/10.1177/0894439317752406>
- Kim, S.-U., Lim, S.-M., Kim, E.-A., Yang, I.-H.: An analysis of eye movement and cognitive load about the editorial design in elementary science textbook. *Univ. J. Educ. Res.* **4**, 570–575 (2016). <https://doi.org/10.13189/ujer.2016.040314>
- Klausch, T., de Leeuw, E.D., Hox, J.J., Roberts, A., de Jongh, A.: Matrix vs. single question formats in web surveys: results from a large scale experiment. Paper presented at the General Online Research Conference, Mannheim, 5–7 March 2012 (2012)
- Knowles, E.S.: Item context effects on personality scales: measuring changes the measure. *J. Pers. Soc. Psychol.* **55**, 312–320 (1988)
- Krejtz, K., Duchowski, A.T., Niedzielska, A., Biele, C., Krejtz, I.: Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLOS ONE* (2018). <https://doi.org/10.1371/journal.pone.0203629>
- Krosnick, J.A.: Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl. Cogn. Psychol.* **5**, 213–236 (1991)
- Krosnick, J.A.: Survey research. *Annu. Rev. Psychol.* **50**, 537–567 (1999)
- Krosnick, J.A., Alwin, D.F.: An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opin. Q.* **51**, 201–219 (1987). <https://doi.org/10.1086/269029>
- Krosnick, J.A., Alwin, D.F.: A test of the form-resistant correlation hypothesis: ratings, rankings, and the measurement of values. *Public Opin. Q.* **52**, 526–538 (1988)
- Leppink, J., Paas, F., Van der Vleuten, C.P.M., Van Gog, T., Van Merriënboer, J.J.G.: Development of an instrument for measuring different types of cognitive load. *Behav. Res. Methods* **45**(1058), 1072 (2013). <https://doi.org/10.3758/s13428-013-0334-1>
- Lipps, O.: Interviewer and respondent survey quality effects in a CATI panel. *Bull. Méthodologie Sociologique* **95**, 5–25 (2007)
- Makransky, G., Terkildsen, T.S., Mayer, R.E.: Role of subjective and objective measures of cognitive processing during learning in explaining the spatial contiguity effect. *Learn. Instr.* **61**, 23–34 (2019). <https://doi.org/10.1016/j.learninstruc.2018.12.001>
- Mayer, R.E.: Cognitive theory of multimedia learning. In: Mayer, R.E. (ed.) *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press, New York (2014)
- Mayer, R.E.: Instruction based on visualizations. In: Mayer, R.E., Alexander, P.A. (eds.) *Handbook of Research on Learning and Instruction*, 2nd edn. Routledge, New York (2017)

- Mayer, R.E., Fiorella, L.: Principles for reducing extraneous processing in multimedia learning: coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In: Mayer, R.E. (ed.) *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press, New York (2014)
- Mayer, R.E., Moreno, R.: Nine ways to reduce cognitive load in multimedia learning. *Educ. Psychol.* **38**, 43–52 (2003)
- Nobre, A.C., Stokes, M.G.: Premembering experience: a hierarchy of time-scales for proactive attention. *Neuron* **104**, 132–146 (2019)
- Paas, F., Renkl, A., Sweller, J.: Cognitive load theory and instructional design: Recent developments. *Educ. Psychol.* **38**, 1–4 (2003)
- Paas, F., Sweller, J.: An evolutionary upgrade of cognitive load theory: using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educ. Psychol. Rev.* **24**, 27–45 (2012). <https://doi.org/10.1007/s10648-011-9179-2>
- Paas, F., van Gog, T., Sweller, J.: Cognitive load theory: new conceptualizations, specifications, and integrated research perspectives. *Educ. Psychol. Rev.* **22**, 115–121 (2010)
- Paccagnella, O.: Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *Eval. Rev.* **30**, 66–85 (2006)
- Rivers, D.: Web surveys for health measurement. Paper presented at Building Tomorrow's Patient-Reported Outcome Measures: The Inaugural PROMIS Conference, Gaithersburg, MD, September 2006
- Selkälä, A., Couper, M.P.: Automatic versus manual forwarding in web surveys. *Soc. Sci. Comput. Rev.* **36**, 669–689 (2018). <https://doi.org/10.1177/0894439317736831>
- Selkälä, A., Reips, U.-D., Viinamäki, L., Suikkanen, A.: Demographics explaining a web survey entries election on the postal invitation letter. In: Conference Presentation, Session: Challenges and Opportunities of Switching to Web, Zagreb, Croatia, 15–19 July 2019. European Survey Research Association (2019)
- Simon, H.A.: *Models of Man*. Wiley, New York (1957)
- Stokes, M., Thompson, R., Nobre, A.C., Duncan, J.: Shape-specific preparatory activity mediates attention to targets in human visual cortex. *PNAS* **106**, 19569–19574 (2009)
- Sundar, S.S.: Social psychology of interactivity in human-website interaction. In: Joinson, A.N., McKenna, K., Postmes, T., Reips, U.-D. (eds.) *Oxford Handbook of Internet Psychology*. Oxford University Press, New York (2007)
- Theeuwes, J.: Spatial orienting and attentional capture. In: Nobre, A.C., Kastner, S. (eds.) *The Oxford Handbook of Attention*. Oxford University Press, New York (2014). <https://doi.org/10.1093/oxfordhb/9780199675111.013.005>
- Toepoel, V., Das, M., Van Soest, A.: Design of web questionnaires: the effects of the number of items per screen. *Field Methods* **21**, 200–213 (2009)
- Tourangeau, R.: Cognitive sciences and survey methods. In: Jabine, T., Straf, M., Tanur, J., Tourangeau, R. (eds.) *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. National Academy Press, Washington, DC (1984)
- Tourangeau, R.: The survey response process from a cognitive viewpoint. *Qual. Assur. Educ.* **26**, 169–181 (2018)
- Tourangeau, R., Conrad, F.G., Couper, M.P.: *The Science of Web Surveys*. Oxford University Press, New York (2013)
- Tourangeau, R., Rips, L.J., Rasinski, K.A.: *The Psychology of Survey Response*. Cambridge University Press, Cambridge (2000)
- Turner, G., Sturgis, P., Martin, D.: Can response latencies be used to detect survey satisficing on cognitively demanding questions? *J. Surv. Stat. Methodol.* **3**, 89–108 (2015)
- van Merriënboer, J.J.G., Kester, L., Paas, F.: Teaching complex rather than simple tasks: balancing intrinsic and germane load to enhance transfer of learning. *Appl. Cogn. Psychol.* **20**, 343–352 (2006). <https://doi.org/10.1002/acp.1250>

- van Merriënboer, J.J.G., Sweller, J.: Cognitive load theory and complex learning: recent developments and future directions. *Educ. Psychol. Rev.* **17**, 147–177 (2005)
- Vannette, D.L., Krosnick, J.A.: Answering questions: a comparison of survey satisficing and mindlessness. In: Ie, A., Ngnoumen, C.T., Langer, E.J. (eds.) *The Wiley Blackwell Handbook of Mindfulness*. Wiley, Chichester (2014). <https://doi.org/10.1002/9781118294895.ch17>
- Whitenton, K.: Minimize cognitive load to maximize usability (2013). <https://www.nngroup.com/articles/minimize-cognitive-load/>
- Yan, T., Olson, K.: Analyzing paradata to investigate measurement error. In: Kreuter, F. (ed.) *Improving Surveys with Paradata. Analytic Uses of Process Information*. Wiley, New York (2013)
- Yan, T., Tourangeau, R.: Fast times and easy questions: the effects of age, experience and question complexity on web survey response times. *Appl. Cogn. Psychol.* **22**, 51–68 (2008). <https://doi.org/10.1002/acp.1331>
- Zagermann, J., Pfeil, U., Reiterer, H.: Measuring cognitive load using eye tracking technology in visual computing. In: *BELIV 2016: Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pp. 78–85 (2016). <https://doi.org/10.1145/2993901.2993908>
- Zhang, C., Conrad, F.G.: Speeding in web surveys: the tendency to answer very fast and its association with straightlining. *Surv. Res. Methods* **8**, 127–135 (2014)
- Zhang, C., Conrad, F.G.: Intervening to reduce satisficing behaviors in web surveys: evidence from two experiments on how it works. *Soc. Sci. Comput. Rev.* **36** (2018). <https://doi.org/10.1177/0894439316683923>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

