FrontlineLearningResearch

It is all in the *surv-eye*: can eye tracking data shed light on the internal consistency in self-report questionnaires on cognitive processing strategies?

Margot Chauliac[a], Leen Catrysse[a], David Gijbels[a], Vincent Donche[a]

[a]University of Antwerp, Belgium

Abstract

Although self-report questionnaires are widely used, researchers debate whether responses to these types of questionnaires are valid representations of the respondent's actual thoughts and beliefs. In order to provide more insight into the quality of questionnaire data, we aimed to gain an understanding of the processes that impact the completion of self-report questionnaires. To this end, we explored the process of completing a questionnaire by monitoring the eye tracking data of 70 students in higher education. Specifically, we examined the relation between eye movement measurements and the level of internal consistency demonstrated in the responses to the questionnaire. The results indicated that respondents who look longer at an item do not necessarily have more consistent answering behaviour than respondents with shorter processing times. Our findings indicate that eye tracking serves as a promising tool to gain more insight into the process of completing self-report questionnaires.

Keywords: eye tracking; cognitive processes; survey research; self-report questionnaires; working memory capacity

1. Introduction

In self-report questionnaires respondents are asked to answer questions about themselves and as an instrument they are widely used to measure beliefs, attitudes, feelings and opinions in diverse fields of research (Singleton & Straits, 2009). This also holds for the domain of research on learning and instruction where self-report questionnaires are often used to map student learning. Important assets of these questionnaires are that they are easy to administer in both small and large groups and that their use is time and cost-effective. However, despite the reliability, validity and advantages self-report questionnaires might offer, a critical stance towards their use is required to gain more insight into students' cognitive processing strategies (Dinsmore & Alexander, 2012). Many researchers argue that respondents are, consciously or unconsciously, not always able to respond accurately to these questions (Schellings, 2011; Schellings & Van Hout-Wolters, 2011). This inability may influence the consistency by which a respondent scores the different items of a questionnaire, and thus the reliability of its outcomes (Richardson, 2004, 2013; Veenman, 2011; Veenman & van Hout-Wolters, 2005).

In order to assess the quality of the retrieved data, it is critical to examine whether the responses to questionnaires are valid representations of respondents' actual thoughts and beliefs (Schwarz, 2007; Tourangeau et al., 2000). Thus, when a specific set of items focuses on the same topic (i.e. a scale mapping a specific belief), the responses to these items need to be representative of the respondent's beliefs. An individual answering pattern on a set of items that is consistent with the underlying scale leads to reliable survey data. When this is not the case, one can start questioning how the respondent completed the survey and to what extent this is related to the consistency of their answers.

Generally, there is a black box concerning the processes in participants' completion of self-report questionnaires. Gaining an understanding of these processes could help to provide additional insight into the quality of survey data. However, this area of interest, and in particular, the process of completing the questionnaires, has been under-examined in the literature so far. In this study, we use eye tracking to examine the processes at play when completing self-report questionnaires that aim to map students' cognitive processing strategies. In particular, we focus on the specific strategies that students use while processing items.

2. Theoretical Framework

2.1. Cognitive processing when completing self-report questionnaires

Surveys have a long history in educational research (Marsden & Wright, 2010; Rossi et al., 1983). Despite its long history, it is only from 1980 onwards that cognitive psychology started to enter the field of survey research. The focus shifted from the outcomes of the questionnaire to the cognitive processing activities that were at play when completing questionnaires (Fowler, 2014; Willis & Miller, 2011). However, despite the development of the cognitive aspects of survey methodology (CASM), the focus was still on examining how cognitive processes could influence the outcomes of the questionnaires, instead of investigating how the underlying processes while completing self-report questionnaires could be related to the reliability of their outcomes.

Following the CASM-movement, multiple theoretical models were developed to grasp the processes at work in the reading of questions and providing answers to these questions (Jobe & Herrmann, 1996): these included the four-stage model by Tourangeau (1984), the autobiographical question-answering model from Schwarz (1990), the flexible processing model by Willis et al. (1991) and the information processing model of self-report item response (Karabenick et al., 2007). All these models share the common feature of attempting to grasp the complexity of completing survey questionnaires by distinguishing important stages that respondents go through in order to generate an

answer. Most researchers agree that respondents must give meaning to a question and be able to retrieve necessary information from their memory. Only then they will be able to make an informed decision and choose a congruent response option (Karabenick et al., 2007; Tourangeau, 1984). All these models are characterised by the fact that the described stages do not have to follow each other in a linear sequence. One can move back and forth so that there may be iterations and overlap between the steps. It is even possible that one or some of the steps are weakly conducted or completely missing. Nevertheless, one can only expect substantive answers when respondents thoroughly conduct all cognitive processes when answering a question (Krosnick, 1991; Krosnick & Alwin, 1987).

In the field of survey research, the comprehension of the question is an important prerequisite for achieving meaningful results. A crucial step is, therefore, to design a questionnaire such that all respondents understand the items in the same way as the researcher intended (Neuert, 2016). Previous research already demonstrated that comprehension problems could arise or that respondents may satisfice while completing the survey (Krosnick & Alwin, 1987). Another important factor that plays a role is the capacity of a respondent's working memory. Working memory concerns the limited amount of information that can be processed and temporarily stored in the memory while performing complex cognitive tasks (Baddeley & Hitch, 1974). Krosnick (1991) argues that working memory is limited and that respondents are unable to give the latter options as much attention as the ones they consider initially. Moreover, respondents may differ in cognitive ability to complete survey questions. This can influence the eventual results (Gathercole & Alloway 2013; Krosnick, 1991).

Until now, a lot of research has been done to gain more insight into the problems that might arise while completing self-report questionnaires (Galesic et al., 2008; Graesser et al., 2006; Lenzner et al., 2011). In the past, researchers made use of cognitive interviewing techniques such as think-aloud protocols and verbal probing to get a grip on the difficulties that might arise (Collins, 2003). The think-aloud protocol is a data-gathering method in which respondents are asked to verbalise their thought processes during or after doing a specific task. Verbal probing is a cognitive interviewing technique where questions are designed to elicit specific information that is usually not provided by respondents. These cognitive interviews provide a suitable methodology for examining the extent to which tools of inquiry capture the experiences of students in a valid and reliable manner (Beatty & Willis, 2007; Desimone & Le Floch, 2004; Presser et al., 2004). However, despite the benefits cognitive interviews have to offer, they do not allow researchers to look directly into processing behaviour while respondents complete in the questionnaire.

2.2. Eye tracking as an eye-opener in survey research

Previous research has shown that eye tracking can help gain more insight into the black box of the processes of completing self-report questionnaires (Galesic et al., 2008; Redline & Lankford, 2001). Via this relatively unobtrusive instrument, one can track the implicit processes at play while completing questionnaires. Eye tracking research has a long tradition in studying cognitive processing during reading and other information processing tasks (Duchowski, 2007; Neuert, 2016; Rayner, 1998). More recently, the technique has also been introduced into the field of survey methodological research to study cognitive processes while answering survey questions (Lenzner et al., 2010; Neuert, 2016).

In previous research, eye tracking has been used to study, among other topics, the visual designs of branching instructions (Redline & Lankford, 2001), different response formats (Lenzner et al., 2014), response order effects (Galesic et al., 2008), the effects of question wording (Graesser et al., 2006; Lenzner et al., 2011) and the cognitive processes associated with answering rating scale questions (Menold et al., 2014). However, these were mainly experimental studies that focused on the aspects of the questionnaire that could lead to difficulties in processing. By investigating the potential burden the questions might bring, one focuses on the possible constraints of the survey. However, the effects these difficulties have on the actual process of completing the questionnaire have not yet been addressed.

The relationship between eye movements and cognitive processing is based on two assumptions: the immediacy assumption and the eye-mind assumption. The immediacy assumption states that a visual stimulus on which the eyes fixate is processed immediately. The eye-mind assumption postulates that as long as the stimulus is fixated, it is mentally processed. Thus, both assumptions suggest that eye

movements provide direct information about what is processed and the amount of cognitive effort that is involved (Just & Carpenter, 1980).

Although eye tracking cannot help in making a concrete distinction between the different stages respondents might go through when completing self-report questionnaires, it does provide insight into the entire process that evolves in the time period between the reading of the stimulus — in this case, the self-report question — and the giving of an answer. The duration a respondent spends processing gives an indication of the cognitive effort the respondent put into the processing. Longer fixation times could, for example, be associated with a deeper and more effortful cognitive processing or may be an indicator of comprehension problems (Holmqvist et al., 2011).

2.3. Present research

In this study, we will include eye tracking as an online measure in order to gain an understanding of the cognitive processes that are active while completing self-report questionnaires. By taking a closer look at eye tracking data, we strive to examine whether the underlying processes are possibly explanatory indicators of the internal consistency by which the respondent completed the questionnaire. After all, internal consistency is one of the most critical prerequisites in obtaining meaningful results from survey data. Consistency is determined by how similar a respondent answers questions that belong to the same scale. Our study aims to answer the following two research questions:

> 1. To what extent is there a relation between the consistency in answering behaviour and eye movement measures when completing a self-report questionnaire?
> 2. To what extent is there a relationship between the consistency in answering behaviour, eye movement measurers and a respondent's working memory capacity when completing a self-report questionnaire?

In the assumption that the time a respondent spends fixating on an area of the survey item more or less corresponds to the time this area is processed (Staub & Rayner, 2007), the time taken to choose an answering option can be an indicator of the cognitive effort that was invested in arriving at this answer or judgment (Fazio, 1990). Therefore, we hypothesise that there could be a link between the cognitive processing taking place when scoring the items of a self-report questionnaire and the internal consistency of the scales. Based on the previous findings on working memory capacity (Krosnick, 1991), we expect an interplay between students' working memory capacity and the cognitive process taking place when completing a questionnaire.

3. Methodology

3.1. Participants

The sample consisted of 92 bachelor students from a social science faculty. Students were recruited during regular lectures and all participated on a voluntary basis. Before the start of the experiment, we received their consent, which was approved by the ethics committee for social sciences and humanities of the participating university. All participants had a normal or corrected-to-normal vision and had Dutch as their native language. Due to issues that are common in eye tracking research (e.g. problems with the calibration of the eye tracker, and a lack of responses to the survey questions [see e.g. Holmqvist et al., 2011]) we lost data from 22 respondents. Data from 10 respondents were excluded due to technical issues; data from 12 respondents were left out because of poor quality of the eye tracking data. After this data cleaning, the data from 70 participants were included in the statistical analyses. To thank the students for their participation, they received two cinema tickets.

3.2. Materials and procedure

The self-report questionnaire data were collected as part of a larger project about learning from texts and the completing of questionnaires where we recorded eye movements to gain insight into the processing behaviour of participants. After the reading of each text, a validated self-report questionnaire was completed to measure students' task-specific processing strategies. A task-specific version of the ILS-SV questionnaire was developed based on the original version (Donche & Van Petegem, 2008; Vermunt & Donche, 2017). This version contained four scales about cognitive processing strategies and consisted of sixteen items that mapped how participants process information when reading a particular text. Students had to read the question, select the answering category of their choice, and state their answer out loud. Answering options ranged from 1 = 'I rarely or never do this' to 5 = 'I almost always do this'. All survey items were answered consecutively without the possibility of changing the given answer.

Apart from completing the self-report questionnaire, the students' working memory capacity was measured by means of the Automated Operation Span Task (Aospan). According to Unsworth et al. (2005), the Aospan is a reliable and valid test for measuring the working memory capacity that can be used in various research domains. Participants were required to solve a series of mathematical operations while trying to retain a set of unrelated letters. The Aospan is mouse-driven, calculates scores automatically and requires little to no intervention from the experimenter (Unsworth et al., 2005). In order to be sure that participants were not only focusing on remembering the letters, a 85% accuracy criterion was imposed for solving the mathematical problems (Unsworth et al., 2005). The Aospan provides two scorings, an absolute credit scoring and a partial credit scoring. Since partial credit scoring is preferred over the absolute all-or-nothing scoring, we made use of the latter (Conway et al., 2005). The mean score for all respondents was 60.11 ($SD = 9.85$). The score for this working memory capacity test was normally distributed, and for further analysis, we made use of standardised scores.

3.3. Eye tracking equipment

To measure students' eye movements, we made use of the Tobii Pro X3-120 eye tracker, which alternates between bright and dark pupil eye tracking in a predefined, systematic way. This eye tracker had a sampling frequency of 120 Hz (binocularly), which made it possible to take a closer look at the fixation durations. The eye tracker was secured to a 17.3-inch monitor with a resolution of 1.920 x 1.080 pixels. Every participant sat at about 60 cm from the screen and the eye tracker. To minimise the influence of student movement, we employed a chinrest. Tobii Technology (Stockholm, Sweden) reported a gaze accuracy of 0.4°, gaze precision of 0.24° and a total system latency of fewer than 11 milliseconds for this eye tracker. The eye movements were recorded with Tobii-Studio (3.4.8) software.

3.4. Consistency in response behaviour

The first indication of consistency in response behaviour is the Cronbach Alpha coefficient, which was calculated for each of the four scales. The consistency levels for the four scales were .67 .68, .69 and .65, respectively (Table 1). These results show an acceptable internal consistency level for four-item scales. Since only a small number of items are used per scale, and given the sensitivity of the Cronbach's Alpha for the number of items, a cut-off value of .60 is considered sufficient (Cortina, 1993; Pallant, 2007). As respondents can differ in the way they score the separate items of a specific scale, thus showing diversity in scoring behaviour across items, we categorised their rating behaviour for each scale. For all respondents, four consistency indicators were created (one for each scale), making distinctions between raters using the same answering category for all items on a scale or raters showing more diversity in the use of answering categories, by making, for instance, use of at least two different answering categories. The consistency indicator ranged from 1 (consistent answering pattern) to 4 (very diverse answering pattern). The questionnaire did not include reversed items, so this could not serve as an explanation for diversity in answering categories.

Table 1

ILS-SV scales, number of items, item examples and reliability (internal consistency)

Scale	Items	Item example	Cronbach's Alpha
Relating and structuring	4	I compare conclusions from different teaching modules with each other.	.67
Critical processing	4	I try to understand the interpretations of experts in a critical way.	.68
Analysing	4	I study each course book chapter point by point and look into each piece separately.	.69
Memorising	4	I learn definitions by heart and as literally as possible.	.65

3.5. Analysis

We used the Tobii fixation filter for fixation identification, which is an implementation of a classification algorithm proposed by Olsson (2007). It uses a velocity threshold (35 pixels per window) and a distance threshold (35 pixels) (Olsen, 2012).

Eye movement data were analysed at the item level. The question and response field for each item in the survey were considered as a combined area of interest (AOI). For each AOI or item in the survey, the total fixation duration and the total fixation count were calculated separately. To control for the length of AOI's, the total fixation duration measure was normalised by calculating a milliseconds-per-character measure (Ariasi et al., 2017; Catrysse et al., 2016; Yeari et al., 2016). The total fixation count measure was normalised by calculating a count-per-character measure. In addition, we logarithmically transformed these measures because they are heavily skewed (Catrysse et al., 2018; Holmqvist et al., 2011; Lo & Andrews, 2015). To check the distribution of the dependent measures, the fitdistrplus package was used (Delignette-Muller & Dutang, 2015). The eye movement data were analysed with linear mixed effects models (LMM) with the lme4 package (Bates et al., 2015) in R (R Core Team, 2014) and with the Rstudio interface. Mixed-effects models are statistical models that incorporate random and fixed effects (Baayen, 2008; Baayen et al., 2008). Subjects, items and scales were considered as crossed random effects (Baayen, 2008; Baayen et al., 2008). The analysis was conducted at the item level and was based on 1,120 data points (70 students each processing 16 items).

Separate models were fitted for the total fixation duration and the total fixation count. Two models per measure were fitted: (1) an LMM with subjects, subscales and items as random effects and consistency in answering behaviour as a fixed effect and (2) an LMM with subjects, subscales and items as random effects and consistency in answering behaviour and working memory capacity as fixed effects. The interactions between the fixed effects were also incorporated into the second model.

4. Results

4.1. The relation between consistency in answering behaviour and eye movement measures

In order to answer the first research question, we report the means and standard deviations for the eye movement measures in Table 2 in relation to the consistency in answering behaviour. For example, students who were very consistent in their answering behaviour on a certain scale, that is, choosing the same answering option for each item, looked on average 8.74 seconds at an item and the

corresponding answering options, and made on average 31.99 fixations on an item and answering options.

Table 2

Descriptive statistics for the number of different answering options per scale in relation to the eye movement measures

	Consistency level indicator 1		Consistency level indicator 2		Consistency level indicator 3		Consistency level indicator 4	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Total fixation duration (seconds)	8.74	4.56	9.02	4.70	8.95	5.21	8.75	5.80
Total fixation count	31.99	16.30	34.15	17.65	34.49	19.99	32.97	22.72

Note: Untransformed eye movement measures reported.

In the next step, we examined the relation between the consistency in answering behaviour and eye movement measures. We analysed the data with linear mixed effect models. For the total fixation duration, the parameter estimates indicated that there was a significant effect of consistency in answering behaviour on the total fixation duration for an item (Table 3). More specifically, the results showed that a student who chose two or three different answering categories looked longer at an item than a student who only opted for one answering category. A student who chose four categories did not look longer at the items than a student who chose only one answering category.

Overall, the parameter estimates showed that students with less consistent scoring behaviour spend more time on processing the items and answering options. However, this was not the case for students who picked four different answering options on a scale. This implies that there seems to be a turning point in the effect of correlation between consistency in answering behaviour and students' eye movement measures.

Table 3

Parameter estimates of the random and fixed effects for the random intercept model for total fixation duration and total fixation count

	Total fixation duration				Total fixation count			
Random effects	Variance	*SD*			Variance	*SD*		
Subject	.05	.23			.05	.23		
Item	.04	.20			.03	.18		
Subscale	.00	.001			.00	.00		
Residual	.09	.30			.10	.32		
Fixed effects	*β*	*SE*	*t*	*pr(>|t|)*	*β*	*SE*	*t*	*pr(>|t|)*
Intercept	4.68	.06	-69.53	**<.001**	-.90	.07	-13.88	**<.001**
Consistency level indicator 2	.09	.04	2.17	**.03**	.10	.04	2.39	**.02**
Consistency level indicator 3	.11	.04	2.68	**.007**	.12	.04	2.75	**.006**
Consistency level indicator 4	.09	.06	1.57	.12	.10	.06	1.60	.11

Note: Significant values are in bold.

For the total fixation count, the estimate of the intercept had a negative value of -0.90. This is due to the log transformation of the count-per-character measure, which causes small values (<1) to turn into negative values. Moreover, we were mainly interested in the potential change in the fixation count, rather than in its absolute value. Therefore, this negative value was not problematic for the interpretation of our results.

The results for the fixation count are similar as for the total fixation duration. A student who chose two or three answering categories made more fixations on an item than a student opting for only one answering category. A student who chose four categories did not make more fixations on the items than a student who picked only one answering category.

4.2. The relation between consistency in answering behaviour, working memory capacity and eye movement measures

To answer the second research question on the relationship between consistency in answering behaviour, working memory capacity and eye movement measures, we updated the mixed effects model of Table 5 and added working memory capacity as a fixed effect in a new model (Table 4). Both for the total fixation duration and total fixation count, we did not find any significant effect of working memory capacity. We can thus conclude that working memory capacity in this study has no interference with students' eye movement measures when completing this self-report questionnaire.

Table 4

Parameter estimates of the random and fixed effects for the random intercept model for total fixation duration and total fixation count including working memory capacity

	Total fixation duration				Total fixation count			
Random effects	Variance	SD			Variance	SD		
Subject	.05	.23			.05	.23		
Item	.04	.20			.03	.18		
Subscale	.00	.00			.00	.00		
Residual	.09	.30			.10	.32		
Fixed effects	β	SE	t	pr(>|t|)	β	SE	t	pr(>|t|)
Intercept	4.68	.07	69.24	**<.001**	-.91	.07	-13.89	**<.001**
WMC	.02	.05	.52	.60	.01	.05	.23	.82
2 AO	.09	.04	2.26	**.02**	.10	.04	2.46	**.01**
3 AO	.11	.04	2.76	**.006**	.12	.04	2.81	**.005**
4 AO	.09	.06	1.50	.13	.09	.06	1.48	.14
WMC*2 AO	-.05	.04	-1.14	.25	-.05	.04	-1.12	.26
WMC*3 AO	-.04	.04	-.93	.35	-.04	.04	-.87	.38
WMC*4 AO	-.07	.05	-1.40	.16	-.09	.05	-1.63	.10

Note: AO: Answering option(s) — WMC: working memory capacity — significant values are in bold.

5. Discussion

Although self-report questionnaires are widely used to map students' processing strategies, there still is a lacuna in the knowledge about the processes at play while respondents complete these questionnaires. By gaining insight into these processes, we want to provide evidence for the debate about the often-reported reliability issues of self-report questionnaires (Richardson, 2004, 2013; Veenman, 2011; Veenman & van Hout-Wolters, 2005). In this exploratory study, we used eye tracking in order to unobtrusively track the processes that are at play while completing a self-report questionnaire

on cognitive processing strategies. Previous research mainly focused on cognitive difficulties that might arise when processing the questionnaire, and therefore the questionnaire's potential limitations to accurately grasp respondents' opinions and beliefs (Galesic et al., 2008; Graesser et al., 2006; Lenzner et al., 2014; Menold et al., 2014; Redline & Lankford, 2001). How these difficulties affect the process of completing the questionnaire, and how they influence the questionnaire's reliability, are two questions that have not been addressed before.

Based on previous research stating that the time a respondent spends fixating on a specific area is more or less equal to the time this area is being processed, the processing time is assumed to be a good indicator of the invested cognitive effort (Fazio, 1990; Staub & Rayner, 2007). Therefore, we believe that there could be a link between the cognitive processing taking place when scoring the items of a survey and the internal consistency of the scored scales from a questionnaire. This concept of consistency is important, given the fact that when a respondent's answering behaviour is not consistent, one can thus start questioning the reliability of the survey data. We first examined the relationship between the consistency in answering behaviour and eye movement measures. Our results demonstrate that the consistency in answering behaviour is significantly related to the total fixation duration for an item. The more a respondent's answers differ in one scale, the longer the respondent looks at the items compared to those who only opt for one answering option. However, no significant difference was found between the respondents choosing one response option and the ones opting for four different answers for items belonging to the same scale. Given these results, there seems to be a turning point in the effect of consistency in answering behaviour. Results suggest that too much pondering over a question does not lead directly to a more consistent answering behaviour. On the contrary, when the respondents spend more time processing a question, they might be trying to process the question more thoroughly to come to an appropriate and thus consistent response, but they just do not succeed in doing so.

Secondly, we aimed to gain more insight into the relation between eye movement measures, answering behaviour and the working memory capacity of the respondent. Previous research on the working memory demonstrated that its capacity is limited and that respondents may therefore not give each answering option as much attention as the one they considered initially (Gathercole & Alloway 2013; Krosnick, 1991). Therefore, we hypothesised that we would find less consistent answering behaviour for the students with a lower working memory capacity. However, both for the total fixation duration as well as for the total fixation count, we did not find any significant effect of working memory capacity. This could be because memory distortions do not play a significant role when this self-report questionnaire is being completed immediately after completing the task that the questionnaire referred to.

6. Limitations and directions for future research

Although our findings show that eye tracking is a promising technique to gain more insight into the process of completing self-report questionnaires, we want to emphasise the exploratory nature of this study and point at some limitations and directions for future research.

The process of completing questionnaires is an extremely complex process. Different theoretical models try to distinguish different stages that possibly play a role when a respondent is cognitively processing a question (see e.g. Karabenick et al., 2007; Tourangeau, 1984). In our study, we considered the question as well as the answering options as one area of interest. This choice allows for an indication of the total time taken until one decides and thus completes the process of filling in the item. More specifically, by focussing on the survey item in its entirety, we took all stages of the different theoretical models into account. In future research, it would be interesting to separate this area into two distinct areas of interest — the question and the answering options — in order to investigate which possible influence each of these areas has on the internal consistency. This would also allow us to further separate the different stages of the theoretical models. However, as these stages do not follow a linear path, separating into different areas of interest will lead to a loss of information. When analysing the question, we could, for example, consider whether different reading processes lead to different outcomes in internal consistency. Hereto, it would also be important to take other eye tracking measures into account.

In our study, we made use of the total fixation duration and the fixation count to map the whole process. However, analysing merely the question would allow us to use other measures such as first pass fixations and second pass fixations (Hyönä et al., 2003; Jarodzka & Brand-Gruwel, 2017) which could possibly shed some light on further difficulties the respondents encountered. Next to looking at the question itself, it could also be clarifying to look at how the respondent processes the different the answering options. The way in which a respondent ponders over a question — merely focusing on one answering option or considering each of the five possibilities — could potentially elucidate their answering behaviour.

Another constraint of this study is that no use was made of complementary data. The use of merely eye tracking data might not provide us with the necessary insight into the reasons why students who respond in a less consistent way take more time to respond to the items, whereas using a multi-method approach to look at the data could help us put different pieces of the puzzle together (Catrysse et al., 2018). As we already know from previous research, a longer reading time can be an indication of several different cognitive processes such as (1) high-level or deeper cognitive processing (Ariasi & Mason, 2011; Holmqvist et al., 2011; Penttinen et al., 2013), (2) strategic attempts to resolve comprehension problems or further text comprehension (Ariasi et al., 2017; Hyönä & Lorch, 2004; Hyönä et al., 2002; Hyönä et al., 2003; Kinnunen & Vauras, 1995), (3) comprehension monitoring (van Gog & Jarodzka, 2013), (4) difficulty with text passages (Rayner et al., 2006) and (5) attempts to reinstate information into working memory in order to elaborate or rehearse that information (Hyönä & Lorch, 2004). However, further research is needed to investigate whether the current insight in the field of text reading also hold for the process of completing survey questionnaires.

A last observation is that when completing the questionnaire, respondents were asked to state their given answer out loud after every question. Knowing that researchers were monitoring their answers could possibly have had an influence on the natural process of completing the questionnaire. For future research, it would therefore be necessary to look at the processes that are at play without verifying for the responses given by respondents. Moreover, it was impossible for the respondents to change their answer on certain questions. Once they provided an answer, the next question was immediately projected without an opportunity for the respondent to change their mind. Considering this in future research, one will be able to search for doubts and changes in the answering process.

7. Conclusions

Notwithstanding certain limitations, our exploratory study was able to show that eye tracking offers important research perspectives that helped us gain more insight into the cognitive processes at play in the process of completing a self-report questionnaire. It also gave us insight into how these processes are related to the consistency by which the survey has been completed. By lifting a corner of the veil that lies over survey research, we now not only know that a longer processing time is not necessarily linked to more consistent answering behaviour, but also that there is a turning point in which longer processing does not lead to more consistency in answering behaviour.

Key points

- The use of eye tracking to record the process of completing self-report questionnaires appears to be a promising tool to gain more insight herein.

- Respondents who look longer at the item in question do not necessarily have more consistent answering behaviour than respondents who spend less time answering the questions.

- When answering self-report questionnaires, there seems to be a turning point in which a longer focus on the item does not lead to a more consistent answering pattern.

References

Ariasi, N., Hyönä, J., Kaakinen, J., & Mason, L. (2017). An eye-movement analysis of the refutation effect in reading science text. *Journal of Computer Assisted Learning, 33*(3), 202-221. https://doi.org/10.1111/jcal.12151

Ariasi, N., & Mason, L. (2011). Uncovering the effect of text structure in learning from a science text: An eye-tracking study. *Instructional science, 39*(5), 581-601. https://doi.org/10.1007/s11251-010-9142-5

Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language, 59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47-89): Academic Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Beatty, P., & Willis, G. (2007). Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly, 71*(2), 287-311. https://doi.org/10.1093/poq/nfm006

Catrysse, L., Gijbels, D., & Donche, V. (2018). It is not only about the depth of processing: What if eye am not interested in the text? *Learning and Instruction, 58*, 284-294. https://doi.org/10.1016/j.learninstruc.2018.07.009

Catrysse, L., Gijbels, D., Donche, V., De Maeyer, S., Van den Bossche, P., & Gommers, L. (2016). Mapping processing strategies in learning from expository text: an exploratory eye tracking study followed by a cued recall. *Frontline learning research, 4*(1), 1-16. https://doi.org/10.14786/flr.v4i1.192

Collins, D. (2003). Pretesting survey instruments: an overview of cognitive methods. *Quality of life research, 12*(3), 229-238. https://doi.org/10.1023/A:1023254226592

Conway, A., Kane, M., Bunting, M., Hambrick, D., Wilhelm, O., & Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786. https://doi.org/10.3758/bf03196772

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104. https://doi.org/10.1037/0021-9010.78.1.98

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software, 64*(4), 1-34. https://doi.org/10.18637/jss.v064.i04

Desimone, L., & Le Floch, K. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational evaluation policy analysis, 26*(1), 1-22. https://doi.org/10.3102/01623737026001001

Dinsmore, D., & Alexander, P. (2012). A Critical Discussion of Deep and Surface Processing: What It Means, How It Is Measured, the Role of Context, and Model Specification. *Educational psychology review, 24*(4), 499-567. https://doi.org/10.1007/s10648-012-9198-7

Donche, V., & Van Petegem, P. (2008). The validity and reliability of the short inventory of learning patterns. In E. Cools, H. van den Broeck, & T. Redmond (Eds.), *Style and cultural differences: how can organisations, regions and countries take advantage of style differences* (pp. 49-59). Ghent: Vlerick Leuven Ghent Management School.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*. London: Springer.

Fazio, R. (1990). Multiple Processes by which Attitudes Guide Behavior: The Mode Model as an Integrative Framework. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23, pp. 75-109). New York: Academic Press.

Fowler, F. (2014). *Survey research methods - 5th edition*. Thousand Oaks: Sage publications.

Galesic, M., Tourangeau, R., Couper, M., & Conrad, F. (2008). Eye-tracking data: New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly, 72*(5), 892-913. https://doi.org/10.1093/poq/nfn059

Gathercole, S., & Alloway , T. (2013). *De invloed van het werkgeheugen op het leren: Handelingsgerichte adviezen voor het basisonderwijs*. Amsterdam: SWP, Amsterdam.

Graesser, A., Cai, Z., Louwerse, M., & Daniel, F. (2006). Question Understanding Aid (QUAID) a web facility that tests question comprehensibility. *Public Opinion Quarterly, 70*(1), 3-22. https://doi.org/10.1093/poq/nfj012

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Hyönä, J., & Lorch, R. (2004). Effects of topic headings on text processing: Evidence from adult readers' eye fixation patterns. *Learning and Instruction, 14*(2), 131-152. https://doi.org/10.1016/j.learninstruc.2004.01.001

Hyönä, J., Lorch, R., & Kaakinen, J. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology, 94*(1), 44-55. https://doi.org/10.1037//0022-0663.94.1.44

Hyönä, J., Lorch, R., & Rinck, M. (2003). Eye Movement Measures to Study Global Text Processing. In R. Hyönä (Ed.), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp. 313-334). Amsterdam: Elsevier Science.

Jarodzka, H., & Brand-Gruwel, S. (2017). Tracking the reading eye: towards a model of real-world reading. *Journal of Computer Assisted Learning, 33*(3), 193-201. https://doi.org/10.1111/jcal.12189

Jobe, J., & Herrmann, D. (1996). Implications of models of survey cognition for memory theory. *Basic applied memory research, 2*, 193-205. https://doi.org/10.1023/A:1023279029852

Just, M., & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review, 87*(4), 329. https://doi.org/10.1037/0033-295X.87.4.329

Karabenick, S., Woolley, M., Friedel, J., Ammon, B., Blazevski, J., Bonney, C., . . . Kempler, T. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist, 42*(3), 139-151. https://doi.org/10.1080/00461520701416231

Kinnunen, R., & Vauras, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning and Instruction, 5*(2), 143-165. https://doi.org/10.1016/0959-4752(95)00009-R

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

Krosnick, J., & Alwin, D. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly, 51*(2), 201-219. https://doi.org/10.1086/269029

Lenzner, T., Kaczmirek, L., & Galesic, M. (2011). Seeing through the eyes of the respondent: An eye-tracking study on survey question comprehension. *International Journal of Public Opinion Research, 23*(3), 361-373. https://doi.org/10.1093/ijpor/edq053

Lenzner, T., Kaczmirek, L., & Galesic, M. (2014). Left Feels Right: A Usability Study on the Position of Answer Boxes in Web Surveys. *32*(6), 743-764. https://doi.org/10.1177/0894439313517532

Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology, 24*(7), 1003-1020. https://doi.org/10.1002/acp.1602

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using Generalized Linear Mixed Models to analyse reaction time data. *Frontiers in Psychology, 6*, 1171. https://doi.org/10.3389/fpsyg.2015.01171

Marsden, P., & Wright, J. (2010). *Handbook of survey research - 2nd edition*. Bingley: Emerald Group Publishing.

Menold, N., Kaczmirek, L., Lenzner, T., & Neusar, A. (2014). How do respondents attend to verbal labels in rating scales? *Field Methods, 26*(1), 21-39. https://doi.org/10.1177/1525822X13508270

Neuert, C. (2016). *Eye tracking in questionnaire pretesting.*

Olsen, A. (2012). The Tobii I-VT fixation filter.

Olsson, P. (2007). Real-time and offline filters for eye tracking.

Pallant, J. (2007). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows - 3th edition*: Maidenhead: Open University Press.

Penttinen, M., Anto, E., & Mikkilä-Erdmann, M. (2013). Conceptual change, text comprehension and eye movements during reading. *Research in Science Education, 43*(4), 1407-1434. https://doi.org/10.1007/s11165-012-9313-2

Presser, S., Couper, M., Lessler, J., Martin, E., Martin, J., Rothgeb, J., & Singer, E. (2004). Methods for Testing and Evaluating Survey Questions. *Public Opinion Quarterly, 68*(1), 109-130. https://doi.org/10.1093/poq/nfh008

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372-422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K., Chace, K., Slattery, T., & Ashby, J. (2006). Eye Movements as Reflections of Comprehension Processes in Reading. *Scientific Studies of Reading, 10*(3), 241-255. https://doi.org/10.1207/s1532799xssr1003_3

Redline, C. D., & Lankford, C. (2001). Eye-movement analysis: a new tool for evaluating the design of visually administered instruments (paper and web). *Proceedings of the Survey Research Methods Section of the American Statistical Association*.

Richardson, J. (2004). Methodological Issues in Questionnaire-Based Research on Student Learning in Higher Education. *Educational psychology review, 16*(4), 347-358. https://doi.org/10.1007/s10648-004-0004-z

Richardson, J. (2013). Research issues in evaluating learning pattern development in higher education. *Studies in Educational Evaluation, 39*(1), 66-70. https://doi.org/10.1016/j.stueduc.2012.11.003

Rossi, P., Wright, J., & Anderson, A. (1983). *Handbook of survey research. Sample surveys: History, current practice, and future prospects*. San Diego: Academic Press.

Schellings, G. (2011). Applying learning strategy questionnaires: problems and possibilities. *Metacognition and Learning, 6*(2), 91-109. https://doi.org/10.1007/s11409-011-9069-5

Schellings, G., & Van Hout-Wolters, B. (2011). Measuring strategy use with self-report instruments: theoretical and empirical considerations. *Metacognition and Learning, 6*(2), 83-90. https://doi.org/10.1007/s11409-011-9081-9

Schwarz, N. (1990). Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction. In *Research methods in personality and social psychology.* (pp. 98-119). Thousand Oaks, CA, US: Sage Publications, Inc.

Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology, 21*(2), 277-287. https://doi.org/10.1002/acp.1340

Singleton, R., & Straits, B. (2009). *Approaches to social research - 5th edition*. Oxford: Oxford University Press.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics*: Oxford University Press.

Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73-100). Washington, DC: National Academy Press.

Tourangeau, R., Rips, L., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Unsworth, N., Heitz, R., Schrock, J., & Engle, R. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498-505. https://doi.org/10.3758/bf03192720

van Gog, T., & Jarodzka, H. (2013). Eye Tracking as a Tool to Study and Enhance Cognitive and Metacognitive Processes in Computer-Based Learning Environments. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 143-156). New York: Springer.

Veenman, M. (2011). Alternative assessment of strategy use with self-report instruments: a discussion. *Metacognition and Learning, 6*(2), 205-211. https://doi.org/10.1007/s11409-011-9080-x

Veenman, M., & van Hout-Wolters, B. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artelt & B. Moschner (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. 77-99): Münster: Waxmann.

Vermunt, J., & Donche, V. (2017). A learning patterns perspective on student learning in higher education: state of the art and moving forward. *Educational psychology review, 29*(2), 269-299. https://doi.org/10.1007/s10648-017-9414-6

Willis, G., & Miller, K. (2011). Cross-cultural cognitive interviewing: Seeking comparability and enhancing understanding. *Field Methods, 23*(4), 331-341. https://doi.org/10.1177/1525822X11416092

Willis, G., Royston, P., & Bercini, D. (1991). The use of verbal report methods in the development and testing of survey questionnaires. *Applied Cognitive Psychology, 5*(3), 251-267. https://doi.org/10.1002/acp.2350050307

Yeari, M., Oudega, M., & van den Broek, P. (2016). The effect of highlighting on processing and memory of central and peripheral text information: evidence from eye movements. *Journal of Research in Reading, 40*(4), 365-383. https://doi.org/10.1111/1467-9817.12072