Repeatedly Measuring Political Interest: Can We Reduce Respondent's Recall Ability and Memory Effects in Surveys Using Memory Interference Tasks?

Melanie Revilla
Universitat Pompeu Fabra (Spain)

Jan Karem Höhne
University of Mannheim (Germany)
Universitat Pompeu Fabra (Spain)

Abstract

It is common practice in behavioural and social sciences to repeat questions, such as in pretest-posttest designs. However, if respondents recall their first answer to a repeated question and use it to decide their second answer, this may cause memory effects, affecting empirical findings. So far, only a few studies have investigated respondents' recall ability and memory effects in surveys. Thus, we conducted an experiment in a probability-based online panel (N = 4,681) where we repeatedly asked a question on political interest. Our results reveal that respondents' recall ability is high and was not reduced by the implementation of memory interference tasks. Memory effects, in contrast, were low (about 7%). They were also not reduced by memory interference tasks.

Keywords: Memory effects, memory interference task, online survey experiment, political interest, recall ability, repeated survey measurements

Introduction

Repeated measurements in surveys are commonly used to estimate the reliability (test-retest) and measurement quality of questions (multitrait-multimethod designs), to evaluate the impact of a treatment (pretest-posttest experiments), or to monitor change over time (panel designs). However, results with repeated measurements can be biased if respondents recall their first answer when answering the repeated question, instead of accomplishing the information retrieval anew (Tourangeau, Rips, & Rasinski, 2000). The problem of memory effects distorting empirical findings is frequently mentioned in the literature (Alwin, 2007; Moser & Kalton, 1972; Saris, Revilla, Krosnick, & Shaeffer, 2010). Nevertheless, only a few studies empirically investigate memory effects and their consequences for surveys with repeated measurements. The scarcity of research is particularly severe regarding memory effects occurring within the same survey. Thus, we conducted an experiment in the probability-based German Internet Panel (GIP) to investigate respondents' recall ability within an online survey. Moreover, we implement memory interference tasks between the repeated measurements to test whether respondents' recall ability can be decreased.

This document is a preprint and thus it may differ from the final version: Revilla, M., & Höhne, J.K. (2020). Repeatedly measuring political interest: Can we reduce respondents' recall ability and memory effects in surveys using memory interference tasks? International Journal of Public Opinion Research. DOI: 10.1093/ijpor/edaa035.

Background

Respondents' Recall Ability Within the Same Survey

Van Meurs and Saris (1990), using data from a representative panel of the Dutch adult population (N = 1,537), asked respondents at the end of a self-administered survey whether they recall their answers to previous questions about political parties and public services. Then, they asked respondents to reproduce their answers. They found that, after about 9 minutes, 70% of the respondents stating that they recalled their previous answer and 36% of those stating that they did not recall it were able to reproduce the initial answer. "It follows that of the respondents who can reproduce the answers correctly after 9 minutes approximately (70%–36% or) 34% is probably due to memory effects while the others reproduce the result because of the stability of their opinion" (van Meurs & Saris, 1990, p. 141). They also found that 1) when the time interval between the two repeated measurements increased, the percentage of respondents who correctly reproduced their previous answer decreased, 2) asking questions about similar topics in the time interval between a question and its repetition decreased respondents' likelihood to correctly reproduce their answer, and 3) respondents giving extreme answers were more likely to correctly reproduce their answer.

Considering the methodological changes that took place since van Meurs and Saris (1990) collected these data, one might wonder whether their results are generalizable to recent surveys. The need for information recall in everyday life changed drastically (Revilla, Ochoa, & Loewe, 2017). Moreover, computer and internet literacy (Eshet-Alkalai & Chajut, 2009) and participation in (online) surveys have substantially increased (Couper, 2017).

Building on van Meurs and Saris's (1990) study, Schwarz, Revilla, and Weber (2020) conducted a lab experiment (N = 115; mainly university students). They asked respondents at the end of an online survey whether they can recall and reproduce their answer to a question on the difficulty/ease of dealing with important problems in life. Using the same estimation procedure as van Meurs and Saris (1990), they found that 17% of the respondents were consistent due to memory.

In addition, Rettig, Höhne, and Blom (2019), using data from the GIP (N = 2,119), also asked respondents whether they can recall and reproduce their answers to two questions about environmentally friendly products and energy saving. They found that 64% of those stating that they recalled their previous answer and 44% of those stating that they did not recall it reproduced the initial answer correctly, suggesting that 20% of the respondents reproduced their answer because of memory effects. The authors also found significant effects of extreme answers and question types on respondents' recall ability.

Memory Interference Tasks

There are at least two processes of forgetting: 1) memory fades over time and 2) memory gets disrupted by events occurring between the "learning" of information and the attempt to recall it (Baddeley, 2014). Adding a task with a similar content between the actual learning and the recall may decrease people's recall ability (McGeoch & McDonald, 1931; Crouse, 1971). This method is known as "retroactive interference of memory" (Henderson, 2005; Baddeley, 2014). Classical tasks used to interfere with memory during psychological experiments include asking

participants to memorize and recall certain contents between the initial learning and the recall request (Crouse, 1971; Baddeley, 2014).

In line with this research, Schwarz, Revilla, and Weber (2020) tested whether including a memory interference task (evaluating the grammatical correctness of sentences) within an online survey reduced the correct reproduction of the answers to the test question. In contrast to their expectation, the authors found no significant reduction of the correct reproduction when adding this interference task.

Research Questions and Contribution

This study addresses the following research questions:

RQ1. What is the proportion of respondents who: a) stated that they recall their previous answer (stated recall), b) were able to correctly reproduce their answer (correct reproduction), and c) how certain are respondents about their ability to recall their previous answer (recall certainty)?

RQ2. Does the implementation of a memory interference task between a question and its repetition reduce a) respondents' stated recall, b) correct reproduction, and c) recall certainty?

RQ3. What other aspects (e.g., extreme answers and time between repetitions) affect a) stated recall, b) correct reproduction, and c) recall certainty?

RQ4. What is the estimated proportion of correct reproduction due to memory?

RQ5. Do memory interference tasks decrease the estimated proportion of correct reproduction due to memory?

Building on the previous literature and particularly the study by Schwarz, Revilla, and Weber (2020), this study contributes to the limited body of research on recall ability and memory effects within the same survey: 1) we conducted a field experiment in a large-scale probability-based online panel. 2) We employed a widely used topic that is repeatedly asked in major social surveys (political interest) so that our results are of direct interest for researchers. In contrast to previous studies, we focused on a concept that is known to be stable over time. This allows us to collect information about the robustness of previous results under different conditions and to broaden the existing knowledge. 3) For similar reasons, we used a shorter (five-point) completely verbalized scale in the test question. For such a scale, a higher recall ability than in previous studies (with longer partially verbalized scales) can be expected. 4) Since the memory interference task employed by Schwarz, Revilla, and Weber (2020) did not reduce correct reproduction, we employed a different type of task, asking respondents to recall words/numbers. 5) We varied the content and position of the task within the survey.

Method

Evaluating Respondents' Recall Ability

The test question on political interest was taken from Beierlein, Kemper, Kovaleva, and Rammstedt (2012): "In general, how interested would you say you are in politics?". It was presented on a single screen with a vertical scale ("very interested," "fairly interested," "somewhat interested," "hardly interested," and "not at all interested"). This question (the fifth one in the survey) was followed by up to 47 questions on a variety of topics. Finally, respondents were asked three follow-up questions adapted from van Meurs and Saris (1990).

The first question asked whether respondents recall their answer to the test question (yes/no). The second one asked respondents to indicate (if they stated recall) or estimate (if they stated no recall) what their answer was. By comparing this answer to the initial answer to the political interest test question, we determined respondents' correct reproduction. The third question asked respondents how confident they were about their recall (0 "not at all certain" to 10 "absolutely certain"). English translations of these questions are available in Appendix A.

Experimental Design

Respondents were randomly assigned to one of the following conditions:

- *Control group*: no interference task, but two substitute questions.
- *Immediate words*: interference task right after the test question asking to memorize and recall five words.
- *Lagged words*: same as *Immediate words group*, but with five survey questions between the test question and the task.
- *Immediate numbers*: same as *Immediate words group* but with five numbers.
- *Lagged numbers*: same as *Lagged words group* but with five numbers.

In the memory interference tasks, respondents were initially asked to memorize the words/numbers. Then, they were asked to type in these words/numbers in a textbox. English translations of the tasks and substitute questions are available in Appendix B. An illustration of the experimental design is available in Appendix C.

Data

The data were collected in wave 42 (July 2019) of the GIP (see Blom et al., 2020). In total, 4,714 respondents started this wave and 33 broke-off before being asked any study-relevant questions, leaving 4,681 respondents for the analyses.[1]

Analyses

To answer *RQ1a* and *RQ1b*, we report the proportions of respondents stating that they recalled their previous answer and correctly reproducing their previous answer, computed over all respondents who answered to the test and follow-up questions. To answer *RQ1c*, we report the average recall certainty, for those answering the test and follow-up questions.

To answer *RQ2*, we test for significance of differences in stated recall, correct reproduction, and recall certainty between the *Control group* and each treatment group (Z-tests for proportions; T-tests for means). To control for other variables (explained below), we also ran logistic regressions with stated recall (1 = yes) and correct reproduction (1 = yes) as binary dependent variables. Additionally, we ran an OLS regression with the dependent variable recall certainty (0 "not at all certain" to 10 "absolutely certain"). To test whether the memory interference tasks affect the three dependent variables, we include as independent variables four

[1] The median age category of these respondents was 50 to 54 years and 48.1% of them were female. In terms of education, 17.5% graduated from a lower secondary school (low education level), 31.2% from an intermediate secondary school (middle education level), and 51.3% from a college preparatory secondary school or university (high education level). To evaluate the sample composition between the experimental groups, we conducted chi-square tests. The results showed no significant differences regarding age, gender, and education.

dummies based on the experimental groups (*Immediate words*, *Lagged words*, *Immediate numbers*, and *Lagged numbers*; 1 = assigned to this group), with the *Control group* serving as reference.

To answer *RQ3*, we added in all regressions three independent variables suggested by previous research:

1) Extreme answer (1 = yes): indicates whether respondents selected an endpoint category when answering the test question.

2) In-between time (in seconds): defined as the start time of the test question minus the end time of the first follow-up question. Following Rettig, Höhne, and Blom (2019), we consider in-between times that were higher than 3 hours as outliers (indicating multitasking and/or interruptions). However, whereas Rettig, Höhne, and Blom (2019) dropped these outliers, we replaced them with the value of 3 hours to keep the sample size as large as possible. In total, 1.1% of the respondents had in-between times higher than 3 hours.[2]

3) Response time (in seconds): defined as the time spent on the survey-page displaying the test question. To deal with outliers, we replaced all response times higher than 60 seconds with the value of 60 seconds. In total, 0.5% of the respondents had longer response times.[3]

Following Rettig, Höhne, and Blom (2019), we also control for the device used to complete the survey using a dummy smartphone (1 = yes) and several respondent characteristics: age (14 ascending categories), female (1 = yes), and education level using two dummies indicating low (1 = yes) and high (1 = yes) education (middle being the reference). However, in contrast to Rettig, Höhne, and Blom (2019), we do not control for panel experience in the regressions since they were no newly recruited panelists in wave 42. In the regression of correct reproduction, we controlled for stated recall and in the regression of recall certainty, we controlled for stated recall and correct reproduction.

To answer *RQ4*, we use the estimation procedure proposed by van Meurs and Saris (1990). First, we calculate the proportion of respondents who stated that they recall their previous answer and reproduced it correctly. Then, we calculate the proportion of respondents who stated that they do not recall their previous answer but still reproduced it correctly. Finally, we calculate the difference between these two proportions.

To answer *RQ5*, we test whether the proportions of respondents who correctly reproduced their previous answer due to memory is significantly different in the *Control group* and each treatment group (Z-tests).

[2] When we drop outliers instead of replacing them, the effect of Lagged numbers on stated recall and the effect of in-between time on recall certainty lose statistical significance ($p = 0.06$ and $p = 0.10$, respectively). All the other results do not change (see the outcomes of the regressions in Table 2).

[3] We include a variable about the time spent on the test question because if respondents take more time to think about their answer to the test question, their recall ability might improve. This variable aims to capture the attention paid by respondents to the test question and the effort respondents invest in the selection of an answer. Thus, response time differs from in-between time. Therefore, we apply a stricter outlier definition for response time than for in-between time. Results do not change when we use 120 seconds as the outlier threshold instead of 60 seconds.

Results

Stated Recall, Correct Reproduction, and Recall Certainty

Levels of stated recall, correct reproduction, and recall certainty

First, Table 1 displays stated recall, correct reproduction, and recall certainty, for all respondents and for each experimental group.

Table 1. Stated recall, correct reproduction, and recall certainty (overall and by experimental group)

	Overall	Control group	Immediate words	Lagged words	Immediate numbers	Lagged numbers
Stated recall (% yes)	90.0	88.8	91.3	89.4	89.3	91.5
Correct reproduction (% same answer)	87.5	85.4	89.3*	86.1	88.1	88.8*
Recall certainty (mean)	8.3	8.2	8.4*	8.3	8.4	8.3

Note. *p < .05. The proportions (%) and means are computed for those who provided an answer to the corresponding questions. Overall, we have 4,627 respondents for the stated recall question, 4,623 for the correct reproduction question, and 4,620 for the recall certainty question. We test for significance of differences between *Control group* and each treatment group (i.e., *Immediate words*, *Lagged words*, *Immediate numbers*, and *Lagged numbers*).

Regarding *RQ1*, the proportions of stated recall (90.0%), correct reproduction (87.5%), and recall certainty (8.3%) are high and also higher than those reported in previous studies (Rettig, Höhne, & Blom, 2019; Schwarz, Revilla, & Weber, 2020; van Meurs & Saris, 1990).

Regarding *RQ2*, the results reveal that the memory interference tasks did not reduce stated recall, irrespective of the task and its position in the survey. For correct reproduction, significant differences were found between the *Control group* and the *Immediate words group* as well as between the *Control group* and the *Lagged numbers group*. However, in all cases, correct reproduction is higher for the groups that received a memory interference task than for the *Control group*. For recall certainty, only one significant difference between the *Control group* and the *Immediate words group* is observed: recall certainty is higher for respondents receiving the memory interference task. Overall, the tasks did not decrease respondents' recall ability.

However, the memory interference tasks can affect respondents' recall ability only if respondents engage in the task (i.e., memorizing and recalling the words/numbers appropriately). Since the survey was self-administered, we could not control respondents' engagement with the task. However, we could evaluate respondents' task performance by counting the number of correct words/numbers entered in the textbox. Most respondents' task performance was good (see Appendix D) suggesting that poor task performance is not the reason for the absence of decrease in recall ability on the test question.

Explaining stated recall, correct reproduction, and recall certainty

To control for other variables and to investigate whether the variables suggested by previous research affect respondents' recall ability, we conducted a series of regressions. Table 2

presents odds ratio for the logistic regressions and coefficients for the OLS regression. The explained variances are low; especially in the logistic regressions.

Table 2. Logistic regressions of stated recall and correct reproduction and OLS regression of recall certainty (standard errors in parentheses)

	Stated recall Odds ratio (SE)	Correct reproduction Odds ratio (SE)	Recall certainty Coefficient (SE)
Immediate words	1.34 (0.21)	1.38 (0.20)*	0.11 (0.08)
Lagged words	1.09 (0.17)	1.06 (0.14)	-0.01 (0.08)
Immediate numbers	1.07 (0.16)	1.26 (0.17)	0.11 (0.08)
Lagged numbers	1.38 (0.22)*	1.33 (0.19)*	-0.05 (0.08)
Extreme answer	3.79 (1.02)***	0.91 (0.13)	0.86 (0.09)***
In-between time	0.99 (0.00)***	0.99 (0.00)	-0.00 (0.00)**
Response time	1.00 (0.01)	0.98 (0.01)*	0.00 (0.00)
Smartphone	0.91 (0.11)	0.85 (0.09)	-0.05 (0.07)
Age	1.01 (0.00)**	1.00 (0.00)	-0.01 (0.00)***
Female	0.84 (0.08)	1.11 (0.10)	-0.18 (0.05)**
Low education	0.76 (0.11)*	0.85 (0.11)	-0.22 (0.08)**
High education	1.12 (0.13)	1.33 (0.14)**	0.21 (0.06)**
Stated recall	NA	1.71 (0.22)***	1.66 (0.09)***
Correct reproduction	NA	NA	1.02 (0.08)***
Constant	5.20 (1.28)***	3.40 (0.84)***	6.41 (0.17)***
Pseudo-R^2/ adjusted R^2	0.0362	0.0154	0.1472
Observations	4,618	4,614	4,611

Note. *p < .05, **p < .01, ***p < .001. Pseudo-R^2 for logistic and adjusted R^2 for OLS regressions. NA stands for Not Applicable.

Regarding *RQ2*, the results of the regressions are in line with the conclusions drawn when comparing the proportions of stated recall, correct reproduction, and recall certainty between the *Control group* and treatment groups. There are only three significant effects of the memory interference tasks: *Immediate words* on correct reproduction and *Lagged numbers* on stated recall and correct reproduction.

Regarding *RQ3*, we found significant effects of extreme answer and in-between time on stated recall and recall certainty, but not on correct reproduction. However, the sizes of the odds ratio (stated recall) and coefficients (recall certainty) of in-between time are negligibly small. The same applies to response time, where we also observe a negligibly small significant effect on correct reproduction.

Memory Effects

Table 3 displays the proportions of respondents who correctly reproduced their answer for those who answered “yes” to the “stated recall” question (first row) and those who answered “no”

(second row). The estimated memory effects (difference between the first and second rows) are shown in the last row.

Table 3. Correct reproduction of respondents stating recall and respondents stating no recall and estimated memory effects (overall and by experimental group)

	Overall	Control group	Immediate words	Lagged words	Immediate numbers	Lagged numbers
Correct reproduction when stated recall = yes (%)	88.3	86.1	89.7*	87.0	89.3*	89.1
Correct reproduction when stated recall = no (%)	80.9	79.8	85.2	78.6	77.8	84.8
Estimated memory effects (%)	7.4	6.3	4.5	8.4	11.5**	4.3

Note. *p<.05, **p<.01. The proportions (%) are computed only for those who provided an answer to the corresponding question. Overall, we have 4,162 respondents stating recall and 461 stating no recall. To test for significant differences in the estimated memory effects, we used the n of the group stating recall (yes).

The estimated memory effects are lower than those reported in previous studies (Rettig, Höhne, & Blom, 2019; Schwarz, Revilla, & Weber, 2020, van Meurs & Saris, 1990). Furthermore, they do not differ significantly across the *Control group* and the treatment groups with one exception: the *Immediate numbers group* shows a significantly higher memory effect than the *Control group*. Overall, the memory interference tasks used did not reduce the correct reproduction due to memory.

Discussion and Conclusion

The aim of this study was to investigate respondents recall ability and the size of memory effects when questions are repeated within the same online survey. For this purpose, we conducted an experiment in the GIP using a test question on political interest. Furthermore, we implemented different memory interference tasks to investigate whether respondents' recall ability and memory effects decrease.

Regarding *RQ1*, we found that stated recall, correct reproduction, and recall certainty were higher than in previous studies. A potential explanation for the higher proportion of correct reproduction in our study compared to the study by Schwarz, Revilla, and Weber (2020) could be that the latter did not contain any extreme answers, while in our study extreme answers occurred for 11.3% of the respondents. However, Rettig, Höhne, and Blom (2019) had an even higher proportion of extreme answers (16.6%) but a lower correct reproduction. Thus, another potential reason is that our test question used a five-point, completely verbalized scale, whereas the three other studies used test questions with ten- or eleven-point, end-verbalized scales.

Having less response categories may facilitate recalling the previous answer. It also increases the likelihood of correct reproduction that can be attributed to chance. Finally, another possible reason is that our test question measured a more stable concept.

Regarding *RQ2*, we did not find any evidence that the implementation of memory interference tasks between the repeated measurements reduces respondents' stated recall, correct reproduction, or recall certainty. In some cases, the memory interference tasks even increased recall ability. This corroborates the findings reported by Schwarz, Revilla, and Weber (2020), even though they used a different type of memory interference task. However, more refined research is needed on the use of memory interference tasks in reducing respondents' recall ability in surveys, as these findings contradict those reported in cognitive psychological research on the processes of forgetting (Baddeley, 2014).

Moving to *RQ3*, we found significant effects of extreme answer, in-between time, and response time. However, these effects do not exhibit a systematic pattern across stated recall, correct reproduction, and recall certainty. In addition, the effects of the two time-based variables are negligibly small. Mixed support was also found for these variables in previous research, suggesting that there is a need for further investigation of the link between these variables and recall ability.

Concerning *RQ4*, the estimated proportion of correct reproduction due to memory is 7.4%. This is substantially smaller than the values reported in previous research: 17% in Schwarz, Revilla, and Weber (2020), 20% in Rettig, Höhne, and Blom (2019), and 34% in van Meurs and Saris (1990). Possible explanations for the lower memory effects in our study are the stable concept used in the test question and the high level of correct reproduction among respondents stating no recall, which might be related to the shorter and completely verbalized scale.

Finally, the results for *RQ5* revealed that the memory interference tasks did not decrease the estimated proportion of correct reproduction due to memory. The only significant difference indicates an increase in memory effects. This might be linked to the low proportion of correct reproduction due to memory.

This study has some limitations. First, we employed one single test question and did not control for the question type as Rettig, Höhne, and Blom (2019) did. Second, we could not control for multitasking and/or distractions during survey completion, so the in-between time and response time are imprecise measures of the concepts of interest (see Höhne & Schlosser, 2018). Although a lab setting, such as the one used by Schwarz, Revilla, and Weber (2020), is associated with several methodological disadvantages, it has the advantage of allowing for the proper control for multitasking and/or distractions and for monitoring respondents' engagement with the memory interference task. Third, our experiment was conducted in the same online panel (GIP) as the one used by Rettig, Höhne, and Blom (2019). Thus, there is a chance that respondents were aware that a repetition of the questions can take place, which, in turn, may have affected their response behavior. However, since the two studies were separated by seven months (three survey waves), we do not expect this effect to be strong. Fourth, in contrast to van Meurs and Saris (1990), we neither varied the length of the survey nor the question topics between the test question and its repetition. Fifth, the estimation procedure proposed by van

Meurs and Saris (1990) only allows the determination of memory effects on an aggregated (group) level. Thus, no explanations on an individual (respondent) level are possible.

Despite its limitations, this study significantly contributes to the scarce literature on respondents' recall ability and memory effects within the same survey: 1) it provides new evidence for a widely used concept in social science research and adjacent research fields (political interest), 2) it provides evidence for a prevalently used scale (five-point, completely verbalized), and 3) it tests different interference tasks (memorizing the words/numbers). Since memory effects are lower than in previous studies, this suggests that the concept of interest and/or the scale characteristics can affect the size of memory effects. Thus, we recommend that researchers keep this in mind when deciding on the scale to be used. Furthermore, this study shows that even a more stable concept is not free of memory effects and that respondents' recall ability and memory effects cannot be easily reduced by memory interference tasks. Thus, researchers should account for the occurrence of memory effects in future studies even when interested in more stable concepts (e.g. when analyzing a multitrait-multimethod experiment, they could introduce time factors in their analyses in addition to method factors). Otherwise, memory effects may bias results that are crucial for the scholarly understanding of political behavior. There is a high need for future research exploring how to best implement memory interference tasks for reducing memory effects. It is crucial to properly understand the mechanisms behind our findings that are not in line with cognitive psychological research on the processes of forgetting. Finally, research exploring new alternatives to reduce memory effects in surveys with repeated measurements is highly needed.

References

Alwin, D. F. (2007). *Margins of Error: A Study of Reliability in Survey Measurement*. Hoboken, NJ: Wiley & Sons.

Baddeley, A. (2014). *Essentials of Human Memory (Classic Edition)*. New York, NY: Psychology Press.

Beierlein, C., Kemper, C. J., Kovaleva, A., & Rammstedt, B. (2012). PEKS – Political Efficacy Kurzskala. In Leibniz-Zentrum für Psychologische Information und Dokumentation (Ed.), Elektronisches Testarchiv (PSYNDEX Tests-Nr. 9006492). Trier: ZPID. https://doi.org/10.23668/psycharchives.419

Blom, A. G., Fikel, M., Friedel, S., Höhne, J. K., Krieger, U., Rettig, T., Wenz, A., & SFB 884 "Political Economy of Reforms" (2020). *German Internet Panel, Wave 42 (July 2019)*. GESIS Data Archive, Cologne (Germany). ZA7591 Data file Version 1.0.0, https://doi.org/10.4232/1.13465

Couper, M. P. (2017). New developments in survey data collection. *Annual Review of Sociology, 43*, 121–145. https://doi.org/10.1146/annurev-soc-060116-053613

Crouse, J. H. (1971). Retroactive interference in reading prose materials. *Journal of Educational Psychology, 62*, 39-44. https://doi.org/10.1037/h0030652

Eshet-Alkalai, Y., & Chajut, E. (2009). Changes over time in digital literacy. *CyberPsychology & Behavior, 12*, 713-715. http://doi.org/10.1089/cpb.2008.0264

Henderson, J. (2005). *Memory and Forgetting*. London: Routledge.

Höhne, J. K., & Schlosser, S. (2018). Investigating the adequacy of response time outlier definitions in computer-based web surveys using paradata SurveyFocus. *Social Science Computer Review, 36,* 369–378. https://doi.org/10.1177/0894439317710450

McGeoch, J. A., & McDonald, W. T. (1931). Meaningful relation and retroactive inhibition. *The American Journal of Psychology*, *43*, 579–588. DOI:10.2307/1415159.

Moser, C. A., & Kalton, G. (1972). *Survey Methods in Social Investigation*. New York, NY: Basic Books.

Rettig, T., Höhne, J. K., & Blom, A. G. (2019). Recalling survey answers: A comparison across question types and different levels of online panel experience. Under review.

Revilla, M., Ochoa, C, & Loewe, G. (2017). Using passive data from a meter to complement survey data in order to study online behavior. *Social Science Computer Review*, *35*, 521–536. https://doi.org/10.1177/0894439316638457

Saris, W. E., Revilla, M., Krosnick, J. A., & Shaeffer, E. M. (2010). Comparing questions with agree/disagree response options to questions with construct-specific response options. *Survey Research Methods, 4*, 61–79. DOI: 10.18148/srm/2010.v4i1.2682

Schwarz, H., Revilla, M., & Weber, W. (2020). Memory effects in repeated survey questions – Reviving the empirical investigation of the independent measurements assumption. *Survey Research Method*, *14*(3), 325-344 https://doi.org/10.18148/srm/2020.v14i3.7579

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.

van Meurs, A., & Saris, W. E. (1990). Memory effects in MTMM studies. In W. E. Saris & A. van Meurs (Eds.), *Evaluation of Measurement Instruments by Meta-Analysis of Multitrait Multimethod Studies*, (pp. 134–146). Amsterdam, NL: North Holland.