

## Assessing time-based prospective memory online: A comparison study between laboratory-based and web-based testing

Authors	Laera,G.; Hering,A.; Kliegel,M.
Published in	Quarterly Journal of Experimental Psychology
DOI	<a href="https://doi.org/10.1177/17470218231220578">10.1177/17470218231220578</a>
Publication Date	2024
Document Version	publishersversion
Link	<a href="https://research.tilburguniversity.edu/en/publications/6be5377a-d3e7-44a8-92cc-f3a43f614d7b">https://research.tilburguniversity.edu/en/publications/6be5377a-d3e7-44a8-92cc-f3a43f614d7b</a>
Citation	Laera, G, Hering, A & Kliegel, M 2024, 'Assessing time-based prospective memory online : A comparison study between laboratory-based and web-based testing', Quarterly Journal of Experimental Psychology, vol. 77, no. 11, pp. 2214-2227. <a href="https://doi.org/10.1177/17470218231220578">https://doi.org/10.1177/17470218231220578</a>
Download Date	2026-06-13 06:40:49
Rights	<p>General rights</p> <p>Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.</p> <ul style="list-style-type: none"> <li>- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.</li> <li>- You may not further distribute the material or use it for any profit-making activity or commercial gain</li> <li>- You may freely distribute the URL identifying the publication in the public portal"</li> </ul> <p>Take down policy</p> <p>If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.</p>

# Assessing time-based prospective memory online: A comparison study between laboratory-based and web-based testing



Gianvito Laera<sup>1,2,3</sup> , Alexandra Hering<sup>1,2,4</sup>  
and Matthias Kliegel<sup>1,2,3</sup>

Quarterly Journal of Experimental Psychology  
1–14

© Experimental Psychology Society 2024



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/17470218231220578

[qjep.sagepub.com](https://qjep.sagepub.com)



## Abstract

Prospective memory (PM, i.e., the ability to remember and perform future intentions) is assessed mainly within laboratory settings; however, in the last two decades, several studies have started testing PM online. Most part of those studies focused on event-based PM (EBPM), and only a few assessed time-based PM (TBPM), possibly because time keeping is difficult to control or standardise without experimental control. Thus, it is still unclear whether time monitoring patterns in online studies replicate typical patterns obtained in laboratory tasks. In this study, we therefore aimed to investigate whether the behavioural outcome measures obtained from the traditional TBPM paradigm in the laboratory—accuracy and time monitoring—are comparable with an online version in a sample of 101 younger adults. Results showed no significant difference in TBPM performance in the laboratory versus online setting, as well as no difference in time monitoring. However, we found that participants were somewhat faster and more accurate at the ongoing task during the laboratory assessment, but those differences were not related to holding an intention in mind. The findings suggest that, although participants seemed generally more distracted when tested remotely, online assessment yielded similar results in key temporal characteristics and behavioural performance as for the laboratory assessment. The results are discussed in terms of possible conceptual and methodological implications for online testing.

## Keywords

Delayed intentions; time-based; online assessment; monitoring

Received: 27 February 2023; revised: 24 November 2023; accepted: 27 November 2023

Prospective memory (PM) is the ability to carry out delayed intentions while performing a background activity, referred to as the ongoing task (OT). There are two types of PM tasks: time-based PM (TBPM) involves performing actions at specific future times (e.g., taking medication every 2 hr), while event-based PM (EBPM) relates to executing intentions when specific events occur (e.g., taking medication during a meal) (Einstein & McDaniel, 1990). PM is crucial in daily activities like medication management, financial tasks, and meal preparation (Haas et al., 2020; Hering et al., 2018; Laera et al., 2023; Woods et al., 2015). Thus, PM is an important psychological construct that is studied in different fields of psychology, such as clinical, lifespan, and applied psychology, as well as neuroscience and neuropsychology (Boag et al., 2019; Burgess et al., 2003; Einstein

et al., 1995; Kvavilashvili et al., 2009; Loft et al., 2021; Okuda et al., 2007; Suchy et al., 2020).

<sup>1</sup>Cognitive Aging Lab (CAL), Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland

<sup>2</sup>Centre for the Interdisciplinary Study of Gerontology and Vulnerability, University of Geneva, Geneva, Switzerland

<sup>3</sup>LIVES—Overcoming Vulnerability: Life Course Perspectives, Swiss National Centre of Competence in Research, Lausanne, Switzerland

<sup>4</sup>Department of Developmental Psychology, Tilburg School for Social and Behavioral Sciences, Tilburg University, Tilburg, The Netherlands

### Corresponding author:

Gianvito Laera, Cognitive Aging Lab (CAL), Faculty of Psychology and Educational Sciences, University of Geneva, 28 Boulevard du Pont d'Arve, 1205 Geneva, Switzerland.

Email: [Gianvito.Laera@unige.ch](mailto:Gianvito.Laera@unige.ch)

PM has been assessed mainly within laboratory settings (Einstein et al., 1995; Einstein & McDaniel, 1990; Mattli et al., 2014; Park et al., 1997), although alternative approaches, such as memory diaries and naturalistic settings, have been explored (Crovitz & Daniel, 1984; Haas et al., 2020; Kvavilashvili & Fisher, 2007; Raskin et al., 2018). In the TBPM task, participants must remember to perform actions, with key measures including TBPM accuracy (proportion of correct responses), time monitoring (frequency of clock checks), and OT performance accuracy and reaction times for correct trials (Huang et al., 2014; Labelle et al., 2009). Time monitoring has been identified as a crucial part for TBPM performance, as it is a marker of attentional resources allocated towards the PM task (Labelle et al., 2009; Varley et al., 2021). Research shows that individuals who perform on time tend to check the clock few times at the task's outset and strategically increase clock checks as the PM target time nears, forming a "J-shaped" curve (Joly-Burra et al., 2022; Labelle et al., 2009; Mäntylä et al., 2006; Mioni et al., 2017, 2020; Vanneste et al., 2016); indeed, many studies showed that strategic time monitoring strongly relates to PM accuracy and reflects strategic behaviour (Ceci & Bronfenbrenner, 1985; Harris & Wilkins, 1982; Mäntylä et al., 2006; Mioni et al., 2012, 2020; Mioni & Stablum, 2014; Vanneste et al., 2016). Furthermore, studies suggest a positive link between time estimation skills and strategic time monitoring, implying the ability to strategically use internal time representations (Labelle et al., 2009; Mioni & Stablum, 2014; Vanneste et al., 2016). Traditionally, TBPM has been mainly studied in controlled laboratory settings, as this provides the highest degree of control, especially with respect to how and when participants may monitor the time. However, due to COVID-19 limitations and accessibility issues, remote online testing has gained popularity. Laboratory testing involves logistics and may exclude certain participants who are not willing or able to travel (Backx et al., 2020; Germine et al., 2012). Therefore, alternatives like remote testing have become valuable for cognitive assessment (Backx et al., 2020; Feenstra et al., 2017; Finley & Penningroth, 2015; Germine et al., 2012; Logie & Maylor, 2009).

Internet-based methods have become essential for experimental psychologists to assess cognitive functions remotely in participants' homes (Backx et al., 2020; Feenstra et al., 2017; Finley & Penningroth, 2015). However, differences between online and lab settings must be considered (Bauer et al., 2012; Hoskins et al., 2010; Parsons et al., 2018; Skitka & Sargis, 2006). Social contact created by the presence of the experimenter(s) in the laboratory may not only affect the cognitive performance but can also be an additional explanatory source for the participants regarding tasks' instructions (Hoskins et al., 2010). Another difference is the testing environment, which can be fully controlled in the laboratory, but it is uncontrolled elsewhere (Bauer et al., 2012; Skitka &

Sargis, 2006). In addition, variations in hardware, software, internet speed, and connectivity can affect results, especially reaction times (Parsons et al., 2018). Despite these disparities, online assessments reliably test cognitive processes such as executive function, memory, psycholinguistics, and attention (Backx et al., 2020; Feenstra et al., 2017; Germine et al., 2012). However, researchers must be cautious in online sample selection and control for demographic factors such as age, gender, and education (Uittenhove et al., 2023). Notably, remote testing of TBPM presents unique challenges due to environmental cues that influence performance and are hard to control online. For example, in the laboratory, it is recommended to control for any temporal cue that could affect time monitoring, such as removing any clock from the testing room or closing windows to remove any temporal influence provided by the day–night cycle (Barner et al., 2019; Esposito et al., 2015; Rothen & Meier, 2017). Given that such level of control is not achievable in online settings, it is possible that participants use other temporal cues, rather than the clock provided by the task's procedure, to estimate the occurrence of the PM target time.

Only a few studies have explored online testing for PM, particularly EBPM (Gilbert, 2015b; Gladwin et al., 2020; Horn & Freund, 2021; Scarampi & Gilbert, 2020). However, the majority of these studies did not include any laboratory measures of PM that would allow comparisons with online assessment. One study found comparable data between online and lab settings, indicating that the online setting did not show data loss (Finley & Penningroth, 2015). In the context of TBPM, two studies used online testing (Gilbert, 2015a; Zuber et al., 2022). In the study by Gilbert (2015a), participants were asked to perform a TBPM task (i.e., press a button every 30 s) while engaged in the lexical decision task; however, the authors did not include an assessment in the laboratory. Another study compared TBPM in both settings, showing similar accuracy and time monitoring but lower OT accuracy (Zuber et al., 2022). The authors argued that participants could have been less attentive to the OT when tested online; no measure was provided either on OT reaction times or on PM cost. Nonetheless, although the authors used a traditional TBPM task in the laboratory, they did not examine this paradigm online; instead, they assessed—both online and in the laboratory—a TBPM task embodied into a serious game that had higher cognitive demands and allocation of attentional resources compared with the traditional version of the TBPM paradigm (Zuber & Kliegel, 2020). Consequently, it is still unknown whether traditional TBPM task can be used in an online assessment to yield reliable and valid results that resemble the empirical evidence obtained in the experimental laboratory setting. It is also unknown if online distractibility in OT is due to the setting or maintaining an intention. To investigate this, examining PM costs separately from OT performance is crucial. PM costs represent interference in OT

performance when attention is divided between PM and OT tasks (Anderson et al., 2019; Peper & Ball, 2022; Smith, 2003). No online PM study has yet explored changes in PM costs during the assessment. This could help determine if increased online distractibility is due to intention maintenance or contextual factors.

In this study, we therefore aimed to investigate whether the outcome measures obtained from the traditional TBPM paradigm in the laboratory and in the online assessment are comparable in a sample of younger adults. Participants completed a 2-min TBPM task while concurrently doing a lexical decision task; they can check the time by pressing the spacebar (see Conte & McBride, 2018; Del Missier et al., 2021; Park et al., 1997). One group performed the task in the laboratory, while one group performed the same task in a fully self-administered online assessment remotely from the participants' home. Based on the previous evidence (Zuber et al., 2022), we predicted better overall OT accuracy in the laboratory, as distractions are typically lower; however, we expected no differences in time monitoring and TBPM accuracy. Moreover, we predicted a difference in PM costs across assessments, with higher costs in the laboratory, because environmental distractions should be better controlled in the laboratory setting, and attentional resources should be selectively devoted to the maintenance of the intention in mind and to the execution of the PM task. Diversely, participants could be generically more distracted during the OT execution when assessed online, but such an increase in distractibility should not be related to holding an intention in mind; therefore, PM costs should be lower online compared with the laboratory assessment. All the analyses were controlled for the effect of age, gender, and education, as such variables can cause differences across assessments (Uittenhove et al., 2023). This article has been pre-registered in the Open Science Framework repository (<https://doi.org/10.17605/OSF.IO/9V5FX>).

## Method

### Power analyses

This study was powered to detect small-to-moderate differences in behavioural performances between assessments (laboratory versus online). The power to detect differences between assessments was examined using the software G\*power 3 (Faul et al., 2007). We calculated the Cohen's  $d$  effect sizes (Cohen, 1969, pp. 278–280; Lakens, 2013) for PM and OT performance, as well as for time monitoring, from the only previous study that investigated the difference between laboratory and online assessment in TBPM (Zuber et al., 2022), which showed that PM accuracy and time monitoring were similar in both laboratory and online settings, whereas OT performance was moderately but significantly lower in the online assessment (Cohen's  $d$ : 0.20). Thus, to replicate the finding in the

literature, in the power analysis we used an effect size of 0.20; the power analysis indicated that detecting an effect size of 0.20, at 80% power (2-tailed  $\alpha$  at 0.05), would require a sample of 52 participants in an independent sample test with normal distribution. To increase the statistical power, as suggested by previous online studies (Finley & Penningroth, 2015; Logie & Maylor, 2009), we chose to have an initial sample whose size ( $N=134$ ) was more than double the ideal size provided by the power analysis ( $N=52$ ). The power analysis indicated that detecting an effect size of 0.20, using an independent sample with the normal distribution of 134 individuals, resulted in a statistical power of 96% (2-tailed  $\alpha$  at 0.05).

### Participants

One hundred thirty-four participants participated in the study (age range: 18–35 years;  $M_{\text{age}}=25$ ;  $SD_{\text{age}}=4.02$ ; 80 females); 67 participants took part in the laboratory assessment, while 67 participants took part in the online assessment. Participants who did the laboratory assessment were recruited using flyers, whereas participants who did the online assessment were recruited using Prolific ([www.prolific.co](http://www.prolific.co)), an online platform where participants receive payment for the completion of web-based experiments. Thirty-three participants (24.6% of the sample size) reported having a history of neurological or major psychiatric disease within the last 5 weeks (e.g., epilepsy, depression, and anxiety) or taking psychotropic drugs or others affecting the central nervous system. These participants were excluded.

The eligible sample comprised 101 individuals (age range: 18–35 years;  $M_{\text{age}}=25$ ;  $SD_{\text{age}}=4.08$ ; 64 females;  $M_{\text{education}}=17$  years;  $SD_{\text{education}}=4.05$ ); 52 performed the laboratory assessment, while 49 performed the online assessment. In the eligible sample, 6 participants (5.9% of the final sample size) reported being left-handed, whereas 95 participants (90.1%) reported being right-handed; 4 participants (4.0%) reported being ambidextrous. All participants gave their informed consent before participating in the study, which was conducted in accordance with the Declaration of Helsinki, and the protocol had been approved by the ethics commission of the Faculty of Psychology and Educational Sciences of the University of Geneva (PSE.20201102.04). Remuneration for the participation in both laboratory and online assessment was set at 10 CHF per hour, and it was delivered according to the actual duration taken by each participant to finish the experiment; as the experimental procedure took on average 25 min (minimum = 12 min, maximum = 77 min), participants were paid on average 4.16 CHF.

### Tasks and questionnaires

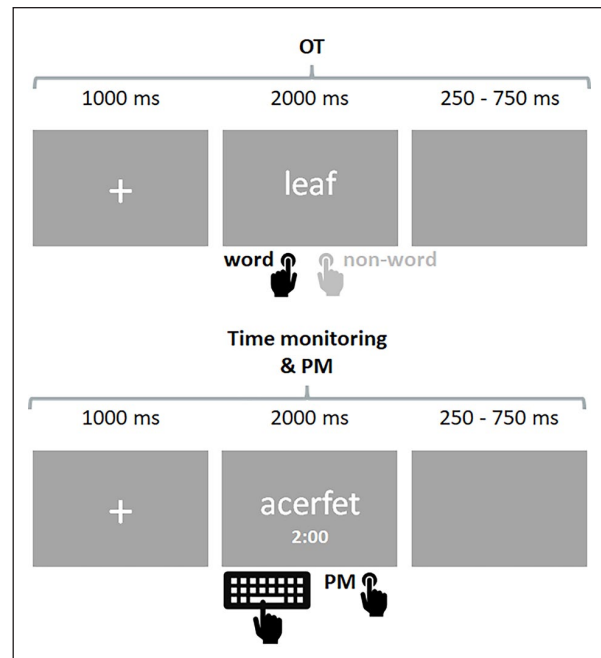
All the tasks have been programmed using Psychopy version 2021.2.3 (Peirce et al., 2019) and hosted by Pavlovla

(<https://pavlovia.org/>; Bridges et al., 2020); all the procedure was administered in French. The stimuli and the code of the experimental procedure are available in the Open Science Framework (<https://doi.org/10.17605/OSF.IO/9VVSFX1>). The English demo version of the experiment is available at the following link: <https://pavlovia.org/run/Laera/time-based-prospective-memory-demo>. An illustration of the OT and the TBPM task is represented in Figure 1.

**Ongoing task.** Participants performed a lexical decision task as OT (Meyer & Schvaneveldt, 1971), and they were required to indicate if a string of letters presented on the screen formed a word or not. Each OT trial started with a fixation cross (1,000 ms) followed by the stimulus (2,000 ms) and a subsequent blank period screen that lasted randomly between 250 and 750 ms. The random blank period avoided any temporal regularity related to the OT trials, which has been demonstrated to potentially work as a temporal cue supporting time monitoring (Guo & Huang, 2019; Heathcote et al., 2015).

Participants performed two blocks: one in which they carried out the lexical decision task alone, that is without any PM task on top of it (i.e., OT baseline), followed by an identical block in which they carried out the same OT, but this time they were asked to perform the TBPM task while doing the OT (i.e., OT with TBPM task). To keep participants engaged with the tasks, we included an additional check during the task: if participants did not respond to more than three OT trials in a row, the OT stopped, showing the following message: “It looks like you have stopped to give answers to the requested task. Please resume the task by pressing the ‘p’ key on your keyboard. Thanks for your collaboration.”; once the participants pressed “p” on their keyboard, the OT continued. If participants pressed the “p” key more than three times during the tasks, s/he was subsequently excluded from the analysis. Only two participants reported having pressed the “p” key during the TBPM task (no one pressed the key during the OT baseline), indicating that, overall, all participants kept engaged with the tasks.

Within both the OT baseline and the OT with TBPM, all stimuli (words and non-words) had between five and eight letters; we selected 278 stimuli (139 words and 139 non-words) based on their highest frequency scores following the rules of Ferrand for French words (Ferrand et al., 2010). We chose these stimuli to reduce the cognitive load related to the OT, to ensure that the effect of the assessment (laboratory and online) was free from confounding effects related to the complexity of the OT. All stimuli (words and non-words) were randomly delivered between the OT baseline and the OT with TBPM, as well as among the practice blocks before each task; overall, the OT comprised 32 stimuli, lasting maximum 2 min. The TBPM task administered online comprised between 136 and 158 OT stimuli (the variable range of OT stimuli was due to the



**Figure 1.** Ongoing task (OT) trial and time-based prospective memory TBPM task.

Graphical representation of the figure of the OT (above) and of the PM task (below). The OT was a lexical decision task (participants pressed “1” for words and “2” for non-words). During the PM task, participants were asked to press “ENTER” each 2 min, while having the possibility to freely check the clock by pressing “SPACEBAR.”

random inter-trial-interval), lasting maximum 8 min and 30 s; the TBPM task administered in the laboratory comprised between 168 and 194 OT stimuli, lasting maximum 10 min and 30 s; 52 stimuli were used for the practice blocks in both assessments.

**TBPM task.** In both the laboratory and the online assessment, the TBPM task was to remember to press the ENTER key on the keyboard every 2 min; moreover, participants were always free to check the clock as often as they wanted by pressing the SPACEBAR; if they did so, a digital clock showing the current time (format: “mm: ss”) appeared on the screen for 3 s. In total, five PM responses were collected for the laboratory assessment, whereas four responses were collected for the online assessment. We reduced the number of PM responses to 4 for the online assessment to limit the overall duration of the procedure because relatively long procedures (i.e., closer to 30 min) can increase the rate of experiments withdrawn by the participants online (Finley & Penningroth, 2015; Logie & Maylor, 2009).

#### Questionnaires

**Socio-demographic questionnaire.** Before the beginning of the OT baseline block, participants were required to answer a socio-demographic questionnaire, asking them

information about age, gender, education (mandatory and university), French language fluency, and three questions related to the exclusion criteria (i.e., 3 questions concerning the presence—within the last 5 weeks—of neurological, major psychiatric disease, and psychotropic drugs' intake, respectively).

**Follow-up questionnaire.** After the completion of the TBPM task, participants were asked to answer a brief follow-up questionnaire. For both laboratory and online assessments, there was a question related to the use of strategy during the TBPM task, and the participants were asked to give binary responses (“yes” or “no”); in the case of a “yes,” the participants were asked to provide a short statement of clarification. Specifically, the question was: “If you think about the previous task, where we asked you to press the ENTER key every 2 minutes having the possibility to check the clock, have you used a strategy to control the passage of time during this task? What strategy did you use?”

Only for participants participating in the online testing, we added more questions identical to the ones comprised within the socio-demographic questionnaire administered earlier (i.e., age, gender, education, mother-tongue language, and exclusion criteria);<sup>2</sup> we chose to repeat such questions to control for discrepancies between the two questionnaires, to increase the chance that participants reported the right information about socio-demographic background and exclusion criteria (Finley & Penningroth, 2015). We chose to exclude the participants who reported more than three discrepant answers; two participants (1.5% of the sample size) reported four answers that differed between the socio-demographic and the follow-up questionnaire; however, those participants were already excluded because they also reported to have a history of neurological or major psychiatric disease, or to take psychotropic drugs within the last 5 weeks.

## Procedure

Both the laboratory and the online assessment followed the same procedure. Prior to participation, all relevant information concerning the experimental procedure and data access were provided in written form and participants provided informed consent to participate in the study. Next, participants filled the socio-demographic questionnaire. Following this, participants were introduced to the OT baseline; however, before passing to the practice block, they went through an “instructional check” quiz (i.e., participants had to answer correctly to questions on the task's instructions before proceeding; Finley & Penningroth, 2015). If participants responded correctly to all the questions of the instructional check quiz, they performed a short practice session of the OT, which comprised eight trials (four words and four non-words). Once participants reached an OT accuracy of at least 80%, the OT baseline block was administered. When they completed the OT,

participants were introduced to the TBPM task. Then, they performed a new instructional check quiz including the instructions of the TBPM task. As for the OT baseline, if participants responded correctly to all the questions of the instructional check quiz, the practice block was administered, which lasted approximately 2 min, allowing the participants to familiarise with the TBPM task. If participants correctly performed the PM response and reached an OT accuracy of at least 80%, the TBPM task started. Following this, participants fulfilled the follow-up questionnaire, after which they were debriefed about the aims and background of the study before quitting the experiment.

## Results

Data and results are stored in the Open Science Framework repository, and they can be downloaded from this link: <https://doi.org/10.17605/OSF.IO/9VSFX> (reference Jamovi file: “Results—main.jmv”). The analyses were carried out using *Jamovi* version 2.3.21.0 (The Jamovi Project, 2021). Descriptive statistics of the OT and TBPM performance, as well as time monitoring, are reported in Table 1. For ANOVA analyses, we calculated the effect sizes using partial eta squared values ( $\eta^2_p$ ): partial eta squared values ( $\eta^2_p$ ) of 0.0099, 0.0588, and 0.1379 were considered benchmarks for small, medium, and large effect sizes, respectively (Richardson, 2011). All the post hoc *t*-tests were carried out applying Bonferroni's correction to the *p*-values (indicated in the text as  $p_{adj}$ ); the rejection level for inferring statistical significance was set at  $p < .05$ . All analyses were controlled for the effect of age, gender, and education (in years), because the two samples were significantly different among each other: specifically, Welch's *t*-tests showed that, compared with online assessment ( $M_{age} = 26.29$ ,  $SD_{age} = 4.11$ ;  $M_{education} = 17.47$ ,  $SD_{education} = 3.84$ ), participants assessed in the laboratory were younger ( $M = 22.98$ ,  $SD = 3.37$ ),  $t(92.98) = -4.41$ ,  $p < .001$ , and reported less years of education ( $M = 15.39$ ,  $SD = 3.31$ ),  $t(94.92) = -2.92$ ,  $p = .004$ ; moreover, the number of female participants was significantly higher in the sample of participants assessed in the laboratory ( $N = 42$ ) compared with the sample assessed online ( $N = 22$ ),  $t(91.44) = 3.96$ ,  $p < .001$ . Before fitting the regression models, multi-collinearity among predictors was assessed, as it is very common in meta-regression (Berlin & Antman, 1992; Mansfield & Helms, 1982). As a rule of thumb, we establish substantial multi-collinearity if predictors showed a correlation  $r \geq .80$ . Overall, all predictors correlated with each other, but the degree of the correlation did not warrant the exclusion of these variables (maximum  $r = .47$ ,  $p < .001$ ).

### TBPM accuracy and time monitoring

TBPM accuracy was measured as the mean proportion of correct PM responses; each response was considered

**Table 1.** Descriptive statistics for time-based prospective memory accuracy, ongoing task (OT) performance, and time monitoring.

Assessment	TBPM acc.	OT accuracy		OT RTs (in seconds; correct trials only)		PM cost		Time monitoring				
		OT baseline	OT (TBPM)	OT baseline	OT (TBPM)	Acc.	RTs	t1	t2	t3	t4	
M	Laboratory	0.98	0.97	0.99	0.663	0.697	0.02	0.034	0.70	1.15	1.73	3.44
	Online	0.93	0.93	0.94	0.886	0.884	0.01	-0.002	0.62	1.13	1.59	3.34
SD	Laboratory	0.06	0.05	0.01	0.132	0.114	0.06	0.092	0.73	0.56	1.03	1.20
	Online	0.18	0.09	0.07	0.254	0.201	0.06	0.181	0.58	0.79	1.13	1.58

Mean and standard deviation for time-based prospective memory accuracy, OT performance (as accuracy and reaction times for correct trials), prospective memory cost (as accuracy and reaction times for correct trials), and time monitoring per 30-s interval (t1 vs. t2 vs. t3 vs. t4) as a function of assessment (laboratory and online). RTs: reaction times; TBPM: time-based prospective memory; OT (baseline): OT performed without time-based prospective memory task; OT (TBPM): OT concomitant with the time-based prospective memory task; t1: time 1 (i.e., first 30-s interval before the PM target time); t2: time 2 (i.e., second 30-s interval before the PM target time); t3: time 3 (i.e., third 30-s interval before the PM target time); t4: time 4 (i.e., fourth and last 30-s interval before the PM target time).

correct if it was made within  $\pm 6$ s around PM target time, equivalent to the 10% of the total PM target time's interval—i.e., 2 min (Vanneste et al., 2016). To investigate the effect of assessment on TBPM accuracy, a hierarchical multiple regression was conducted, with two models. Model 1 included the socio-demographic variables as predictors, namely, age, gender (0=male, 1=female), and education (in years), with TBPM accuracy as the dependent variable. In Model 2, assessment (0=laboratory, 1=online) was included as a predictor variable, with TBPM accuracy as a dependent variable. If the test of the model comparison was statistically significant, the alternative hypothesis was retained, meaning that the Assessment explained a significant part of the variance in the sample that was not captured by the control variables; alternatively, if the test of model comparison was not statistically significant, the null hypothesis was retained, meaning that the Assessment did not explain any significant part of the variance in the sample that was not captured by the control variables. Overall, the results showed that the first model was significant,  $F(3, 81)=3.14$ ,  $p=.030$ ,  $R^2_{adj}=.07$ ; among the predictors, only education was significantly associated with TBPM accuracy, namely, that participants with higher education did worse at the TBPM task ( $\beta=-0.30$ ,  $t=-2.49$ ,  $p=.015$ ). The second model,  $F(4, 80)=2.71$ ,  $p=.036$ ,  $R^2_{adj}=.08$ , which included Assessment ( $\beta=-0.28$ ,  $t=-1.17$ ,  $p=.244$ ) did not show significant improvement from the first model,  $\Delta F(1, 80)=1.38$ ,  $p=.244$ ,  $\Delta R^2=.02$ ; the negative effect of education was still significant in model 2 ( $\beta=-0.30$ ,  $t=-2.46$ ,  $p=.016$ ).<sup>3</sup>

For the analysis on time monitoring, we carried out a mixed ANOVA controlling for the effect of age, gender, and education. The between-subject independent variable was Assessment (laboratory vs. online), whereas the within-subject independent variable was Time (t1 vs. t2 vs. t3 vs. t4). The dependent variable was time monitoring behaviour, represented graphically in Figure 3, which was measured as the mean frequency of clock checks over 4 intervals of 30s each (i.e., “t1” represented the first 30-s

interval before the PM target time; “t2” represented the second interval; “t3” represented the third interval; and “t4” represented the fourth and last interval). This analysis aimed to investigate whether participants used the clock strategically during the TBPM task, and if the strategic time monitoring interacted with the assessment. Results showed a significant main effect of Time,  $F(2.01, 193.12)=11.76$ ,  $p<.001$ ,  $\eta^2p=0.11$ , but no significant main effect of the Assessment ( $p=.831$ ,  $\eta^2p=0.001$ ), as well as no interaction effect Time \* Assessment ( $p=.938$ ,  $\eta^2p=0.001$ ). None of the control variables (age, gender, and education) affected significantly time monitoring ( $p>.05$ ). In both Assessment conditions, the “J-shaped” monitoring pattern emerged (Figure 3). Post hoc analyses for the main effect of Time revealed that participants checked the clock less frequently in t1 ( $M=0.66$ ,  $SD=0.65$ ) compared with t2 ( $M=1.14$ ,  $SD=0.68$ ),  $t(99)=-8.99$ ,  $p_{adj}<.001$ , to t3 ( $M=1.66$ ,  $SD=1.08$ ),  $t(99)=-13.58$ ,  $p_{adj}<.001$ , and to t4 ( $M=3.39$ ,  $SD=1.39$ ),  $t(99)=-25.05$ ,  $p_{adj}<.001$ . Similarly, participants checked the clock less in t2 compared with t3,  $t(99)=-7.42$ ,  $p_{adj}<.001$ , and to t4,  $t(99)=-22.01$ ,  $p_{adj}<.001$ . Clock check frequency was significantly lower in t3 than t4,  $t(99)=-20.65$ ,  $p_{adj}<.001$ .

### OT performance and PM cost

The OT accuracy was measured as the mean proportion of correct responses (i.e., number of correct responses divided by the total number of OT trials), as well as RTs (in seconds) for correct OT trials. To investigate the effect of assessment on OT performance, a hierarchical multiple regression was conducted with two models, separately for OT block (i.e., baseline vs. during the TBPM task) as well as for accuracy and RTs; in total, four hierarchical regression analyses were conducted.<sup>4</sup> In all analyses, Model 1 included only socio-demographic variables as predictors, namely, age, gender (0=male, 1=female), and education (in years); in Model 2, assessment (0=laboratory, 1=online) was included as predictor variable. The results

of the analysis of OT accuracy during the baseline block showed that the first model was significant,  $F(3, 97)=4.84$ ,  $p=.003$ ,  $R^2_{adj}=.10$ . Model 1 showed that participants with higher education had lower OT accuracy during the baseline block ( $\beta=-0.27$ ,  $t=-2.53$ ,  $p=.13$ ); moreover, older participants tended to have a higher accuracy ( $\beta=0.23$ ,  $t=2.08$ ,  $p=.4$ ), and women performed better than men ( $\beta=0.24$ ,  $t=2.47$ ,  $p=.015$ ). The second model,  $F(4, 96)=4.81$ ,  $p=.001$ ,  $R^2_{adj}=.13$ , which included Assessment ( $\beta=-0.44$ ,  $t=-2.06$ ,  $p=.042$ ) showed a significant improvement from the first model,  $\Delta F(1, 96)=4.25$ ,  $p=.042$ ,  $\Delta R^2=.04$ ; the effect of education was still significant in Model 2 ( $\beta=-0.26$ ,  $t=-2.41$ ,  $p=.18$ ), as well as the effect of age ( $\beta=0.30$ ,  $t=-2.64$ ,  $p=.010$ ), whereas the effect of gender was no longer significant ( $p=.085$ ). The results of the analysis of OT accuracy during the TBPM block showed that the first model was not significant,  $F(3, 97)=2.57$ ,  $p=.059$ ,  $R^2_{adj}=.05$ ; among the predictors, only gender was significantly associated with OT accuracy, indicating that women performed better than men during the TBPM block ( $\beta=-0.22$ ,  $t=2.20$ ,  $p=.031$ ). The second model,  $F(4, 96)=6.88$ ,  $p<.001$ ,  $R^2_{adj}=.19$ , which included Assessment ( $\beta=-0.89$ ,  $t=-4.29$ ,  $p<.001$ ) showed significant improvement from the first model,  $\Delta F(1, 96)=18.40$ ,  $p<.001$ ,  $\Delta R^2=.15$ ; the effect of age became significant in Model 2, indicating that older participants tended to have a higher accuracy ( $\beta=0.24$ ,  $t=2.26$ ,  $p=.026$ ), whereas the effect of gender was no longer significant ( $p=.355$ ).

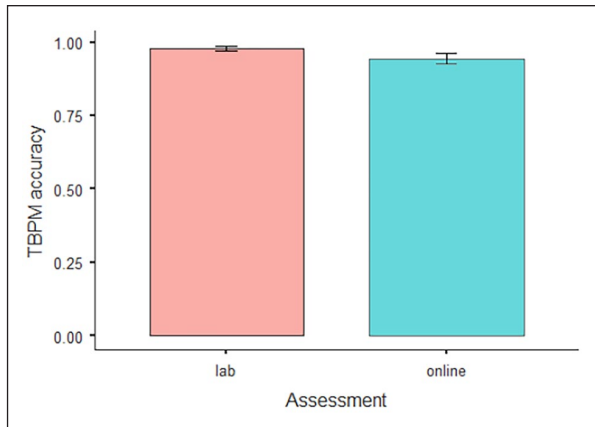
The results of the analysis of RTs for correct OT trials during the baseline block showed that the first model was significant,  $F(3, 97)=4.37$ ,  $p=.006$ ,  $R^2_{adj}=.09$ ; none of the predictors were significantly associated with RTs for correct OT trials during the baseline block. The second model,  $F(4, 96)=8.31$ ,  $p<.001$ ,  $R^2_{adj}=.23$ , which included Assessment, showed significant improvement from the first model,  $\Delta F(1, 96)=17.90$ ,  $p<.001$ ,  $\Delta R^2=.14$ ; among the predictors, only Assessment was significantly associated with RTs for correct OT trials during the baseline block ( $\beta=0.86$ ,  $t=4.23$ ,  $p<.001$ ). The results of the analysis of RTs for correct OT trials during the TBPM block showed that the first model was significant,  $F(3, 97)=5.28$ ,  $p=.002$ ,  $R^2_{adj}=.11$ ; among the predictors, only gender was significantly associated with RTs for correct OT trials during the TBPM block, indicating that women were faster than men ( $\beta=-0.21$ ,  $t=-2.26$ ,  $p=.026$ ). The second model,  $F(4, 96)=9.04$ ,  $p<.001$ ,  $R^2_{adj}=.24$ , which included Assessment showed significant improvement from the first model,  $\Delta F(1, 96)=17.61$ ,  $p<.001$ ,  $\Delta R^2=.13$ ; Among the predictors, only assessment was associated with RTs for the correct OT trials during the TBPM block ( $\beta=-0.84$ ,  $t=-4.20$ ,  $p<.001$ ), whereas the effect of gender was no longer significant ( $p=.312$ ).

The PM costs were calculated by subtracting OT performance (accuracy and RTs for correct trials) in the OT

baseline block (i.e., the lexical decision task performed with no intention) from the OT performance during the TBPM block (i.e., with a PM intention); positive values indicated that participants were more accurate, but slower in terms of RTs, at the OT during the TBPM task compared with the OT baseline block. To investigate the effect of assessment on PM costs, a hierarchical multiple regression was conducted with two models, separately for OT block (i.e., accuracy and RTs for correct OT trials); in total, two hierarchical regression analyses were conducted. In all analyses, Model 1 included only socio-demographic variables as predictors, namely, age, gender (0=male, 1=female), and education (in years); in Model 2, assessment (0=laboratory, 1=online) was included as a predictor variable. The results of the analysis of PM costs for OT accuracy rates showed that the first model was not significant,  $F(3, 97)=1.89$ ,  $p=.136$ ,  $R^2_{adj}=.03$ ; none of the predictors were significantly associated with PM cost ( $p>.05$ ). The second model,  $F(4, 96)=1.77$ ,  $p=.142$ ,  $R^2_{adj}=.03$ , which included Assessment ( $\beta=-0.27$ ,  $t=-1.17$ ,  $p=.244$ ) did not show a significant improvement from the first model,  $\Delta F(1, 96)=1.38$ ,  $p=.244$ ,  $\Delta R^2=.01$ ; as in Model 1, none of the predictors were significantly associated with OT accuracy ( $p>.05$ ). The results of the analysis of PM costs as RTs for correct OT trials showed that the first model was not significant,  $F(3, 97)=0.46$ ,  $p=.713$ ,  $R^2_{adj}=.02$ ; none of the predictors were significantly associated with PM cost ( $p>.05$ ). The second model,  $F(4, 96)=0.69$ ,  $p=.599$ ,  $R^2_{adj}=.01$ , which included Assessment ( $\beta=-0.24$ ,  $t=-1.18$ ,  $p=.241$ ) did not show a significant improvement from the first model,  $\Delta F(1, 96)=1.39$ ,  $p=.241$ ,  $\Delta R^2=.02$ ; as in Model 1, none of the predictors were significantly associated with RTs for correct OT trials ( $p>.05$ ).

## Discussion

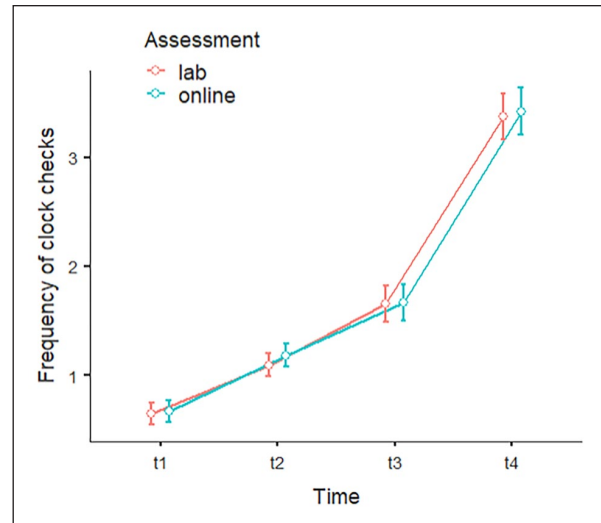
In this study, we aimed to investigate whether assessing a traditional TBPM paradigm (Einstein et al., 1995; Labelle et al., 2009; Park et al., 1997; Vanneste et al., 2016) in the laboratory and in an online assessment yield comparable results in a sample of younger adults. Given the previous evidence, we expected better OT accuracy during the laboratory compared with the online assessment, but no differences in time monitoring and TBPM accuracy (Zuber et al., 2022). Furthermore, we expected higher PM costs in the laboratory, because environmental distractions should be better controlled in the laboratory setting, and attentional resources should be selectively devoted to the maintenance of the intention in mind and to the execution of the PM task; when assessed online, participants could be generically more distracted during the OT execution, but such increase in distractibility should not be related to holding an intention in mind; therefore, PM costs should be lower in the online compared with the laboratory assessment.



**Figure 2.** Time-based prospective memory (TBPM) accuracy. Graphical representations of TBPM accuracy (maximum score = 1; error bars indicate standard error of the means). Error bars indicate the standard error of the means.

Regarding TBPM accuracy (Figure 2), our results indicated no significant differences across assessments, confirming previous evidence showing comparable accuracy levels among assessments (Zuber et al., 2022). Furthermore, we observed no noteworthy distinctions in time monitoring between laboratory and online assessments, with both conditions revealing a consistent “J-shaped” monitoring pattern (Figure 3), similar to previous lab studies (Einstein et al., 1995; Labelle et al., 2009; Mioni & Stablum, 2014). This finding is particularly important as it suggests that potential environmental cues (like the presence of clocks or the day–night cycle; Barner et al., 2019; Esposito et al., 2015; Rothen & Meier, 2017), which cannot be controlled in online assessments, did not significantly affect time monitoring in TBPM online experiments. Future studies could replicate our results including additional self-reported measures in the online assessment related to the presence of clocks in the room in which participants are performing the task or related to information regarding the time of the day during which they are performing the task. In our study, we did not ask such self-reported measures, but they would allow a more fine-grained analysis on the influence of temporally relevant environmental cues on time monitoring within uncontrolled online settings. Regarding the implications for concurrent cognitive processes in TBPM, our results indicate that the ability to strategically use internal representations of time duration over time is not affected by online assessment. Future research may explore this further by including additional measures related to time perception and task-switching to pinpoint which cognitive functions are more affected by online assessment (Germine et al., 2012; Mioni & Stablum, 2014; Varley et al., 2021).

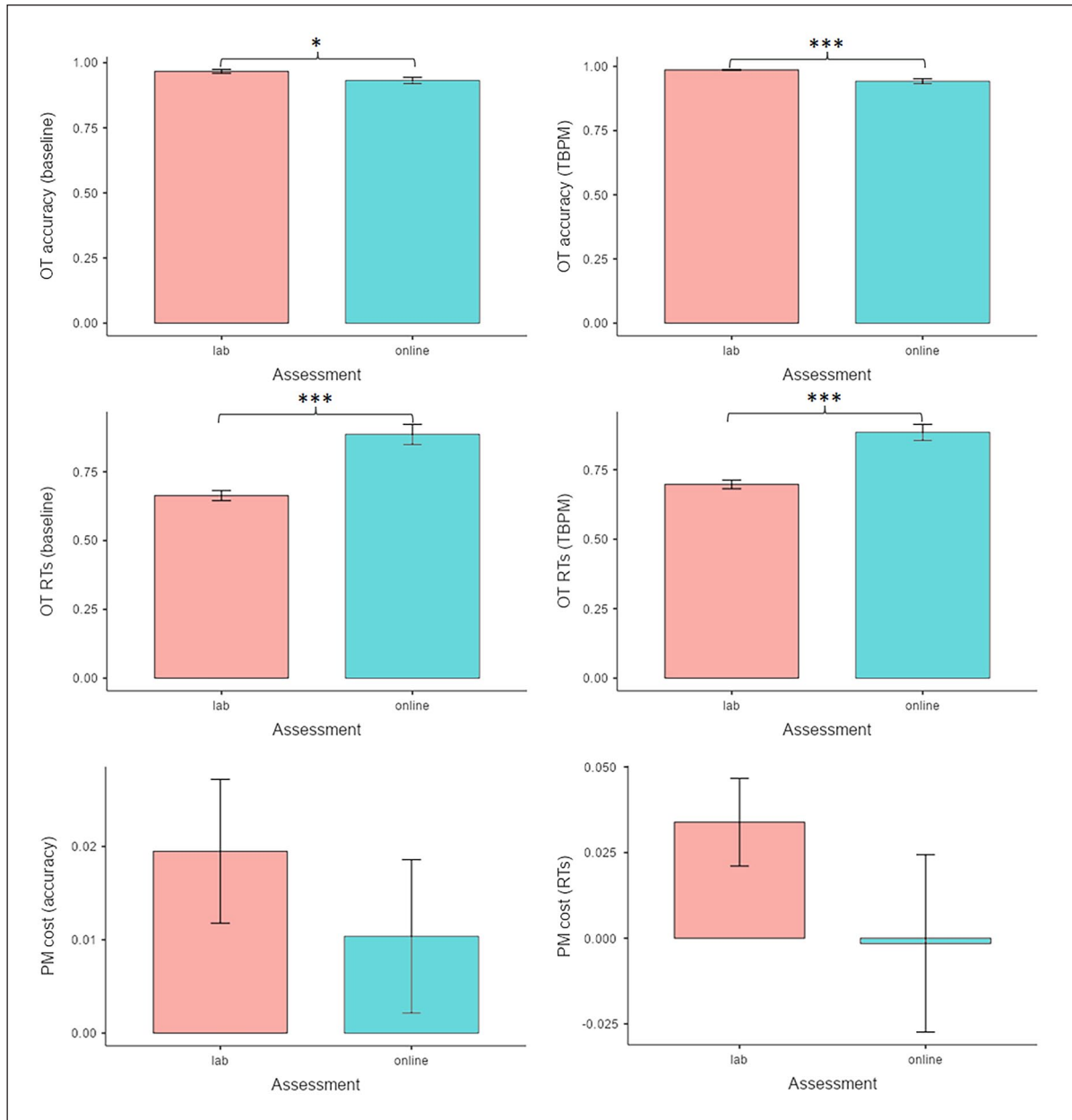
Results on OT performance showed that participants were more accurate (Figure 4, upper panels) and faster (Figure 4, middle panels) in the laboratory compared with



**Figure 3.** Time monitoring.

Graphical representations of time monitoring as the frequency of clock checks frequency per 30 s interval as a function of the assessment (laboratory vs. online). Error bars indicate standard error of the means; t1: time 1 (i.e., first 30-s interval before the PM target time); t2: time 2 (i.e., second 30-s interval before the PM target time); t3: time 3 (i.e., third 30-s interval before the PM target time); t4: time 4 (i.e., fourth and last 30-s interval before the PM target time).

the online assessments, consistent with previous evidence (Zuber et al., 2022). This suggested that online participants were more easily distracted, aligning with previous research (Logie & Maylor, 2009; Zuber et al., 2022). No PM cost was found in the two samples (Figure 4, lower panels) suggesting not only that holding an intention in mind did not affect OT performance (see also the Online Supplementary materials for more information), but also that, even though online participants may have been more distractible during the OT, this did not affect their ability to maintain and execute time-based intentions. Unexpectedly, when OT was performed concurrently with the TBPM task, participants were more accurate (Figure 4, left lower panel), contrary to previous studies showing decreased accuracy in such conditions (Hicks et al., 2005; Huang et al., 2014; McBride & Flaherty, 2020; Peper & Ball, 2022). However, the results in terms of reaction times were descriptively consistent with the presence of a cost related to holding an intention in mind, at least for laboratory-tested participants (Figure 4, right lower panel). Although not statistically significant, these results were in line with our predictions, at least descriptively, indicating that, for participants assessed in the laboratory, attentional resources were devoted mainly to the maintenance of the intention in mind and to the execution of the PM task, whereas participants assessed online were generically more distracted during the OT execution, but such increase in distractibility was not related to holding an intention in mind. However, the reader should bear in mind that we did not ask explicitly participants whether they were distracted;



**Figure 4.** Ongoing task (OT) performance and prospective memory (PM) cost.

Graphical representations of OT performance in terms of accuracy (upper panels; maximum accuracy = 1), and reaction times for correct trials (middle panels) divided by baseline block (left upper and middle panels) and TBPM block (right upper and middle panels) as a function of assessment (laboratory vs. online); PM cost is depicted in the lower panels, in terms of both cost in OT accuracy (left lower panel) and reaction times (right lower panel) as a function of assessment (laboratory vs. online). Error bars indicate the standard error of the means. RTs: reaction times. \* $p < .05$ . \*\*\* $p < .001$ .

future studies are needed to do so in order (1) to investigate the degree of self-reported distractibility across assessments and (2) to assess whether these subjective reports do correlate with the behavioural performance.

There might be at least three reasons that can explain such results. One reason could be related to the fact that participants always performed the OT baseline before the TBPM task; such a procedure might have caused learning effects that can explain the increase in OT accuracy during

the TBPM task compared with the OT baseline block (Guo et al., 2019). A second reason could be related to the fact that the OT baseline comprised fewer trials (32) compared with the TBPM block (from 136 to 194), which can affect the magnitude of the PM costs, especially for reaction times (Lachaud & Renaud, 2011; Sternberg, 2010). Thus, future studies could replicate our results including an OT baseline block with the same amount of trials comprised in the TBPM task, counterbalancing the task execution to

avoid potential learning effects that can affect the presence, direction, and magnitude of the PM cost. Finally, a third reason is related to the presence of PM cost in the TBPM paradigm, which has been questioned by several studies: although there is still a debate on the nature of PM cost and on how attentional resources and processes are distributed between OT and PM task (e.g., Anderson et al., 2019), some studies showed that attentional resources and processes are quite independent in TBPM tasks (i.e., the elaboration of the OT does not prevent participants to perform the TBPM task correctly). Diversely, this is not the case for EBPM, in which the OT stimuli always contain the PM cue: in this case, the elaboration of the OT does prevent participants from performing the EBPM task correctly<sup>5</sup> (Conte & McBride, 2018; McBride & Flaherty, 2020; Zuber & Kliegel, 2020). Therefore, in this sample of participants, perhaps PM cost might have been attenuated in the presence of TBPM tasks. Future studies can compare EBPM and TBPM between laboratory and online assessments to further elucidate these assumptions.

All analyses in this study included age, gender, and education to ensure that differences in assessments were not influenced by socio-demographic factors (Uittenhove et al., 2023). We found that age positively predicted OT accuracy, aligning with prior research suggesting older individuals perform better in lexical decision tasks (e.g., Ratcliff et al., 2004) due to increased vocabulary (Baltes et al., 2007). Surprisingly, education negatively affected accuracy in both the OT baseline and TBPM tasks, contrary to typical findings (Ihle et al., 2022; Wang et al., 2009). However, our study focused on a limited age range of younger adults, warranting caution in interpreting these results (Joly-Burra et al., 2022). Moreover, it is important to notice that the relationship between education and TBPM performance may be influenced by other factors, such as differences in socio-economic status, lifestyle, or occupation; further research is needed to fully understand this relationship (Uittenhove et al., 2023).

This study has several limitations. One possible limitation concerns some methodological differences between laboratory and online assessment (e.g., difference in the number of PM tasks within the TBPM block, or the presence of instruction quizzes in the online assessment). As mentioned above, these methodological differences were designed to maximise the quality of the data obtained online, and we believe that, in this way, this study can provide a first guidance for future studies that aim to assess the laboratory-based TBPM paradigm online, not only concerning specific conceptual aspects (e.g., we showed that participants online were likely to use no other temporal sources than the clock provided by the task's procedure to estimate the occurrence of the PM target time) but also in terms of the methodological design of the paradigm, as we showed that four PM cues are already enough to provide a comparable performance between assessment

settings. Nonetheless, future studies are needed to further replicate the results of this study using identical procedures. Another consideration is the duration of the PM target time and the overall experimental procedure length. We used a 2-min PM target time to keep the procedure within 30 min, given that participants tend to disengage after this time online (e.g., Logie & Maylor, 2009). It remains unknown how participants would perform when the PM target time is longer (e.g., 5 min) and the procedure exceeds 30 min. Future studies can increase the duration of the PM target time and the overall duration of the experimental procedure to investigate whether participants still show similar PM performance and clock-checking between laboratory and online assessments. Another limitation of this study concerns the fact that we did not explicitly inquire about cheating, but we noted instances where participants stopped the task, possibly to cheat or earn money without engagement. Future research should replicate our findings and investigate cheating more thoroughly by asking participants whether and how they cheated. Finally, we did not mention removing any time-related external factor in the instructions, because any reference to "temporal" environmental in the instructions could have gained salience for the participants and, in turn, might have pushed them to cheat, paradoxically. Instead, we provide an instruction slide with the following statement: "You are about to begin the experiment. It is ideal to start the experience in an enclosed, quiet, and comfortable place isolated from any type of distraction," so without any explicit reference on the involvement of time in the tasks. Yet, despite the lack of such preventive measures, we did not find any differences in TBPM (as both accuracy and time monitoring), suggesting that, if present in the sample, participants who cheated did not affect the behavioural pattern of TBPM performance and time monitoring. However, this is still an interesting research question for future studies, which can replicate our results including instructions that explicitly ask participants (1) to remove time-related environmental cues and (2) to indicate which time-related cues were present (and eventually removed) from the surrounding environment.

Overall, this is the first study that investigated whether a traditional TBPM task yielded comparable results when administered online and in the laboratory. Importantly, we demonstrated that the important pattern of strategic time monitoring emerges in a comparable way for both settings, leading to comparable TBPM performance across assessments, and even in a between-subjects design. Hence, given that the pattern of results did not change among the assessments, and considering the lack of full control during online testing, the present results may lend initial and cautious support in favour of studying TBPM remotely. In this regard, although participants assessed online were more distracted during the OT execution compared with the participants assessed in the laboratory, such an increase

in distractibility was not related to holding an intention in mind, as shown by the absence of the PM cost. Future research that wants to include online assessment along with the traditional laboratory assessment can easily control for the main effect of the assessment's setting.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: MK acknowledges funding from the Swiss National Science Foundation (SNSF).

### ORCID iD

Gianvito Laera  <https://orcid.org/0000-0002-0589-7334>

### Supplementary material

The supplementary material is available at: [qjep.sagepub.com](http://qjep.sagepub.com).

### Data accessibility statement



The data and materials from the present experiment are publicly available at the Open Science Framework website: <https://osf.io/9vsfx>

### Notes

1. The experiment's code is also available at the following link: <https://gitlab.pavlovia.org/Laera/time-based-prospective-memory-demo>.
2. The only question that differed between the socio-demographic and the follow-up questionnaire online was the question about age: in the socio-demographic questionnaire, we asked to provide the age in years, whereas in the follow-up questionnaire, we asked to provide their date of birth. Then, we took only the year, calculated the actual age according to this information, and compared it with the age (in years) provided by the participant during the socio-demographic questionnaire.
3. We carried out further analysis on TBPM accuracy without the fifth—and last—PM cue (i.e., making the comparison between laboratory and online assessment 4 vs. 4). Overall, the results do not change, allowing us to conclude that the results' pattern was not due to different numbers of PM tasks within the block (for more information, see the Online Supplementary materials).
4. Further analyses were conducted using mixed-design ANOVAs on both OT accuracy and RTs for correct trials, as well as on measures of PM cost, with two further repeated measures: (1) the OT block (i.e., OT baseline vs. OT during TBPM task) and (2) the time (t1 vs. t2 vs. t3 vs. t4; similarly to the analysis on time monitoring). The purpose was to investigate whether the results changed when the model took into account the effect of OT block over time on OT performance and intention-related attentional processes. Overall, the results do not change; moreover, no significant PM cost was found (see the Online Supplementary materials for more information).
5. Event-based PM is assessed in the laboratory by asking people to remember to perform a specific action in response to a particular external cue while engaged in the OT (Einstein & McDaniel, 1990; Park et al., 1997). The PM cue is embodied within the OT stimuli (e.g., if the OT is a lexical decision task, the PM cue could be a specific word or a syllable within the string of letters); hence, the elaboration of the OT material is essential to perform the PM task correctly; diversely, the OT does not contain the PM cue in the TBPM paradigm (i.e., the PM cue is time itself, which is embodied in the clock and, as such, is completely independent of the OT material); hence, participants do not need to elaborate the OT material to perform the TBPM task (Zuber & Kliegel, 2020).

### References

- Anderson, F. T., Strube, M. J., & McDaniel, M. A. (2019). Toward a better understanding of costs in prospective memory: A meta-analytic review. *Psychological Bulletin, 145*(11), 1053–1081. <https://doi.org/10.1037/bul0000208>
- Backx, R., Skirrow, C., Dente, P., Barnett, J. H., & Cormack, F. K. (2020). Comparing web-based and lab-based cognitive assessment using the Cambridge Neuropsychological Test Automated Battery: A within-subjects counterbalanced study. *Journal of Medical Internet Research, 22*(8), Article e16792. <https://doi.org/10.2196/16792>
- Baltes, P. B., Lindenberger, U., & Staudinger, U. M. (2007). Life span theory in developmental psychology. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 569–664). John Wiley & Sons. <https://doi.org/10.1002/9780470147658.chpsy0111>
- Barner, C., Schmid, S. R., & Diekelmann, S. (2019). Time-of-day effects on prospective memory. *Behavioural Brain Research, 376*, 112179. <https://doi.org/10.1016/j.bbr.2019.112179>
- Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: Joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *The Clinical Neuropsychologist, 26*(2), 177–196. <https://doi.org/10.1080/13854046.2012.663001>
- Berlin, J. A., & Antman, E. M. (1992). Advantages and limitations of meta-analytic regressions of clinical trials data. *Controlled Clinical Trials, 13*(5), 422. [https://doi.org/10.1016/0197-2456\(92\)90151-O](https://doi.org/10.1016/0197-2456(92)90151-O)
- Boag, R. J., Strickland, L., Heathcote, A., Neal, A., & Loft, S. (2019). Cognitive control and capacity for prospective memory in complex dynamic environments. *Journal of Experimental Psychology: General, 148*(12), 2181–2206. <https://doi.org/10.1037/xge0000599>
- Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ, 8*, Article e9414. <https://doi.org/10.7717/peerj.9414>

- Burgess, P. W., Scott, S. K., & Frith, C. D. (2003). The role of the rostral frontal cortex (area 10) in prospective memory: A lateral versus medial dissociation. *Neuropsychologia*, *41*(8), 906–918.
- Ceci, S. J., & Bronfenbrenner, U. (1985). “Don’t Forget to Take the Cupcakes out of the Oven”: prospective memory, strategic time-monitoring, and context. *Child Development*, *56*(1), 152–164. <https://doi.org/10.2307/1130182>
- Cohen, J. (1969). *Statistical power analysis for the behavioural sciences*. Academic Press. <https://doi.org/10.4324/9780203771587>
- Conte, A. M., & McBride, D. M. (2018). Comparing time-based and event-based prospective memory over short delays. *Memory*, *26*(7), 936–945. <https://doi.org/10.1080/09658211.2018.1432662>
- Crovitz, H. F., & Daniel, W. F. (1984). Measurements of everyday memory: Toward the prevention of forgetting. *Bulletin of the Psychonomic Society*, *22*, 413–414. <https://doi.org/10.3758/BF03333861>
- Del Missier, F., Stragà, M., Visentini, M., Munareto, G., & Mäntylä, T. (2021). Intention framing in time-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(8), 1226–1245. <https://doi.org/10.1037/xlm0000991>
- Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 717–726.
- Einstein, G. O., McDaniel, M. A., Richardson, S. L., Guynn, M. J., & Cunfer, A. R. (1995). Aging and prospective memory: Examining the influences of self-initiated retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 996–1007. <https://doi.org/10.1037/0278-7393.21.4.996>
- Esposito, M. J., Occhionero, M., & Cicogna, P. (2015). Sleep deprivation and time-based prospective memory. *Sleep*, *38*(11), 1823–1826. <https://doi.org/10.5665/sleep.5172>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Feenstra, H. E. M., Vermeulen, I. E., Murre, J. M. J., & Schagen, S. B. (2017). Online cognition: Factors facilitating reliable online neuropsychological test results. *The Clinical Neuropsychologist*, *31*(1), 59–84. <https://doi.org/10.1080/13854046.2016.1190405>
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2010). The French lexicon project: Lexical decision data for 38,840 French words and 38,840 pseudo words. *Behavior Research Methods*, *42*(2), 488–496. <https://doi.org/10.3758/BRM.42.2.488>
- Finley, A. J., & Penningroth, S. L. (2015). Online versus in-lab: Pros and cons of an online prospective memory experiment. *Advances in Psychology Research*, *113*, 135–162.
- Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*(5), 847–857. <https://doi.org/10.3758/s13423-012-0296-9>
- Gilbert, S. J. (2015a). Strategic offloading of delayed intentions into the external environment. *Quarterly Journal of Experimental Psychology*, *68*(5), 971–992. <https://doi.org/10.1080/17470218.2014.972963>
- Gilbert, S. J. (2015b). Strategic use of reminders: Influence of both domain-general and task-specific metacognitive confidence, independent of objective memory ability. *Consciousness and Cognition*, *33*, 245–260. <https://doi.org/10.1016/j.concog.2015.01.006>
- Gladwin, T. E., Jewiss, M., Banic, M., & Pereira, A. (2020). Associations between performance-based and self-reported prospective memory, impulsivity and encoding support. *Acta Psychologica*, *206*, 103066. <https://doi.org/10.1016/j.actpsy.2020.103066>
- Guo, Y., & Huang, X. (2019). Time-based prospective memory has plasticity in behavior under different monitoring conditions. *Current Psychology*, *40*, 3386–3392. <https://doi.org/10.1007/s12144-019-00270-5>
- Guo, Y., Liu, P., & Huang, X. (2019). The practice effect on time-based prospective memory: The influences of ongoing task difficulty and delay. *Frontiers in Psychology*, *10*, Article 2002. <https://doi.org/10.3389/fpsyg.2019.02002>
- Haas, M., Zuber, S., Kliegel, M., & Ballhausen, N. (2020). Prospective memory errors in everyday life: Does instruction matter? *Memory*, *28*(2), 196–203. <https://doi.org/10.1080/09658211.2019.1707227>
- Harris, J., & Wilkins, A. (1982). Remembering to do things: A theoretical framework and an illustrative experiment. *Human Learning: Journal of Practical Research Applications*, *1*, 123–136.
- Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! A delay theory of prospective memory costs. *Psychological Review*, *122*(2), 376–410. <https://doi.org/10.1037/a0038952>
- Hering, A., Kliegel, M., Rendell, P. G., Craik, F. I. M., & Rose, N. S. (2018). Prospective memory is a key predictor of functional independence in older adults. *Journal of the International Neuropsychological Society*, *24*(6), 640–645. <https://doi.org/10.1017/S1355617718000152>
- Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). Task interference in time-based, event-based, and dual intention prospective memory conditions. *Journal of Memory and Language*, *53*(3), 430–444. <https://doi.org/10.1016/j.jml.2005.04.001>
- Horn, S. S., & Freund, A. M. (2021). How do gain and loss incentives affect memory for intentions across adulthood? *The Journals of Gerontology: Series B*, *76*(4), 711–721. <https://doi.org/10.1093/geronb/gbaa140>
- Hoskins, L. L., Binder, L. M., Chaytor, N. S., Williamson, D. J., & Drane, D. L. (2010). Comparison of oral and computerized versions of the word memory test. *Archives of Clinical Neuropsychology*, *25*(7), 591–600. <https://doi.org/10.1093/arclin/acq060>
- Huang, T., Loft, S., & Humphreys, M. S. (2014). Internalizing versus externalizing control: Different ways to perform a time-based prospective memory task. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(4), 1064–1071. <https://doi.org/10.1037/a0035786>
- Ihle, A., Gouveia, É. R., Gouveia, B. R., Marques, A., Marconcin, P., de Maio Nascimento, M., Jurema, J., Tinôco, M. A., & Kliegel, M. (2022). Cognitive functioning mediates the association of cognitive reserve with health-related quality of life. *Sustainability*, *14*(2), 826. <https://doi.org/10.3390/su14020826>

- The Jamovi Project. (2021). *Jamovi* (Version 1.6) [Computer software]. <https://www.jamovi.org/>
- Joly-Burra, E., Haas, M., Laera, G., Ghisletta, P., Kliegel, M., & Zuber, S. (2022). Frequency and strategicness of clock-checking explain detrimental age effects in time-based prospective memory. *Psychology and Aging, 37*(5), 637–648. <https://doi.org/10.1037/pag0000693>
- Kvavilashvili, L., & Fisher, L. (2007). Is time-based prospective remembering mediated by self-initiated rehearsals? Role of incidental cues, ongoing activity, age, and motivation. *Journal of Experimental Psychology: General, 136*(1), 112–132. <https://doi.org/10.1037/0096-3445.136.1.112>
- Kvavilashvili, L., Kornbrot, D. E., Mash, V., Cockburn, J., & Milne, A. (2009). Differential effects of age on prospective and retrospective memory tasks in young, young-old, and old-old adults. *Memory, 17*(2), 180–196. <https://doi.org/10.1080/09658210802194366>
- Labelle, M. A., Graf, P., Grondin, S., & Gagné-Roy, L. (2009). Time-related processes in time-based prospective memory and in time-interval production. *European Journal of Cognitive Psychology, 21*(4), 501–521. <https://doi.org/10.1080/09541440802031000>
- Lachaud, C. M., & Renaud, O. (2011). A tutorial for analyzing human reaction times: How to filter data, manage missing values, and choose a statistical model. *Applied Psycholinguistics, 32*(2), 389–416. <https://doi.org/10.1017/S0142716410000457>
- Laera, G., Joly-Burra, E., Zuber, S., Ballhausen, N., Künzi, M., Ihle, A., da Silva Coelho, C., Haas, M., Mikneviciute, G., Tinello, D., Kliegel, M., & Hering, A. (2023). Do executive functions explain older adults' health-related quality of life beyond event-based prospective memory? *Aging, Neuropsychology, and Cognition, 30*, 135–149. <https://doi.org/10.1080/13825585.2021.1989368>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, Article 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Loft, S., Dismukes, K., & Grundgeiger, T. (2021). Prospective memory in safety-critical work contexts. In J. Rummel (Ed.), *Current issues in memory: Memory research in the public interest* (pp. 192–207). Routledge.
- Logie, R. H., & Maylor, E. A. (2009). An Internet study of prospective memory across adulthood. *Psychology and Aging, 24*(3), 767–774. <https://doi.org/10.1037/a0015479>
- Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. *The American Statistician, 36*(3a), 158–160. <https://doi.org/10.1080/00031305.1982.10482818>
- Mäntylä, T., Carelli, M. G., & Forman, H. (2006). Time monitoring and executive functioning in children and adults. *Journal of Experimental Child Psychology, 96*(1), 1–19. <https://doi.org/10.1016/j.jecp.2006.08.003>
- Mattli, F., Schnitzspahn, K. M., Studerus-Germann, A., Brehmer, Y., & Zöllig, J. (2014). Prospective memory across the lifespan: Investigating the contribution of retrospective and prospective processes. *Aging, Neuropsychology, and Cognition, 21*(5), 515–543. <https://doi.org/10.1080/13825585.2013.837860>
- McBride, D. M., & Flaherty, M. (2020). Comparing costs in time-based and event-based prospective memory. *Memory, 28*(7), 918–925. <https://doi.org/10.1080/09658211.2020.1798463>
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227–234. <https://doi.org/10.1037/h0031564>
- Mioni, G., Grondin, S., McLennan, S. N., & Stablum, F. (2020). The role of time-monitoring behaviour in time-based prospective memory performance in younger and older adults. *Memory, 28*(1), 34–48. <https://doi.org/10.1080/09658211.2019.1675711>
- Mioni, G., Santon, S., Stablum, F., & Cornoldi, C. (2017). Time-based prospective memory difficulties in children with ADHD and the role of time perception and working memory. *Child Neuropsychology, 23*(5), 588–608. <https://doi.org/10.1080/09297049.2016.1172561>
- Mioni, G., & Stablum, F. (2014). Monitoring behaviour in a time-based prospective memory task: The involvement of executive functions and time perception. *Memory, 22*(5), 536–552. <https://doi.org/10.1080/09658211.2013.801987>
- Mioni, G., Stablum, F., McClintock, S. M., & Cantagallo, A. (2012). Time-based prospective memory in severe traumatic brain injury patients: The involvement of executive functions and time perception. *Journal of the International Neuropsychological Society, 18*(4), 697–705. <https://doi.org/10.1017/S1355617712000306>
- Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Yamadori, A., Frith, C. D., & Burgess, P. W. (2007). Differential involvement of regions of rostral prefrontal cortex (Brodmann area 10) in time- and event-based prospective memory. *International Journal of Psychophysiology, 64*(3), 233–246. <https://doi.org/10.1016/j.ijpsycho.2006.09.009>
- Park, D. C., Hertzog, C., Kidder, D. P., Morrell, R. W., & Mayhorn, C. B. (1997). Effect of age on event-based and time-based prospective memory. *Psychology and Aging, 12*(2), 314–327.
- Parsons, T. D., McMahan, T., & Kane, R. (2018). Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms. *The Clinical Neuropsychologist, 32*(1), 16–41. <https://doi.org/10.1080/13854046.2017.1337932>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Peper, P., & Ball, H. (2022). *Strategic monitoring improves prospective memory: A meta-analysis (Revision 1)* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/ajsvh>
- Raskin, S. A., Shum, D. H. K., Ellis, J., Pereira, A., & Mills, G. (2018). A comparison of laboratory, clinical, and self-report measures of prospective memory in healthy adults and individuals with brain injury. *Journal of Clinical and Experimental Neuropsychology, 40*(5), 423–436. <https://doi.org/10.1080/13803395.2017.1371280>
- Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the

- lexical-decision task. *Psychology and Aging*, *19*(2), 278–289. <https://doi.org/10.1037/0882-7974.19.2.278>
- Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. <https://doi.org/10.1016/j.edurev.2010.12.001>
- Rothen, N., & Meier, B. (2017). Time-of-day affects prospective memory differently in younger and older adults. *Aging, Neuropsychology, and Cognition*, *24*(6), 600–612. <https://doi.org/10.1080/13825585.2016.1238444>
- Scarampi, C., & Gilbert, S. J. (2020). The effect of recent reminder setting on subsequent strategy and performance in a prospective memory task. *Memory*, *28*(5), 677–691. <https://doi.org/10.1080/09658211.2020.1764974>
- Skitka, L. J., & Sargis, E. G. (2006). The Internet as psychological laboratory. *Annual Review of Psychology*, *57*(1), 529–555. <https://doi.org/10.1146/annurev.psych.57.102904.190048>
- Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: Investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 347–361. <https://doi.org/10.1037/0278-7393.29.3.347>
- Sternberg, S. (2010). *Reaction-time experimentation*. <https://www.sas.upenn.edu/~saul/rt.experimentation.pdf>
- Suchy, Y., Ziemnik, R. E., Niermeyer, M. A., & Brothers, S. L. (2020). Executive functioning interacts with complexity of daily life in predicting daily medication management among older adults. *The Clinical Neuropsychologist*, *34*(4), 797–825. <https://doi.org/10.1080/13854046.2019.1694702>
- Uittenhove, K., Jeanneret, S., & Vergauwe, E. (2023). From lab-testing to web-testing in cognitive research: Who you test is more important than how you test. *Journal of Cognition*, *6*(1), Article 1. <https://doi.org/10.5334/joc.259>
- Vanneste, S., Baudouin, A., Bouazzaoui, B., & Taconnat, L. (2016). Age-related differences in time-based prospective memory: The role of time estimation in the clock monitoring strategy. *Memory*, *24*(6), 812–825. <https://doi.org/10.1080/09658211.2015.1054837>
- Varley, D., Henry, J. D., Gibson, E., Suddendorf, T., Rendell, P. G., & Redshaw, J. (2021). An old problem revisited: How sensitive is time-based prospective memory to age-related differences? *Psychology and Aging*, *36*(5), 616–625. <https://doi.org/10.1037/pag0000625>
- Wang, Y., Cui, J., Chan, R. C. K., Deng, Y., Shi, H., Hong, X., Li, Z., Yu, X., Gong, Q., & Shum, D. (2009). Meta-analysis of prospective memory in schizophrenia: Nature, extent, and correlates. *Schizophrenia Research*, *114*(1–3), 64–70. <https://doi.org/10.1016/j.schres.2009.07.009>
- Woods, S. P., Weinborn, M., Li, Y. R., Hodgson, E., Ng, A. R. J., & Bucks, R. S. (2015). Does prospective memory influence quality of life in community-dwelling older adults? *Aging, Neuropsychology, and Cognition*, *22*(6), 679–692. <https://doi.org/10.1080/13825585.2015.1027651>
- Zuber, S., Haas, M., Framorando, D., Ballhausen, N., Gillioz, E., Künzi, M., & Kliegel, M. (2022). The Geneva Space Cruiser: A fully self-administered online tool to assess prospective memory across the adult lifespan. *Memory*, *30* (2), 117–132. <https://doi.org/10.1080/09658211.2021.1995435>
- Zuber, S., & Kliegel, M. (2020). Prospective memory development across the lifespan: An integrative framework. *European Psychologist*, *25*(3), 162–173. <https://doi.org/10.1027/1016-9040/a000380>